

## Original article

# A deep-learning approach for reservoir evaluation for shale gas wells with complex fracture networks

Hongyang Chu<sup>1,2,3</sup>, Peng Dong<sup>2</sup>\*, W. John Lee<sup>3</sup>

<sup>1</sup>*School of Civil and Resources Engineering, University of Science and Technology Beijing, Beijing 100083, P. R. China*

<sup>2</sup>*State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing 102249, P. R. China*

<sup>3</sup>*Harold Vance Department of Petroleum Engineering, Texas A&M University, College Station 77843, USA*

### Keywords:

Deep-Learning  
time series analysis  
multivariate input  
shale gas  
reservoir evaluation

### Cited as:

Chu, H., Dong, P., Lee, W. J. A deep-learning approach for reservoir evaluation for shale gas wells with complex fracture networks. *Advances in Geo-Energy Research*, 2023, 7(1): 49-65. <https://doi.org/10.46690/ager.2023.01.06>

### Abstract:

The complex fracture networks in shale gas reservoirs bring greater challenges and uncertainties to the modeling in reservoir evaluation. As the emerging potential technology, deep learning can be usefully applied to many aspects of reservoir evaluation. To further conduct the reservoir evaluation in rate transient analysis, this work proposes a data-driven proxy model for accurately evaluating the horizontal wells with complex fracture networks in shales. The production time, variable bottom hole pressure, and the fracture networks properties are used as input variables, while the output variable refers to the production for the forecast time period. The data from boundary element method is used to generate the proxy model for the learning process. The method of shuffled cross-validation is used to increase the model's accuracy and generalizability. The proxy model is coupled with recently developed deep learning methods such as attention mechanism, skip connection, and cross-validation to address the time series analysis problem for multivariate operating and physical parameters. Results demonstrate that the attention mechanism is robust. The operating parameters analysis shows that the attention mechanism has the ability to analyze variable pressure drop/flowrate data. Sensitivity analysis also indicates that the model takes into account the geometric characteristics of fracture network. The model reliability is proved by a case study from Marcellus shale. The computation time of the trained attention mechanism model is approximately 0.3 s, which equates to 3.8% of the physical model's running time.

## 1. Introduction

Due to advancements in horizontal well technology and hydraulic fracturing treatment, shale gas has gained increased attention as an unconventional gas resource (Denney, 2009; Nobakht et al., 2013; Zhao et al., 2022). Forecasting production is a critical task in shale gas studies, as it can aid in estimating in-place resources and facilitate reservoir evaluation (Neal and Mian, 1989; Lee and Sidle, 2010; Zhao and Du, 2020). There are several challenges associated with forecasting production from shale gas reservoirs, including transportation issues caused by the complex storage and transportation mechanisms for shale gas, which includes gas diffusion and adsorption (Clarkson et al., 2013), and formation

complexity caused by hydraulic fracturing treatment, which involves complex hydraulic fracture networks (Curtis, 2002; Fink et al., 2017; Hamdi et al., 2020).

The complex mechanism of gas transport in shales complicates the physical model. Carlson and Mercer (1991) demonstrated that, in contrast to natural gas stored in open pores in conventional gas reservoirs, the majority of gas in shale reservoirs is adsorbed. The gas flow in the shale matrix is primarily controlled by molecular diffusion. Kucuk and Sawyer (1980) developed analytical and numerical solutions for shale reservoirs that take the Klinkenberg effect and gas desorption in the shale matrix into account but ignore shale gas diffusion and adsorption. Gao et al. (1994) proposed a model that incorporates sorption effects and describes the desorption

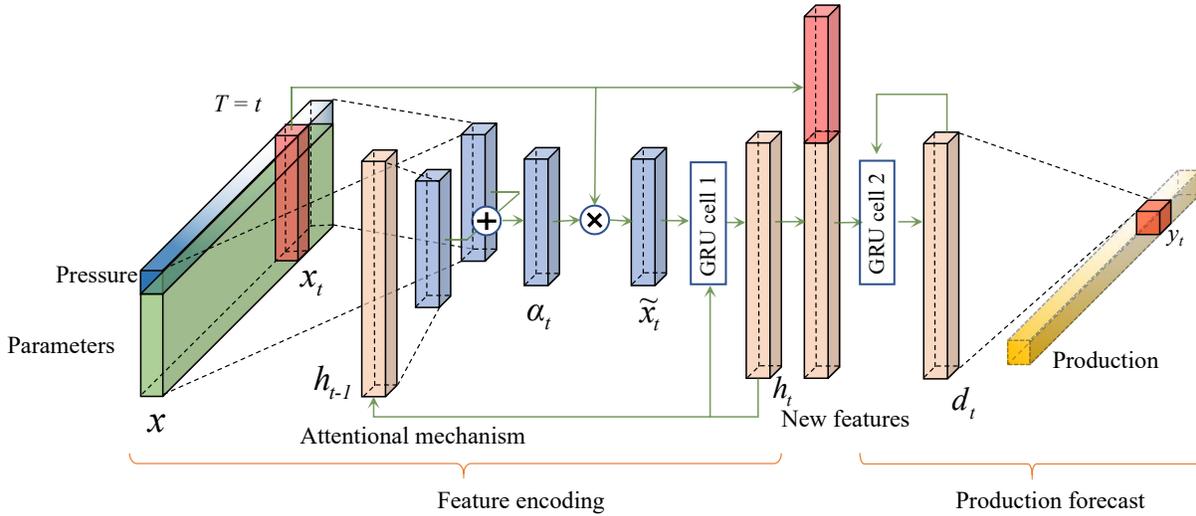
behaviors of shale gas using the Langmuir isotherm. Ozkan et al. (2010) proposed the first dual-mechanism model for hydraulically fractured horizontal wells (HFHW) in shales, taking into account Knudsen diffusive flow in matrix nanopores and stress-dependent effects in the natural fractures system. The proposed model is suitable for analytical and numerical modeling, but it ignores the gas adsorption effect. More numerical methods have been developed to account for a variety of physics in shales. Cipolla et al. (2011) proposed a numerical model for shale gas that described gas desorption, stress-dependent fractures, and complex fracture networks. Their research examined the effect of gas adsorption on the production profile and ultimate gas recovery in shales in detail. Freeman et al. (2013) proposed a numerical model for gas desorption, multi-porosity and permeability fields, and HFHW using the finite difference method. Subsequently, the simulator was used to clarify various flow regimes encountered during production. Wang (2014) developed a multiple-mechanism shale gas model using the boundary element method (BEM), Laplace transform, and Gaussian elimination method to incorporate gas adsorption/diffusion effects and HFHW into the derived model. The HFHW model incorporates diffuse/viscous flow, adsorption, and stress-sensitive effects. Additionally, Sun et al. (2015) discussed the development of a numerical model that incorporates the shale gas mechanistic of desorption, diffusion, convection, and multi-porosity/permeability fields. The preceding significant works provide valuable tools for forecasting shale gas well production performance.

The results from numerical methods are accurate. However, numerical simulation tools based on physics are typically cumbersome, time-consuming, and occasionally costly, deviating from the practical premise of reservoir evaluation (Aminian and Ameri, 2009). Mesh refinement methods make numerical simulation tools take hours or even days to complete the reservoir evaluation for a well with complex fracture networks. Additionally, the majority of research is limited to HFHW in shale gas reservoirs. Microseismic events demonstrate the formation of complex fracture networks during large-scale hydraulic fracturing treatment (Warpinski et al., 2009; Cipolla et al., 2011). Complex fracture networks lead to the difficulties for the numerical and analytical modeling. For analytical modeling, it can only be analyzed by the simplified method of linear flow or elliptical flow. For numerical modeling, it requires a more refined gridding technique, and the meshing-size differences will deteriorate the numerical model convergence when simulating fracture networks with millimeter-scale widths. Due to the fact that these phenomena add to the computational burden of the numerical model and the modeling difficulty of the analytical model, the demand for the proxy model is both urgent and attractive.

The state-of-the-art oil and gas production forecast methods can be divided into conventional statistical techniques (including autoregressive moving average, autoregressive conditional heteroskedasticity model, etc), machine learning models (support vector machine, random forest), and deep learning-based methods (recurrent neural network, long short-term memory networks, gate recurrent unit, etc). Recent advances in deep learning techniques have facilitated the generation of

data-driven proxy models based on physics (Jia and Zhang, 2016; Zhang et al., 2019; Yin et al., 2021; Zhong et al., 2021; Liu et al., 2022; Luo et al., 2022). Neural network is frequently used in conjunction with time series analysis (TSA) techniques to solve time series regression tasks, such as production forecasting. Choi et al. (2015) used seasonal autoregressive integrated moving average model to predict future oil production in North Dakota. Jia and Zhang (2016) demonstrated how to use the TSA and artificial neural networks (ANN) to analyze Barnett shale gas well production data. They also admitted that when the input of the ANN is expanded to include additional operational and physical parameters, it may be more beneficial to forecast production. Suhag et al. (2017) made a similar comparison between the data-driven method composed of ANN and TSA and the empirical decline curve analysis (DCA) method. Their model uses well logs, historical production data, and completion data as inputs. One hypothesis is that all wells produce at a constant bottom hole pressure (BHP). Ghahfarokhi et al. (2018) used a multi-layer perceptron neural network to predict gas production using multi-point formation temperature monitoring in distributed temperature sensing and flowing time data. Khan and Louis (2021) used the bottom node and top node pressures of wells as ANN inputs to forecast shale gas production as the targeted output. To perform DCA with ANN and TSA, the input pressure data were also assumed to be constant throughout the manufacturing process. However, the ANN limit's shallow architecture and fully connected features act as impediments to scaling. Alimohammadi et al. (2020) used a deep neural network to develop a multivariate data-driven production forecasting process. As inputs, the pressure in the tubing head and the temperature in the bottom hole were chosen. Wu et al. (2020) used the least squares support vector machine to predict the development indicators in offshore oilfields. Ning et al. (2022) compared the performance of autoregressive integrated moving average model (ARIMA), long-short term memory (LSTM), and Prophet for oil and gas production prediction. Results show that the ARIMA is robust in predicting well production.

The primary objective of this work is to develop a data-driven model that will significantly accelerate the simulation of complex fracture networks in shale reservoirs. The multi-layer gate recurrent unit (GRU) cell with an attention mechanism and a skipconnection are used to solve the TSA problem of complex fracture networks wells. To be more precise, the mersenne twister (MT) generator is used to generate a stochastic initialization that conforms to the normal distribution and thus allows for the incorporation of physical model features. The training and testing data for the attention mechanism (Att-GRU) model are generated using the BEM. To avoid model overfitting, The cross-validation method is used to regularize sample data. The gating mechanism, attention mechanism, and skip connection could all be utilized to improve prediction accuracy and computational efficiency when using time series production data with the Att-GRU model. The results demonstrate that the Att-GRU model can faithfully recapitulate the variable BHP condition in all of its complexities when physical input is multivariate. The rate normalized pressure (RNP) analysis of the case study demonstrates that the proposed



**Fig. 1.** The schematic diagram of the architecture of multi-layer Att-GRU model.

Att-GRU is a viable proxy for forecasting production and evaluating reservoirs.

The novelty in this work is carry out the evaluation by considering complex fracture network and variable bottomhole pressure/production data. Since the complex fracture networks and variable bottomhole pressure/production data are not considered in decline curve analysis, the above mentioned studies can only be used for production forecasting and can not be used for reservoir evaluation in rate transient analysis.

## 2. Methodology

This section first clarifies the physical background and corresponding governing equations. Primary and secondary fractures, reactivated natural fractures, and matrix with stress sensitivity all occur in shales (Gale et al., 2014). Physical models from BEM are used to generate pressure and production data for horizontal wells with complex fracture networks in shales. The detailed derivation can be shown in Appendix. The Att-GRU principle is explained for the problem of complex fracture networks wells in shale reservoirs. Additionally, the initialization and preprocessing of data, the model structure, and the methodology workflow are discussed.

### 2.1 Multi-layer GRU with attention mechanism

Well production, along with various associated variables, is highly dependent on time series data over the long term. The impact of these variables on the production rate must consider their relative importance and the time period during which they affect the production rate. Due to the non-linear nature of the historical data, A new multi-layer Att-GRU is proposed for reservoir evaluation.

#### 2.1.1 Multilayer Att-GRU model

The proposed reservoir evaluation model aims to learn the non-linear mapping from the input matrix to the output vector:

$$\hat{y} = F(\mathbf{X}; \Psi) \quad (1)$$

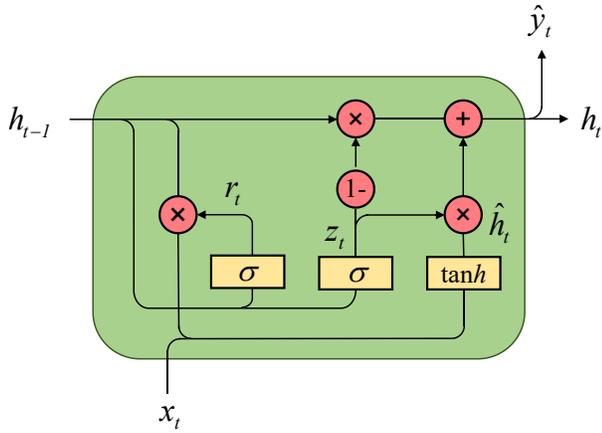
where  $F(\cdot)$  refers to the non-linear mapping function,  $\mathbf{X}^{(n+1) \times T}$  is the input matrix with the dimension of  $(n+1) \times T$ ,  $\Psi$  is the weight tensor determined by model learning process, and  $\hat{y}$  is the forecasted production rate vector. The input matrix in Eq. (2) is composed of the pressure time series vector  $p$  and the variable series data matrix  $\Theta$ :

$$\mathbf{X} = [p; \Theta] \in \mathbf{R}^{(n+1) \times T} \quad (2)$$

where  $p = (p_1, p_2, \dots, p_T) \in \mathbf{R}^T$  is the pressure vector in time series,  $\Theta = (\theta_1, \theta_2, \dots, \theta_T) \in \mathbf{R}^{n \times T}$  is the matrix of  $n$  variables involved in  $T$  data points generated by BEM.  $\theta = (\theta_1^1, \theta_1^2, \dots, \theta_1^n)^\top \in \mathbf{R}^n$  is the transposed vector of variables at time  $t$  with the dimension of  $n \times 1$  and it is assumed to be independent of time ( $\theta_1 = \theta_2, \dots, = \theta_T$ ).

As illustrated in Fig. 1, the Att-GRU model consists of two components: feature encoding and production prediction. According to Cho et al. (2014), the GRU cell is an extension type of the LSTM cell. The GRU has been used to address a variety of TSA issues, including classification and prediction tasks in time series, due to its simple model structure and superior performance. Its objective is to discover the long-term dependencies between time series data and to increase computational efficiency. The first GRU cell in the feature encoding section encodes the input variables at various points in time to create a new feature vector. In comparison to the conventional feature vector for the first GRU cell, the attention mechanism is used to generate a new input vector that incorporates historical and global information, allowing the model to adaptively select related vectors. In the reservoir evaluation section, the encoded new feature vectors and initial feature vectors are fed into the second GRU cell via a skip connection to perform cyclic reservoir evaluation. The skip connection contributes to the performance of the Att-GRU model by resolving the gradient vanishing issue (Mao et al., 2016).

Fig. 2 demonstrates that the hidden state vectors in the first GRU cell are used to encode the input time series vector  $x_t$



**Fig. 2.** The schematic diagram of the GRU cell (modified from Cho et al. (2014)). (The  $\times$  is the product sign, the  $+$  refers to the plus sign, the  $\sigma$  is the sigmoid function and the  $\tan h$  is the hyperbolic tangent function.)

at various time to form a new feature vector  $h_t$ . A GRU cell includes the reset gate  $r_t$  and update gate  $z_t$ . The sigmoid function  $\sigma$  is used as the activation function to produce probability estimates between 0 and 1. The reset gate  $r_t$  determines the combination degree of the new input vector  $x_t$  and hidden state output vector  $h_{t-1}$  at time  $t-1$ , as depicted in Eqs. (4) and (5). The  $z_t$  determines the vector  $h_{t-1}$  at time  $t-1$  are saved to the output vector  $h_t$  at time step  $t$ :

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (3)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (4)$$

$$n_t = \tan h(W_n \cdot [r_t \odot h_{t-1}, x_t]) \quad (5)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot n_t \quad (6)$$

where  $x_t$  is the input vectors at time  $t$ ,  $h_t$  is the hidden state output vectors,  $z_t$  and  $r_t$  are the output vectors of two sigmoid gates,  $W_n$ ,  $W_r$ , and  $W_z$  are the weight vectors determined by model learning process,  $n_t$  is the hidden state intermediate vector,  $\odot$  is the element-wise multiplication,  $\tan h$  refers to the hyperbolic tangent function, and the subscript  $t$  represents the  $t$ -th time step.

To enable the hidden state output vectors to adaptively select various variables in the input matrix  $X$ , the attention mechanism is introduced before the input data enters the first GRU cell. The attention mechanism makes the model automatically learn and calculate the contribution of input data to output data by weighting each scalar in the  $X$  (Hübner et al., 2010; Qin et al., 2017). The specified weight vector  $e_t$  can be determined in the learning process and it is designed as a combination of the  $X$  and the normalized vectors  $\tilde{h}_{t-1}$ , as given in Eq. (7). To measure the importance of input variables, the definition of the normalized weight vector is:

$$e_t = v_e^T \tan h(W_e \tilde{h}_{t-1} + U_e X) \quad (7)$$

$$\alpha_t = \frac{\exp(e_t)}{\sum_{i=1}^{n+1} \exp(e_t^i)} \quad (8)$$

where  $\tilde{h}_{t-1} \in \mathbf{R}^{m \times n+1}$  is the matrix after copying  $n+1$  times  $h_{t-1}$ ,  $m$  is the feature number in the hidden state output vector,  $n$  refers to the number of input variables,  $h_{t-1} \in \mathbf{R}^m$  is the  $m$ -dimensional hidden state output vectors at time  $t-1$ ,  $v_e \in \mathbf{R}^T$ ,  $W_e \in \mathbf{R}^{T \times m}$ , and  $U_e \in \mathbf{R}^{T \times T}$  are the  $T$ ,  $T \times m$ ,  $T \times T$ -dimensional weight matrixes determined by model learning. With the definition of  $\alpha_t$ , the adaptive extraction process of input variables can be expressed as:

$$\tilde{x}_t = \alpha_t \odot x_t \quad (9)$$

The new feature vector  $\tilde{x}_t$  can be used as the input vector of the first GRU cell.  $h_t$  will be used as a new feature vector for reservoir evaluation:

$$h_t = f_1(h_t, \tilde{x}_t) \quad (10)$$

where  $f_1$  refers to the non-linear mapping function in the first GRU cell. In the reservoir evaluation part, the second GRU cell is used to predict the production rate at all-time steps. The input vector  $x_t$  and the hidden state output vectors  $h_t$  of the first GRU cell are connected via skip connection, which is used as the input vector of the second GRU cell (Eq. (11)). The skip connection provides the short path from the bottom to the top layers (Tong et al., 2017):

$$d_t = f_2(d_{t-1}, [h_t; x_t]) \quad (11)$$

where  $f_2$  is the non-linear mapping function in the second GRU cell,  $d_{t-1} \in \mathbf{R}^m$  and  $d_t \in \mathbf{R}^m$  are the  $m$ -dimensional hidden state output vectors of the second GRU cell at time  $t-1$  and  $t$ . Further, the weight matrix  $W_d$  with the dimension of  $T \times m$  and the weight vector  $b_w$  with the dimension of  $T \times 1$  are used to non-linearly transform the  $d_t$  with the dimension of  $M \times 1$  into a vector with the dimension  $T \times 1$ . The weight vector  $v_y^T$  with the dimension of  $T \times 1$  and the weight  $b_v$  are used to linearly transform the  $T \times 1$ -dimensional vector to finally obtain the predicted  $\hat{y}_t$  at time  $t$ :

$$\hat{y}_t = v_y^T \text{relu}(W_d d_t + b_w) + b_v \quad (12)$$

where  $v_y \in \mathbf{R}^T$ ,  $W_d \in \mathbf{R}^{T \times m}$ ,  $b_w \in \mathbf{R}^T$ , and  $b_v \in \mathbf{R}^1$  are weight vectors determined by Att-GRU training,  $d_t \in \mathbf{R}^m$  refers to the hidden state output vectors of the second GRU cell at time  $t$ ,  $\text{relu}$  is the activation function, and  $\hat{y}_t$  is the normalized predicted rate at time  $t$ .

### 2.1.2 Neural network training

The mean squared error (MSE) between the predicted output value and the actual value is optimized as the loss function:

$$L(y, \hat{y}) = \frac{1}{N} \sum_{i=1}^N \sum_{t=1}^T (\hat{y}_t^i - y_t^i)^2 \quad (13)$$

where  $N$  is the sample number,  $T$  refers to the number of input data points,  $\hat{y}_t^i$  and  $y_t^i$  are the  $i$ -th sample's predicted and actual production rates at time  $t$ ,  $\hat{y}$  and  $y$  are the predicted and actual production rate vectors for  $N$  samples. For the weight tensor,

the Adam optimizer in Eq. (14) is used to update the model weight value:

$$\begin{cases} w \leftarrow w - \frac{\eta}{\sqrt{v} + \varepsilon} u \\ u \leftarrow \frac{\beta_1 u + (1 - \beta_1)g}{1 - \beta_1} \\ v \leftarrow \frac{\beta_2 v + (1 - \beta_1)g^2}{1 - \beta_1^2} \end{cases} \quad (14)$$

where  $\eta$  is the learning rate,  $w$  and  $g$  are the weight tensor and tensor gradient determined by model training,  $u$  and  $v$  are the intermediate variables. In this work, the parameter combination recommended by Kingma and Ba (2015) is used where  $\beta_1$ ,  $\beta_2$  and  $\varepsilon$  are models parameters, and  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\varepsilon = 10^{-8}$ . Additionally, the technique of minibatch gradient descent is used to reduce the memory requirement for Att-GRU training. The minibatch size and initial learning rate are both set to 64 and 0.02. At 1,000, 1,500, 2,000, and 2,500 epochs, the learning rate is reduced to half. To evaluate the performance of the Att-GRU model, the mean absolute error (MAE) and mean absolute percentage error (MAPE) are used. The MAE directly measures the difference between the predicted and actual production rates:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y^i - \hat{y}^i| \quad (15)$$

where  $\hat{y}^i$  and  $y^i$  are the predicted output vector and actual output vector for the  $i$ -th sample. The MAPE is a measure of the method prediction accuracy in statistics. It shows a relative value:

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{y^i - \hat{y}^i}{y^i} \right| \quad (16)$$

## 2.2 Pressure and rate calculation with time series analysis

The BEM demonstrates that the relationship between production performance and operating and physical parameters is complex and non-linear. The variable BHP condition contains extensive field data. As a result, the pressure and rate data include cases of variable pressure drop/variable flowrate in addition to the constant BHP condition.

### 2.2.1 Physical input parameters initialization

Given the offshore environment, unconsolidated sandstone and heavy oil, optimizing horizontal well designs has become an increasingly difficult task. The horizontal well length is dependent on the rate of production, cumulative production, and economic factors (i.e., net present value (NPV) and net present value ratio (NPVR)). According to the analysis of production performance, NPV, and NPVR, the well length for horizontal wells is between 0 and 7,000 feet (Cho, 2001). The optimal length of a horizontal well varies between 1,000 and 2,000 feet (Vicente et al., 2003). With a shale gas price of USD 5/Mcf, the fracture number and half-length ranges are 5-33 and 50-250 m, respectively (Yao et al., 2021). The factors affecting hydraulic fracture conductivity are more complex,

and include effective closure stress, proppant type and size, non-Darcy, and multiphase flow effects (Fredd et al., 2000). According to previous research (Fredd et al., 2000; Yao et al., 2021), the fracture conductivity distribution range is typically 0.1-6,000 mD.ft. The normal distribution assumption is one of the most commonly used probability distribution assumptions. According to the Central Limit Theorem (Kwak and Kim, 2017), when experiments are repeated many times on a large number of random variables, their distributions will be very close to the normal distribution. The MT generator (Matsumoto and Nishimura, 1998) generates random numbers in the distribution ranges depicted in Fig. 3 which shows a normal probability plot. It also establishes that the initial distribution of input parameters is normal.

### 2.2.2 Bottom hole pressure initialization

The variation range of BHP is chosen to be 1-16 MPa. The cases of constant BHP are selected as the original templates in the training samples. The constant BHP cases are extended to the variable BHP cases, which can be roughly classified into four types: (a) constant BHP, (b) variable BHP in stepwise variation, (c) variable BHP in exponential decline and (d) variable BHP in a combination of stepwise and exponential decline, as illustrated in Fig. 4. The BHP curve can be divided into ten stages using stepwise variation. Each stage's duration is consistent with the dirichlet distribution, and the total time required is 10,000 h. Stepwise variation of the BHP value results in a random deviation from the original value plus or minus 1 MPa, which also fits the normal distribution. In particular, for the exponential decline type, the BHP curve meets the mathematical formula of  $y = ae - bx + c + d$ .

To avoid the gradient explosion/vanishing issues in the training process, it is necessary to normalize the input and output data. The input data includes BHP, production time, well length, fracture half-length, fracture conductivity, and fracture number as the input data can be normalized as:

$$\tilde{t} = \frac{t - t_{\min}}{t_{\max} - t_{\min}} \quad (17)$$

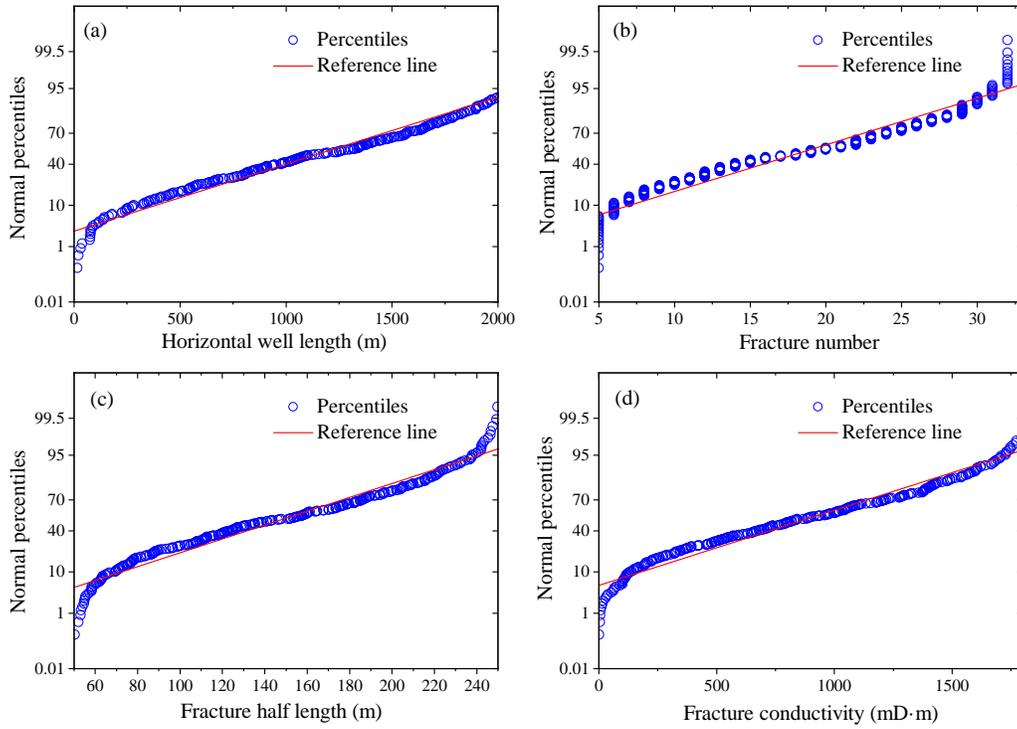
where  $t$  is the original time vector,  $\tilde{t}$  is the normalized time vector,  $t_{\max}$  and  $t_{\min}$  are the maximum and minimum values for the production time. A min-max scaler is selected to normalize the BHP data:

$$\tilde{P} = \frac{P - P_{\min}}{P_{\max} - P_{\min}} \quad (18)$$

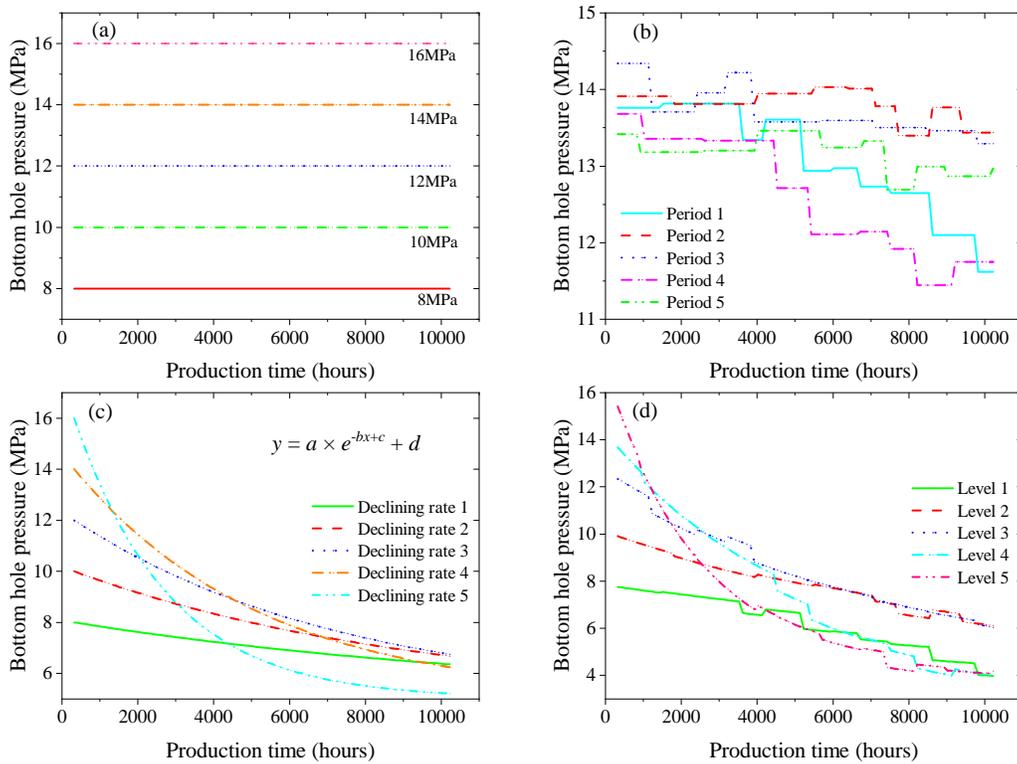
where  $\tilde{P}$  is the normalized pressure vector,  $P$  refers to the original pressure vector,  $P_{\max}$  and  $P_{\min}$  are the maximum and minimum pressures values. The remaining physical input parameters, including fracture half-length, fracture conductivity, fracture number, and well length, also need to be normalized before being imported into the Att-GRU model:

$$\tilde{\theta}_i = \frac{\theta_i - \theta_{i,\min}}{\theta_{i,\max} - \theta_{i,\min}} \quad (19)$$

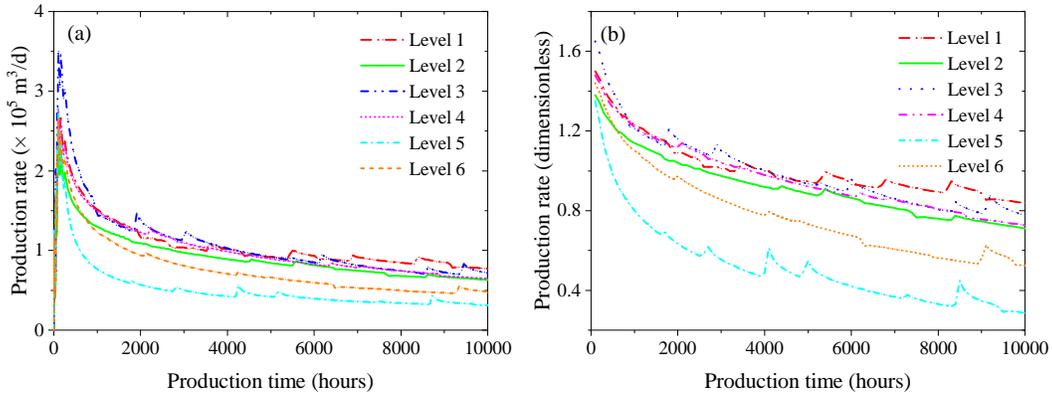
where  $\tilde{\theta}_i$  is the  $i$ -th normalized parameter vector,  $\theta_i$  is the  $i$ -th original parameter vector,  $\theta_{i,\min}$  and  $\theta_{i,\max}$  are the maximum and minimum parameter values.



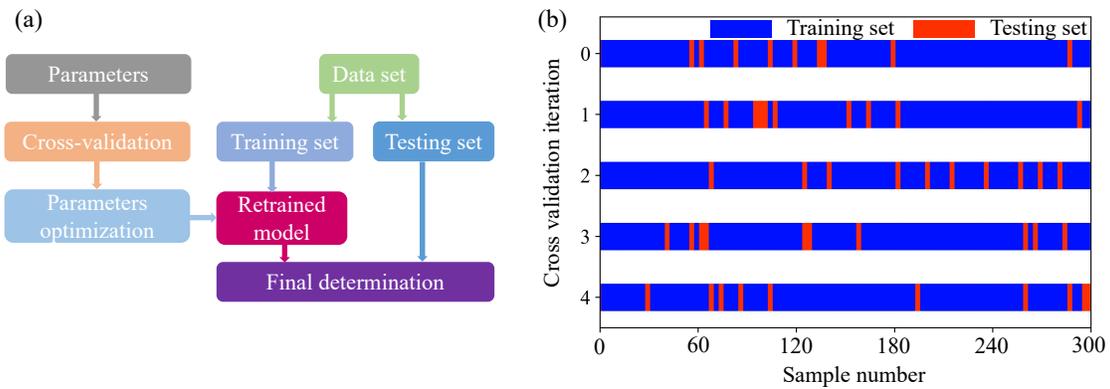
**Fig. 3.** The normal probability plot of input parameters. (a) Horizontal well length, (b) fracture number, (c) fracture half-length and (d) fracture conductivity.



**Fig. 4.** Input BHP plots of shale gas wells. (a) Constant BHP, (b) variable BHP in stepwise variation, (c) variable BHP in exponential decline and (d) variable BHP in the combination of stepwise and exponential decline.



**Fig. 5.** Representative production rate plots of shale gas wells with complex fracture networks. (a) Before normalization and (b) after normalization.



**Fig. 6.** Schematic diagram of the shuffled cross-validation method. (a) Cross-validation method flowchart and (b) training and testing sets distribution in the cross-validation method.

### 2.2.3 Calculation and preprocessing of production rate

The production rate in constant and variable BHP condition calculated by BEM is used as the output data. For the output production data, the logarithmic normalization in Eq. (20) is used to process the output data (Fig. 5):

$$\tilde{y} = \frac{\log Q - \log Q_{\min}}{\log Q_{\max} - \log Q_{\min}} \quad (20)$$

where  $\tilde{y}$  is the normalized production rate vector as the Att-GRU target vector,  $Q$  is the original production rate vector,  $Q_{\min}$  and  $Q_{\max}$  are the maximum and minimum values for the production rate. For the Att-GRU predicted production rate vector, the normalized production rate vector is scaled back to the original range as:

$$Q = \exp[\tilde{y} \times (\log Q_{\max} - \log Q_{\min}) + \log Q_{\min}] \quad (21)$$

## 3. Results and discussion

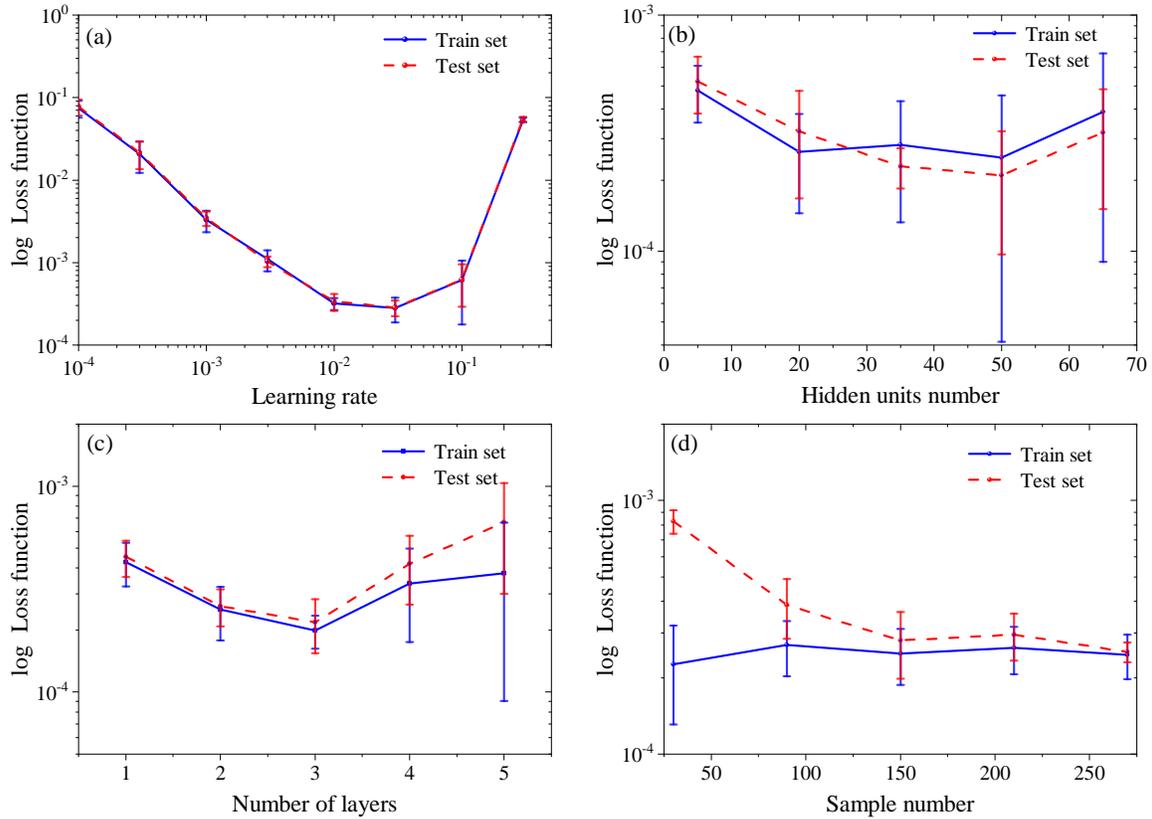
In this part, the physical data in BEM are used to train and test the Att-GRU model. The training procedure, testing results, and parameters analysis are also clarified. The computational burden of the data-driven Att-GRU is compared with the physical model.

## 3.1 Training procedure

### 3.1.1 Model optimization

As illustrated in Fig. 6(a), shuffled cross-validation is used to assess model robustness, avoid model overfitting, and optimize hyperparameters. The learning rate, the number of hidden units, the number of GRU cells, and the sample number are all set to 0.01, 20, 2, and 300 in the initialized model. The shuffled cross-validation technique is used to randomly shuffle and divide the data into equal-sized  $k$  folds (see Fig. 6). Additionally, as illustrated in Fig. 6(b), the sample sizes for the training and testing sets account for approximately 80% and 20% of the total sample size, respectively. Four hyperparameters are optimized separately and the model's stability is evaluated using 5-fold cross-validations. The final Att-GRU model is trained over 4,000 epochs.

The log loss function and error bars on the training and testing sets are selected as the benchmark for model optimization. Fig. 7 and Table 1 show that the log loss functions of the 5-fold cross-validation events on the training and testing sets are similar and eliminated the overfitting issue. The log loss function and error bars analysis demonstrates that the Att-GRU has the smallest log loss function and the highest robustness when the learning rate, the number of hidden layer units, the



**Fig. 7.** Model optimization results using shuffled cross-validation. (a) Learning rate, (b) the number of hidden layer units, (c) the layer number and (d) samples number.

**Table 1.** The optimized parameters and values in model optimization.

Optimized parameters	Interval	Value
Learning rate	$10^{-4}, 3 \times 10^{-4}, 10^{-3}, 3 \times 10^{-3}$ $10^{-2}, 3 \times 10^{-2}, 1 \times 10^{-1}, 3 \times 10^{-1}$	$10^{-2}$
Number of hidden layer units	5, 20, 35, 50, 65	50
Layer number	1, 2, 3, 4, 5	3
Samples number	30, 90, 150, 210, 270	270

number of layers, and the number of samples are set to 0.01, 50, 3, and 270, respectively.

### 3.1.2 Model comparison

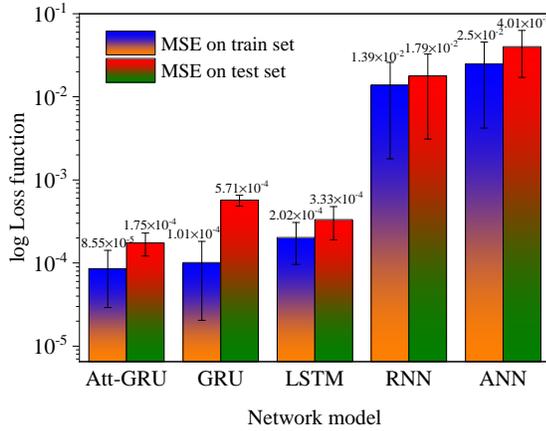
At various points in time, the ANN makes use of independent data vectors, and there is no concept of memory to handle memory tasks. The neurons in a recurrent neural network (RNN) employ a simple feedback-type method in which the previous time's state variables and input data combine to form the current time's input data. One of the reasons for RNN's limited use is the inherent gradient vanishing problem. The GRU can retain memory in the gating mechanism, thereby resolving the issue of long-term dependencies without causing the gradient to vanish. The attention mechanism is used to extract features from the input data based on the multi-layer GRU, allowing the model to adaptively judge the importance

of various variables at different time points. Since the input data is time-dependent, the model requires a large amount of input data. However, at a certain moment, only a small part to improve model performance by preventing gradient vanishing. The Att-GRU model achieves the smallest log loss function MSE and error bar value on the training and testing sets, as illustrated in Fig. 8. In comparison to the LSTM model, the addition of an attention mechanism and a skip connection reduces the MSE and MAPE on the testing set by 47% and 28%, respectively, as shown in Table 2.

The MAPE and error are calculated on the testing data set. Fig. 9 shows that the MAPE of ANN and RNN fluctuates significantly in the entire of training process. As the training progresses, the errors of ANN and RNN are not stable. Compared with ANN and RNN, the MAPE and error of LSTM and GRU can maintain a relatively constant value during the

**Table 2.** Comparison results of MSE and MAPE on various model testing sets.

Methods metric	Att-GRU	GRU	LSTM	RNN	ANN
MSE	$1.75 \times 10^{-4}$	$5.71 \times 10^{-4}$	$3.33 \times 10^{-4}$	$1.79 \times 10^{-2}$	$4.01 \times 10^{-2}$
MAPE	$6.9 \times 10^{-3}$	$1.7 \times 10^{-2}$	$9.6 \times 10^{-3}$	$4.53 \times 10^{-2}$	$8.75 \times 10^{-2}$



**Fig. 8.** Models comparison results of log loss function MSE and error bar during training process.

training process. Overall, the MAPE and error stability are better in the Att-GRU training, and their values are close to 0.

### 3.2 Testing results

The sample data that are not included in the training set are used to verify and test the proposed Att-GRU model’s generalization ability and accuracy. During the Att-GRU model training process, the log loss functions on the training and testing sets are compared. As illustrated in Fig. 10, the log loss function’s initial value is set to 100. The log loss functions exhibit an exponentially decreasing trend as the training epoch length increases. As the log loss function value decreases, the minute fluctuations in the function become more significant. There are a total of 4,000 training epochs. Additionally, Fig. 10 demonstrates that as training epochs increase to 1,000, the log loss functions on the training and testing sets tend to 10<sup>-4</sup>. The five random initializations of the weight tensor in the Att-GRU model demonstrate that the initialization of the weight tensor does not affect the model’s robustness.

#### 3.2.1 Experimental design

The experimental design excludes data on shale gas well production history from the training and testing data sets. The experimental design case consists of 22 primary fractures and 132 secondary fractures, all of which have finite fracture conductivity, and a horizontal well length of 1,462 m. Table 3 contains the remaining parameters. As illustrated in Fig. 11, the shale gas well produces at a stepwise decreasing BHP rate, and the production history can be roughly divided into nine stages. The normalized BHP and production rate upper and lower limits are 1.0-1.3 and 0.8-1.7, respectively. The production data from shale gas wells in BEM are compared to

the data predicted by the Att-GRU model. The data in BEM and the predicted results in Fig. 11 agree reasonably well over the entire production history, indicating that the proposed Att-GRU model has learned the physical model’s underlying physics. The greatest discrepancy between actual and predicted production data occurs in the variable BHP stage.

#### 3.2.2 Error analysis

As illustrated in Fig. 12, the MAPE and error curves used in the experimental design during the training process are all distributed within the distribution interval of the testing set and have values close to 0. Increases in MAPE and error curves correspond to changes in the BHP of production wells. The Att-GRU model has a harder time learning the physical superposition effect introduced by the variable BHP condition.

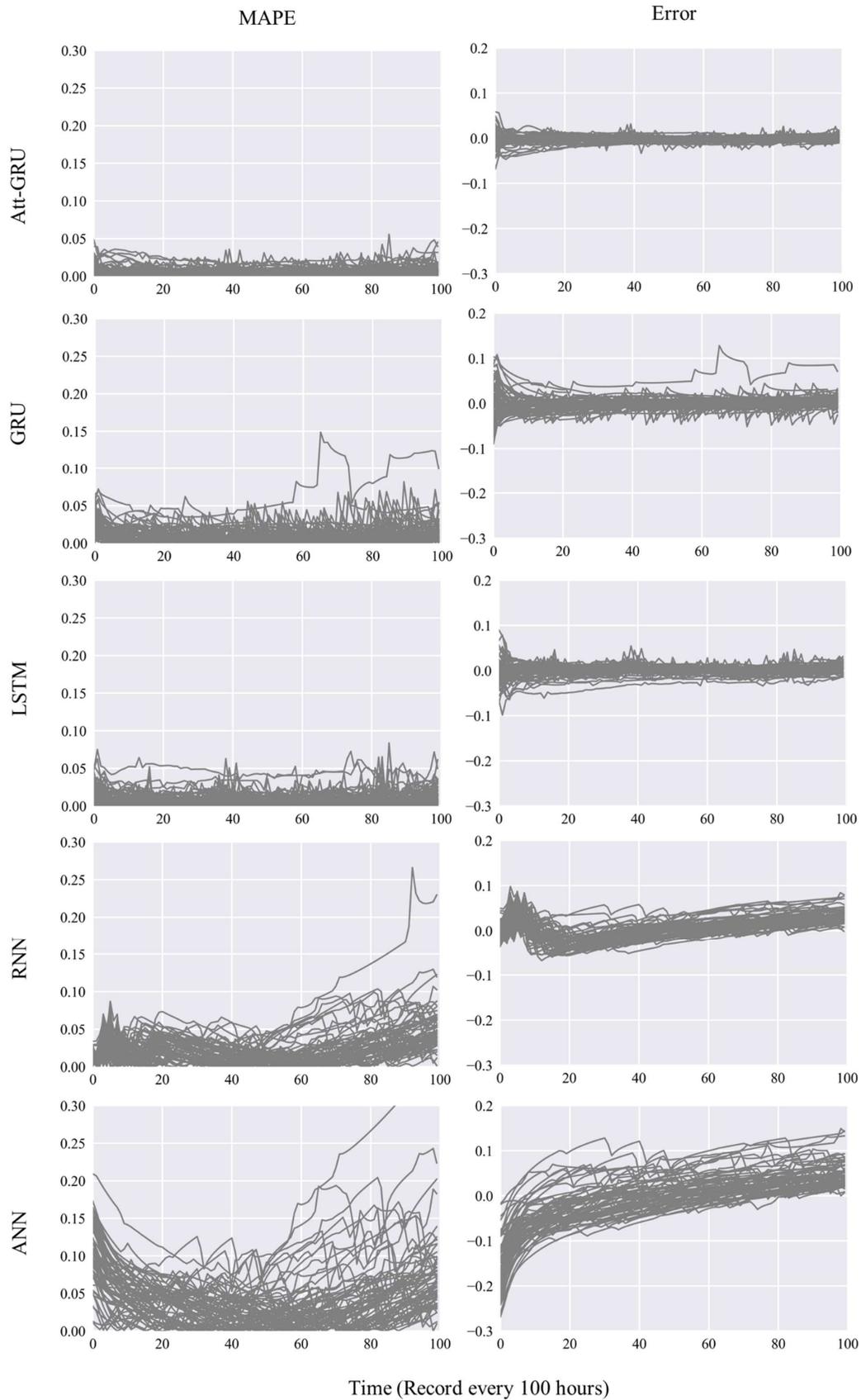
### 3.3 Operating and physical parameters analysis

#### 3.3.1 Variable pressure drop/flowrate analysis

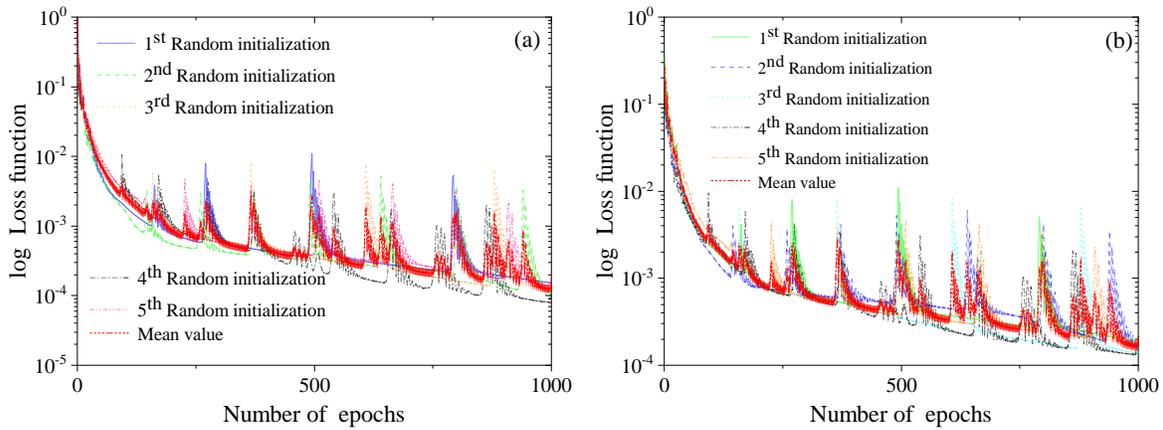
The majority of shales production wells operate under constant and variable BHP conditions. The constant BHP cases are used as the base cases for the variable pressure drop/flow rate analysis. Due to the ultra-low permeability of shale, the depletion strategy is the most likely to be used in the recovery of shale reservoirs. As illustrated in Fig. 13, pressure depletion during recovery has resulted in the decline in the production of shale gas wells with complex fracture networks. To maintain a constant rate of production and to account for field well management strategies, the base cases of constant BHP have been expanded to include various declining BHP forms. The Att-GRU model is used to determine production performance under constant and variable BHP conditions, as illustrated in Fig. 13. For shale gas wells, the remaining parameters are kept constant.

#### 3.3.2 Physical parameters sensitivity analysis

To demonstrate that the proposed Att-GRU successfully captured and learned the physical model’s features, a sensitivity analysis of fracture network parameters is performed. The variation ranges for the half-length and number of primary fractures are chosen to be 80-145 m and 10-22 m, respectively. The shale gas well’s horizontal length is increased from 750 to 1,450 m. The BHP is assumed to be the variable BHP in the combination of stepwise and exponential decline in the parameters sensitivity analysis section, see Fig. 14(a). As shown in Fig. 14(b), the half-length of fractures is positively correlated with the rate and cumulative production of wells. As expected, the rate of production and cumulative output increase as the number of fractures increases (see Fig. 14(c)). Increases in horizontal well length result in an increase in drainage area,



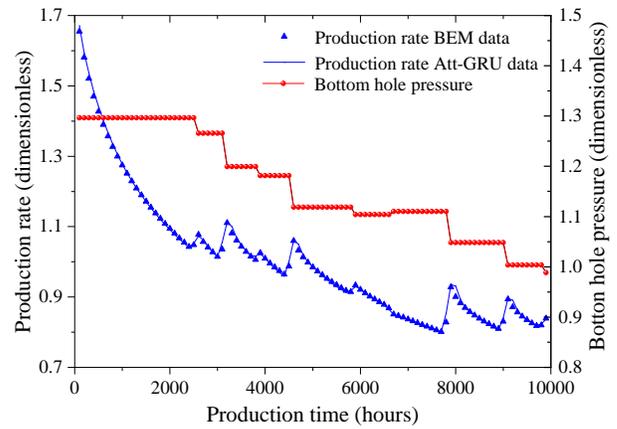
**Fig. 9.** MAPE and error comparison results of the testing data set at various time points in the training process.



**Fig. 10.** The log loss function on the training and testing sets for five random initializations of weight tensor. (a) Training set and (b) testing set.

**Table 3.** The physical parameters of shale gas well with complex fracture networks in experimental design.

Parameter	Experimental case
Well length (m)	$1.462 \times 10^3$
Well radius (m)	$9 \times 10^{-2}$
Skin factor	$8 \times 10^{-3}$
Fracture number	22
Fracture half-length (m)	$1.423 \times 10^2$
Fracture conductivity (mD·m)	$3.383 \times 10^2$
Fracture number	$1.32 \times 10^2$
Fracture half-length (m)	21.1
Fracture conductivity (mD·m)	$2 \times 10^2$
Fracture width (m)	$1 \times 10^{-4}$
Natural fracture permeability (mD)	$3.5 \times 10^{-3}$
Initial pressure (MPa)	23.35
Thickness (m)	32.43
Porosity (%)	6.9
Rock compressibility ( $\text{MPa}^{-1}$ )	$8.27 \times 10^{-4}$
Reservoir temperature ( $^{\circ}\text{C}$ )	48.89
Matrix unit radius (m)	$1.8 \times 10^{-9}$
Gas viscosity (cP)	$2.3 \times 10^{-2}$
Gas density ( $\text{kg}/\text{m}^3$ )	$1.906 \times 10^2$
Gas compressibility ( $\text{MPa}^{-1}$ )	$2.8 \times 10^{-2}$
Z-factor	0.94
Langmuir volume ( $\text{m}^3/\text{ton}$ )	2.4
Langmuir pressure (MPa)	3.23
Diffusion coefficient ( $\text{m}^2/\text{s}$ )	$1 \times 10^{-7}$



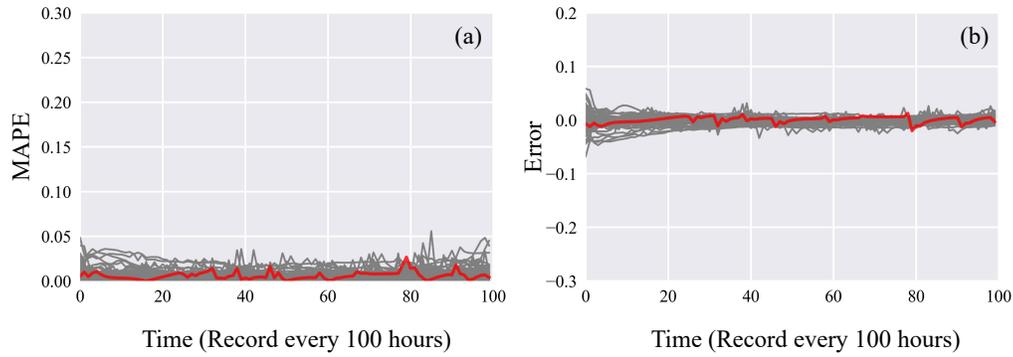
**Fig. 11.** The variable BHP data and production rate data of shale gas well with complex fracture networks in experiment design.

demonstrates that the proposed Att-GRU model has acquired knowledge of the underlying physical properties of complex fracture networks as well as the variable pressure drop/flowrate characteristics observed in field conditions.

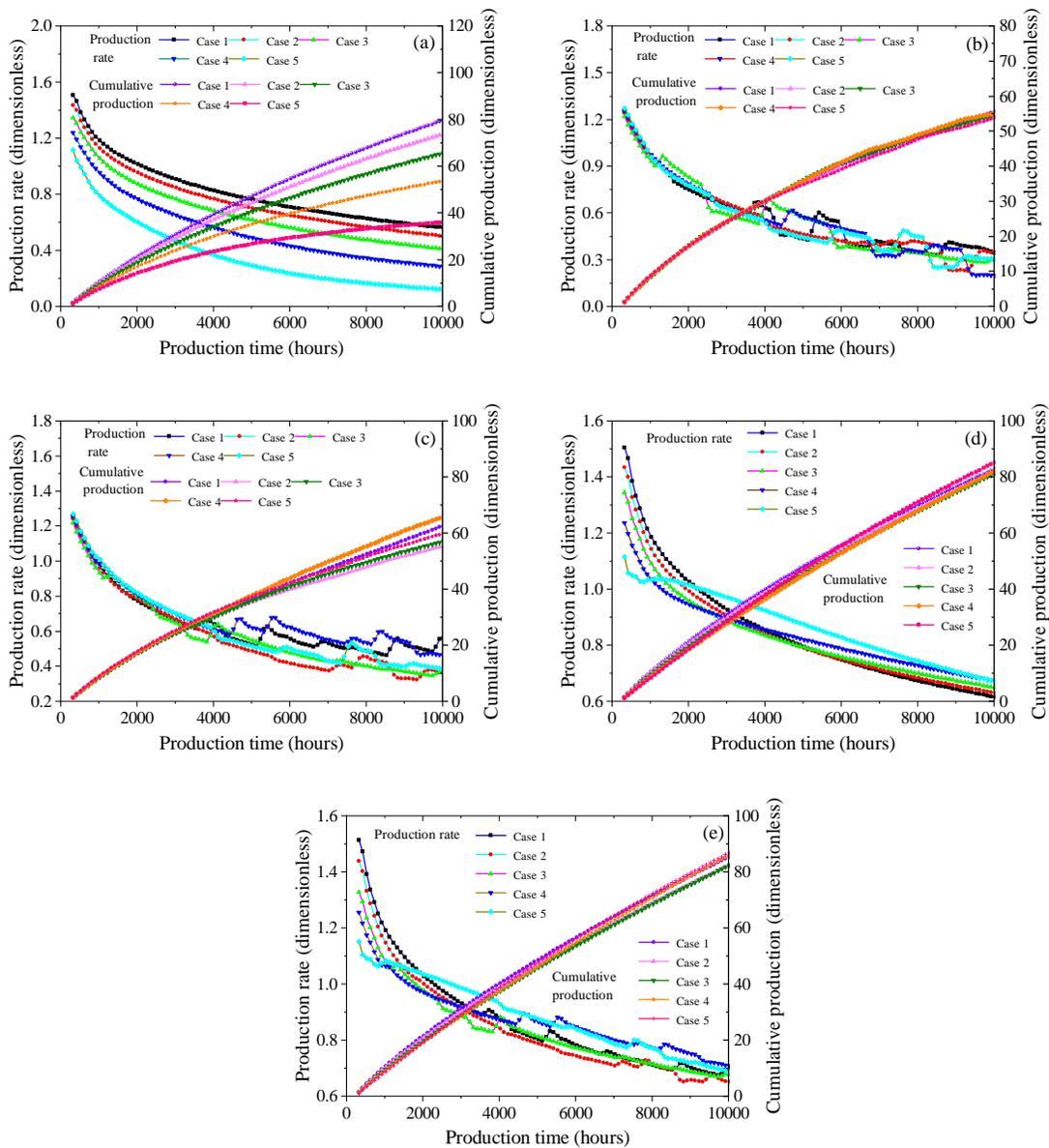
### 3.4 Computation burden

The programs are written by pytorch and it is an open source Python machine learning library. The total computational time of the data-driven Att-GRU model and the physical model are compared using an NVIDIA Quadro P4000 graphics processing unit. The physical model in BEM is a sophisticated numerical model that requires discretization of only the points or segments defining the inner and outer boundaries, rather than the entire area. The performance of BEM calculations is highly dependent on the coefficient matrix and numerical integration. According to Fig. 15, it takes an average of 13 minutes for the physical model to complete a reservoir evaluation task for a type of shale gas well with complex fracture networks. The total time cost increases linearly as the number of running physical models increases. According to the sample size in Fig. 7(d), the Att-GRU model requires at

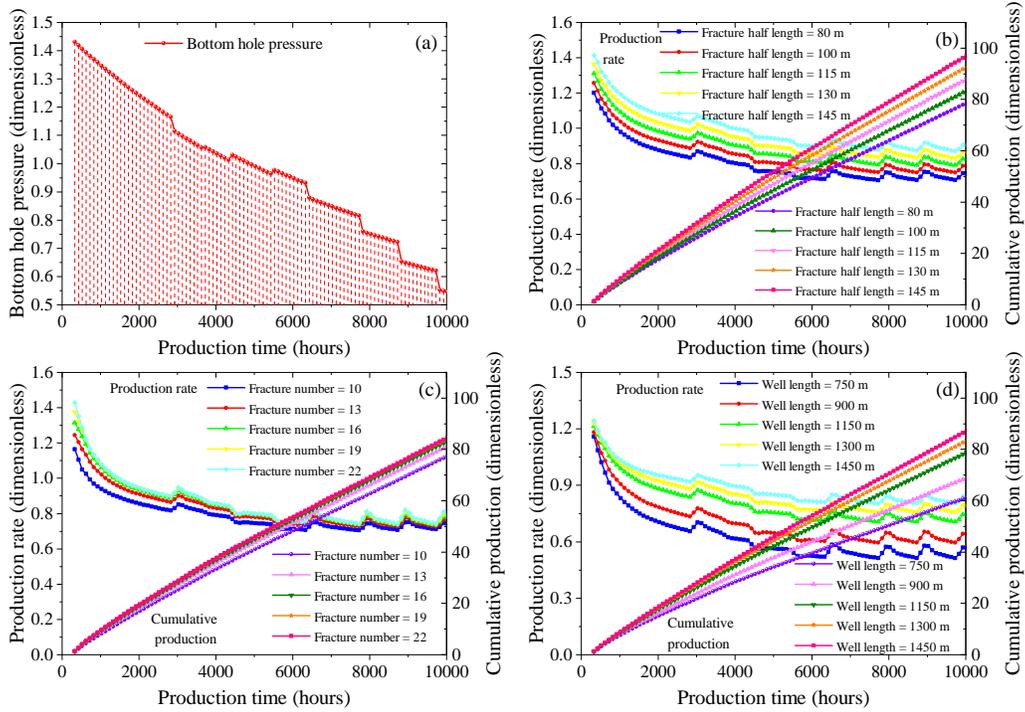
which results in an increase in production rate and cumulative production, as illustrated in Fig. 14(d). Sensitivity analysis



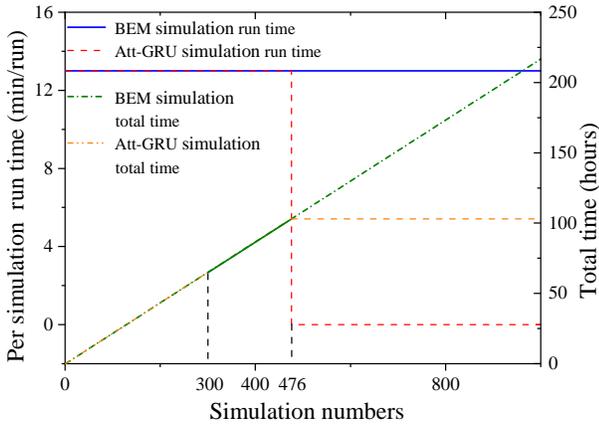
**Fig. 12.** The MAPE and error curves at various time points of the experiment design and testing data set. (a) MAPE curves and (b) error curves.



**Fig. 13.** Production rate and cumulative production plots in various BHP conditions. (a) Constant BHP, (b) variable BHP in stepwise variation, (c) variable BHP in stepwise decline, (d) variable BHP in exponential decline and (e) variable BHP in the combination of stepwise and exponential decline.



**Fig. 14.** Comparison of production rate and cumulative production in various physical parameters. (a) Input BHP data, (b) fracture half-length, (c) fracture number and (d) well length.



**Fig. 15.** The computation burden comparison of data-driven Att-GRU model and the physical model in BEM.

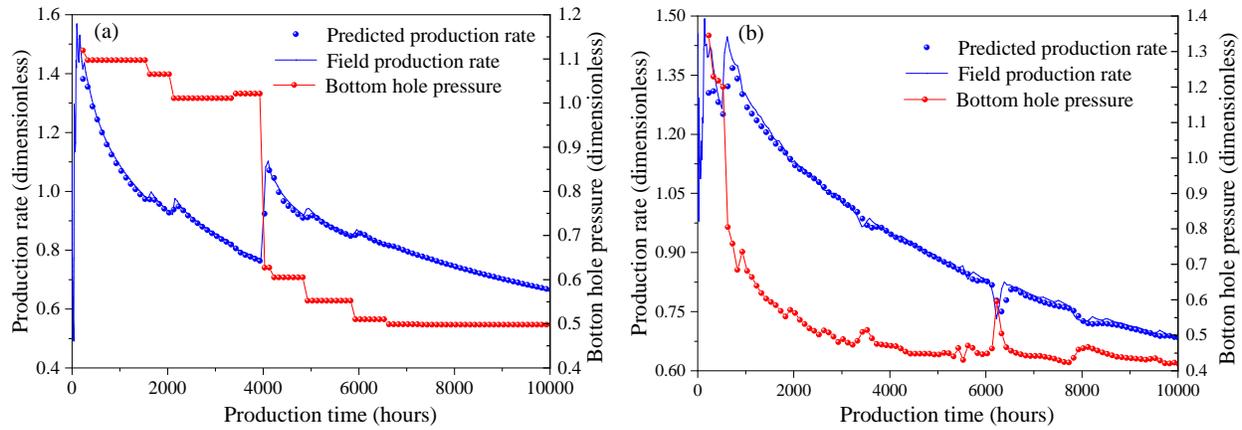
least 300 runs of the physical model, which takes 65 h in total. The Att-GRU model is trained using 5-fold cross-validations, with each cross-validation taking 20 minutes. The total calculation time is 38 h, which corresponds to 176 physical model runs. Once the model training process is complete, the Att-GRU model prediction process is approximately 0.3 s, which equates to 3.8% of the physical model’s running time. Meanwhile, if the physical model is run more than 476 times in a project (including type curve matching in reservoir evaluation), the Att-GRU model’s calculation cost will be lower than that of the physical model (see Fig. 15 and Table 2).

**Table 4.** The well, fracture, reservoir, and shale gas properties in the case study.

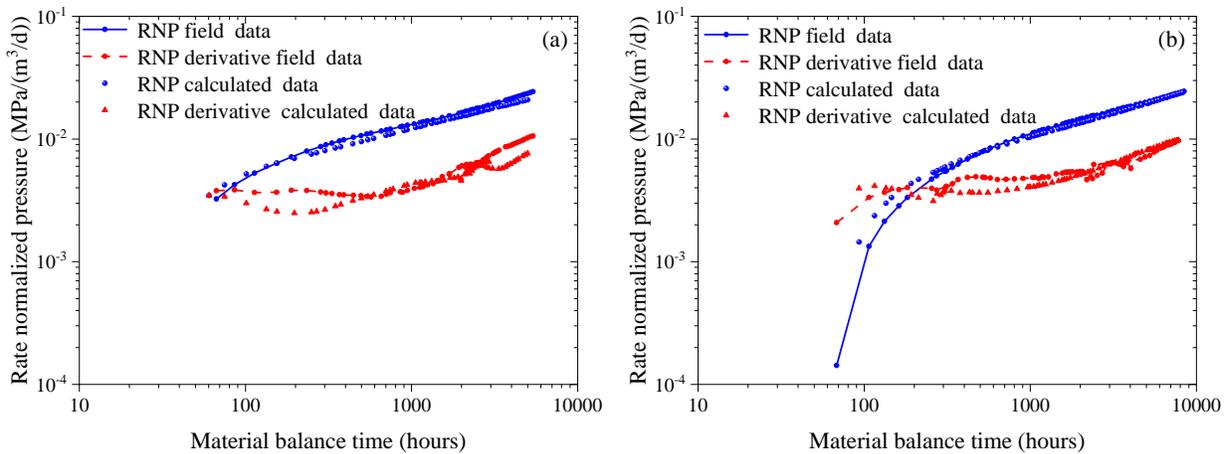
Parameter	Case 1	Case 2
Well length (m)	$1.191 \times 10^3$	$1.272 \times 10^3$
Skin factor	$8 \times 10^{-3}$	$1 \times 10^{-2}$
Fracture number	12	7
Fracture half-length (m)	80	91
Fracture conductivity (mD·m)	$1.023 \times 10^3$	$4.572 \times 10^2$
Fracture number	72	42
Fracture half-length (m)	24.384	22
Fracture conductivity (mD·m)	$2 \times 10^2$	$1.8 \times 10^2$
Natural fracture permeability (mD)	$3.51 \times 10^{-3}$	$3.38 \times 10^{-3}$

#### 4. Case study

To further demonstrate the proposed methodology’s applicability and generalizability, two shale gas wells’ recorded history data from published literature are used as a case study (Thompson et al., 2011). Two shale gas wells are targeting the Marcellus shale and are all horizontal wells with hydraulic fractures. Thompson et al. (2011) matched the history data using the multi-fractured horizontal wells model. The fundamental properties of the reservoir and shale gas are summarized in Table 4. Throughout the long history of production, Fig. 16 indicates that two wells have been classified as variable BHP. The variable BHP data and initial physical



**Fig. 16.** Comparison of production data predicted by the Att-GRU model and the field production data for two shale gas horizontal wells with complex fracture networks. (a) Case 1 and (b) case 2.



**Fig. 17.** Comparison of RNP data predicted by the Att-GRU model and the field RNP data for two shale gas horizontal wells with complex fracture networks. (a) Case 1 and (b) case 2.

parameters are used as the Att-GRU model's input parameters. The production data calculated by the Att-GRU model and the reported field data are plotted in Fig. 16 for comparison using type curve matching. The good agreement between the calculated and field data demonstrates the applicability of the proposed methodology, which can be used to forecast the production of shale gas wells under realistic variable pressure drop/flow rate conditions.

The proposed methodology is capable of performing reservoir evaluations and forecasting the performance of complex fracture network shale gas wells. The predicted output value is used to match the field data by adjusting the Att-GRU model's input parameters. Rate transient analysis can be performed by comparing the RNP data generated by the Att-GRU to the RNP data recorded in historical data using RNP type curve analysis. Fig. 17 demonstrates that an acceptable RNP curve match between predicted and field data can be obtained. Meanwhile, some errors remain in the RNP derivative curve, particularly in the case of rapidly changing BHP values.

## 5. Conclusions

A practical proxy model for shale gas wells with complex fracture networks is developed using the data-driven Att-GRU model and the physical model in BEM. The Att-GRU model can extract features from input data via the attention mechanism, allowing the model to adaptively judge the importance of various variables at various points in time. The Att-GRU model can memorize significant historical data selectively using the gating mechanism. Additionally, the attention mechanism and skip connection enhance the model's prediction performance while preserving its computational efficiency. The Att-GRU model is trained to forecast well performance under both constant and variable BHP conditions. Regularization of the input and output data facilitates model convergence during training. To evaluate the performance of the Att-GRU model, 20% of the sample data is used as the testing set, and the shuffled cross-validation method is used to avoid overfitting. The results indicate that the proposed model is capable of forecasting the production of shale gas wells at various time points, BHP values, and physical parameters. Field cases

further verify the Att-GRU model's practicability in reservoir evaluation. The detailed conclusions are as follows:

1) With the gating mechanism in GRU cell, attention mechanism, and skip connection, the Att-GRU model has the strong ability to deal with TSA issues of multivariate inputs including time, BHP, and physical parameters.

2) The results of the shuffled cross-validation and blind test indicate that the Att-GRU model performed well on both the training and testing sets.

3) Case analysis shows that the Att-GRU can meet the requirements of reservoir evaluation.

4) For early data and variable BHP conditions, sample data constrain the Att-GRU model's capability and place greater demands on sample data, particularly for the early period.

5) This deep learning method adds a new dimension to proxy models, accelerating the simulation of complex fracture networks.

Although the model prediction time is only 0.3 s. The model training time is 38 h. The decrease of model training time will be one of our next research focus.

## Acknowledgements

We are very grateful to all the people who have contributed to this work. This work received funding support from Fundamental Research Funds for the Central Universities (No. QNXM20220011), China Postdoctoral Science Foundations (No. 2021M700391), and National Science Foundation, China (No. 12202042).

## Conflict of interest

The authors declare no competing interest.

**Open Access** This article is distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC-ND) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## References

- Alimohammadi, H., Rahmanifard, H., Chen, N. Multivariate time series modelling approach for production forecasting in unconventional resources. Paper SPE 201571 Presented at SPE Annual Technical Conference and Exhibition, Virtual, 26-29 October, 2020.
- Aminian, K., Ameri, S. Predicting production performance of CBM reservoirs. *Journal of Natural Gas Science and Engineering*, 2009, 1(1-2): 25-30.
- Carlson, E. S., Mercer, J. C. Devonian shale gas production: Mechanisms and simple models. *Journal of Petroleum Technology*, 1991, 43(4): 476-482.
- Chao, G., Lee, W. J., Spivey, J. P., et al. Modeling multilayer gas reservoirs including sorption effects. Paper SPE 29173 Presented at SPE Eastern Regional Meeting, Charleston, West Virginia, 8-10 November, 1994.
- Cho, H. Integrated optimization on long horizontal well length. Paper SPE 68599 Presented at SPE Hydrocarbon Economics and Evaluation Symposium, Dallas, Texas, 2-3 April, 2001.
- Cho, K., Van Merriënboer, B., Gulcehre, C., et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation. Paper Presented at Conference on Empirical Methods in Natural Language Processing, Doha, Qatar, 25-29 October, 2014.
- Choi, J., Roberts, D. C., Lee, E. S. Forecasting oil production in North Dakota using the seasonal autoregressive integrated moving average (S-ARIMA). *Natural Resources*, 2015, 6(1): 16-26.
- Cipolla, C. L., Fitzpatrick, T., Williams, M. J., et al. Seismic-to-simulation for unconventional reservoir development. Paper SPE 146876 Presented at SPE Reservoir Characterisation and Simulation Conference and Exhibition, Abu Dhabi, UAE, 9-11 October, 2011.
- Cipolla, C. L., Lolon, E. P., Erdle, J. C., et al. Reservoir modeling in shale-gas reservoirs. *SPE Reservoir Evaluation & Engineering*, 2010, 13(4): 638-653.
- Clarkson, C. R., Solano, N., Bustin, R. M., et al. Pore structure characterization of North American shale gas reservoirs using USANS/SANS, gas adsorption, and mercury intrusion. *Fuel*, 2013, 103: 606-616.
- Curtis, J. B. Fractured shale-gas systems. *AAPG Bulletin*, 2002, 86(11): 1921-1938.
- Denney, D. Evaluating implications of hydraulic fracturing in shale-gas reservoirs. *Journal of Petroleum Technology*, 2009, 61(8): 53-54.
- Fink, R., Krooss, B. M., Amann-Hildenbrand, A. Stress-dependence of porosity and permeability of the Upper Jurassic Bossier shale: An experimental study. *Geological Society, London, Special Publications*, 2017, 454(1): 107-130.
- Freeman, C. M., Moridis, G., Ilk, D., et al. A numerical study of performance for tight gas and shale gas reservoir systems. *Journal of Petroleum Science and Engineering*, 2013, 108: 22-39.
- Fredd, C. N., McConnell, S. B., Boney, C. L., et al. Experimental study of hydraulic fracture conductivity demonstrates the benefits of using proppants. Paper SPE 60326 Presented at SPE Rocky Mountain Regional/Low-Permeability Reservoirs Symposium and Exhibition, Denver, Colorado, 12-15 March, 2000.
- Gale, J. F. W., Laubach, S. E., Olson, J. E., et al. Natural fractures in shale: A review and new observations. *AAPG Bulletin*, 2014, 98(11): 2165-2216.
- Gao, C., Lee, W. J., Spivey, J. P., et al. Modeling multilayer gas reservoirs including sorption effects. Paper SPE 29173 Presented at SPE Eastern Regional Meeting, Charleston, West Virginia, 8-10 November, 1994.
- Ghahfarokhi, P. K., Carr, T., Bhattacharya, S., et al. A fiber-optic assisted multilayer perceptron reservoir production modeling: A machine learning approach in prediction of gas production from the marcellus shale. Paper URTEC 2902641 Presented at SPE/AAPG/SEG Unconventional Resources Technology Conference, Houston, Texas, 23-25 July, 2018.
- Hamdi, H., Clarkson, C. R., Esmail, A., et al. Huff-n-puff (HNP) pilot design in shale reservoirs using dual-porosity, dual-permeability compositional simulations. *European Association of Geoscientists & Engineers*,

- 2020, 2020: 1-30.
- Hübner, R., Steinhauser, M., Lehle, C. A dual-stage two-phase model of selective attention. *Psychological Review*, 2010, 117(3): 759-784.
- Jia, X., Zhang, F. Applying data-driven method to production decline analysis and forecasting. Paper SPE 181616 Presented at SPE Annual Technical Conference and Exhibition, Dubai, UAE, 26-28 September, 2016.
- Khan, H., Louis, C. An artificial intelligence neural networks driven approach to forecast production in unconventional reservoirs-comparative analysis with decline curve. Paper IPTC 21350 Presented at International Petroleum Technology Conference, Virtual, 23 March-1 April, 2021.
- Kingma, D. P., Ba, J. Adam: A method for stochastic optimization. Paper Presented at 3<sup>rd</sup> International Conference on Learning Representations, San Diego, 7-9 May, 2015.
- Kucuk, F., Sawyer, W. K. Transient flow in naturally fractured reservoirs and its application to Devonian gas shales. Paper SPE 9397 Presented at SPE Annual Technical Conference and Exhibition, Dallas, Texas, 21-24 September, 1980.
- Kwak, S. G., Kim, J. H. Central limit theorem: The cornerstone of modern statistics. *Korean Journal of Anesthesiology*, 2017, 70(2): 144-156.
- Lee, J., Sidle, R. Gas-reserves estimation in resource plays. *SPE Economics & Management*, 2010, 2(2): 86-91.
- Liu, J., Ma, S., Shen, W., et al. Image feature recognition and gas permeability prediction of Gaomiaozhi bentonite based on digital images and machine learning. *Advances in Geo-Energy Research*, 2022, 6(4): 314-323.
- Luo, S., Ding, C., Cheng, H., et al. Estimated ultimate recovery prediction of fractured horizontal wells in tight oil reservoirs based on deep neural networks. *Advances in Geo-Energy Research*, 2022, 6(2): 111-122.
- Mao, X., Shen, C., Yang, Y. B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Advances in Neural Information Processing Systems*, 2016, 29: 2802-2810.
- Matsumoto, M., Nishimura, T. Mersenne twister: A 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 1998, 8(1): 3-30.
- Neal, D. B., Mian, M. A. Early-time tight gas production forecasting technique improves reserves and reservoir description. *SPE Formation Evaluation*, 1989, 4(1): 25-32.
- Ning, Y., Kazemi, H., Tahmasebi, P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Computers & Geosciences*, 2022, 164: 105126.
- Nobakht, M., Ambrose, R., Clarkson, C. R., et al. Effect of completion heterogeneity in a horizontal well with multiple fractures on the long-term forecast in shale-gas reservoirs. *Journal of Canadian Petroleum Technology*, 2013, 52(6): 417-425.
- Ozkan, E., Raghavan, R., Apaydin, O. G. Modeling of fluid transfer from shale matrix to fracture network. Paper SPE 134830 Presented at SPE Annual Technical Conference and Exhibition, Florence, Italy, 19-22 September, 2010.
- Pedrosa, O. A. Pressure transient response in stress-sensitive formations. Paper SPE 15115 Presented at SPE California Regional Meeting, Oakland, California, 2-4 April, 1986.
- Qin, Y., Song, D., Chen, H., et al. A dual-stage attention-based recurrent neural network for time series prediction. Paper Presented at Proceedings of the 26<sup>th</sup> International Joint Conference on Artificial Intelligence, 19 August, 2017.
- Stehfest, H. Algorithm 368: Numerical inversion of Laplace transforms [D5]. *Communications of the ACM*, 1970, 13(1): 47-54.
- Suhag, A., Ranjith, R., Aminzadeh, F. Comparison of shale oil production forecasting using empirical methods and artificial neural networks. Paper SPE 187112 Presented at SPE Annual Technical Conference and Exhibition, San Antonio, Texas, 9-11 October, 2017.
- Sun, H., Chawathe, A., Hoteit, H., et al. Understanding shale gas flow behavior using numerical simulation. *SPE Journal*, 2015, 20(1): 142-154.
- Thompson, J. M., M'Angha, V. O., Anderson, D. M. Advancements in shale gas production forecasting-A marcellus case study. Paper SPE 144436 Presented at North American Unconventional Gas Conference and Exhibition, The Woodlands, Texas, 14-16 June, 2011.
- Tong, T., Li, G., Liu, X., et al. Image super-resolution using dense skip connections. Paper Presented at 2017 IEEE International Conference on Computer Vision, Venice, Italy, 22-29 October, 2017.
- Van Everdingen, A. F., Hurst, W. The application of the Laplace transformation to flow problems in reservoirs. *Journal of Petroleum Technology*, 1949, 1(12): 305-324.
- Vicente, R., Sarica, C., Ertekin, T. Horizontal well design optimization: A study of the parameters affecting the productivity and flux distribution of a horizontal well. Paper SPE 84194 Presented at SPE Annual Technical Conference and Exhibition, Denver, Colorado, 5-8 October, 2003.
- Wang, H. T. Performance of multiple fractured horizontal wells in shale gas reservoirs with consideration of multiple mechanisms. *Journal of Hydrology*, 2014, 510: 299-312.
- Warpinski, N. R., Mayerhofer, M. J., Vincent, M. C., et al. Stimulating unconventional reservoirs: Maximizing network growth while optimizing fracture conductivity. *Journal of Canadian Petroleum Technology*, 2009, 48(10): 39-51.
- Wu, C., Wang, S., Yuan, J., et al. A prediction model of specific productivity index using least square support vector machine method. *Advances in Geo-Energy Research*, 2020, 4(4): 460-467.
- Yao, J., Li, Z., Liu, L., et al. Optimization of fracturing parameters by modified variable-length particle-swarm optimization in shale-gas reservoir. *SPE Journal*, 2021, 26(2): 1032-1049.
- Yin, F., Xue, X., Zhang, C., et al. Multifidelity genetic transfer: an efficient framework for production optimization. *SPE Journal*, 2021, 26(4): 1614-1635.
- Zhao, K., Du, P. A new production prediction model for

- multistage fractured horizontal well in tight oil reservoirs. *Advances in Geo-Energy Research*, 2020, 4(2): 152-161.
- Zhao, J., Sun, M., Pan, Z., et al. Effects of pore connectivity and water saturation on matrix permeability of deep gas shale. *Advances in Geo-Energy Research*, 2022, 6(1): 54-68.
- Zhang, L., Wang, S., Zhang, K., et al. Cooperative artificial bee colony algorithm with multiple populations for interval multiobjective optimization problems. *IEEE Transactions on Fuzzy Systems*, 2019, 27(5): 1052-1065.
- Zhong, Z., Sun, A. Y., Ren, B., et al. A deep-learning-based approach for reservoir production forecast under uncertainty. *SPE Journal*, 2021, 26(3): 1314-1340.