Journal of Petroleum Science and Engineering 186 (2020) 106682

Contents lists available at ScienceDirect

Journal of Petroleum Science and Engineering

journal homepage: http://www.elsevier.com/locate/petrol

Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model

Xuanyi Song [a], Yuetian Liu [a,*], Liang Xue [a], Jun Wang [b], Jingzhe Zhang [c], Junqiang Wang [a], Long Jiang [b], Ziyan Cheng [b]

[a] State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum, Beijing 102249, China
[b] SINOPEC Research Institute of Exploration and Development Company, Dongying 257015, China
[c] Mork Family Department of Chemical Engineering and Materials Science, University of Southern California, Los Angeles, CA 90089-1211, USA

ARTICLE INFO

Keywords:
Deep learning
LSTM
Time sequence data
PSO
Production forecasting

ABSTRACT

Oil production forecasting is one of the most critical issues during the exploitation phase of the oilfield. The limitations of traditional approaches make time-series production prediction still challenging. With the development of artificial intelligence, high-performance algorithms make reliable production prediction possible from the data perspective. This study proposes a Long Short-Term Memory (LSTM) neural network based model to infer the production of fractured horizontal wells in a volcanic reservoir, which addresses the limitations of traditional method and shows accurate predictions. The LSTM neural network enables to capture the dependencies of the oil rate time sequence data and incorporate the production constraints. Particle Swarm Optimization algorithm (PSO) is employed to optimize the essential configuration of the LSTM model. For evaluation purpose, two case studies are carried out with the production dynamics from synthetic model and from the Xinjiang oilfield, China. Towards a fair evaluation, the performance of the proposed approach is compared with traditional neural networks, time-series forecasting approaches, and conventional decline curves. The experiment results show that the proposed LSTM model outperforms other approaches.

1. Introduction

Nowadays petroleum still remains strategic energy of the world. It is believed that the oil and gas resources are more easily recovered from the existing oilfield (Amirian et al., 2018). The oil production rate over time is the important basis to manage and maintain the oilfield. The adjustment for the development methods can be made in time to make the reservoir production as much as possible and obtain a maximum economic profit when knowing the production change in advance. However, the prediction of time-series oil production rate is challenging due to its nonlinear variation. The incorporation of the production constrains make it more complex.

There are three major methods to infer the time-series production rate. Reservoir numerical simulation is the most common approach to give a liquid rate forecast. The construction of an exact numerical model is tedious and time consuming (Nwaobi and Anandarajah, 2018). The process includes establishing a geological model, a numerical model, and history matching (Clarkson et al., 2015; Kalra et al., 2018), which requires formation data, rock fluid data and so on. Analytical models are also used to describe the underground flow mechanism of hydrocarbons. However, the abstraction from reservoir to physical model and mathematic model needs to introduce a bunch of assumptions about boundary conditions, compressibility, capillary force, thermal effects and so on (Zhang et al., 2019; Clarkson and Qanbari, 2016; Du et al., 2017). Although the assumptions enable us to simplify the problem and get solutions, they narrow down the application scope of the models. Besides, the analytical approach sometimes relies on physical experiments which are long lasting and expensive. Traditional decline curve though can give a time-series rate curve, the production constraints and other factors that influence the well performance cannot be taken into consideration. Furthermore, the oil rate from the above models are all ideal curves. The practical oil rate data are not as smooth as their output at all. The complex phase behavior during flowing, the adjustment of production constraints, and process of well stimulation make the oil rate change in a sophisticated way which cannot be represented by conventional models.

* Corresponding author.
E-mail address: lyt51s@163.com (Y. Liu).

https://doi.org/10.1016/j.petrol.2019.106682
Received 4 July 2019; Received in revised form 8 November 2019; Accepted 10 November 2019
Available online 12 November 2019

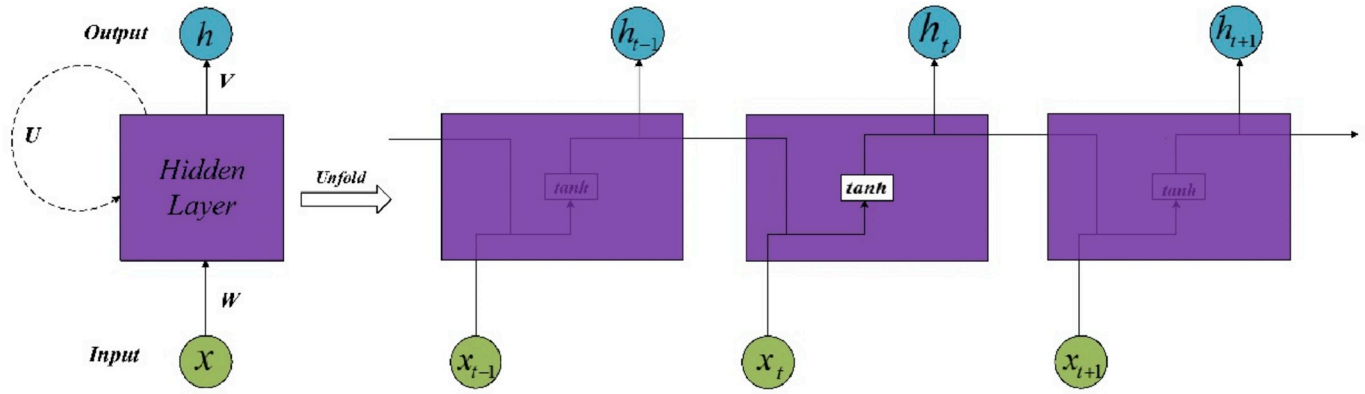


Fig. 1. The structure of traditional RNN.

Recently, with the development of artificial intelligence, more and more excellent algorithms are introduced into the petroleum industry, which provides new ideas for complicated problems. Guideline on how to build a robust predictive model has been illustrated by Schuetter et al. (2018) taking Random Forest (RF), Support Vector Machine (SVM), Gradient-boosting machine, and multidimensional Kriging as examples. As for the production problem, attempts have been made. The first-year production was explored by Luo et al. (2018) and Wang et al. (Wang and Chen, 2019). Accurate oil recovery forecasting models were built using SVM (Panja et al., 2018; Han and Bian, 2018). The production quality of the Permian Basin Wolfcamp was investigated based on SVM and RF (Zhong et al., 2015). However, these approaches are all about point data prediction, which means that they can only forecast the initial rate, cumulate production, or the recovery factor, not the time-series production rate.

Cao et al. (2016) utilized ANN to explore the relation between tubing head pressure and production rate. Then the time-series oil rate can be predicted according to the variation of tubing head pressure of the corresponding time. However, tubing head pressure coincides with the oil rate, meaning the data of liquid rate and tubing head pressure are recorded at the same time. So, it is not suitable for future prediction. Ahmadi et al. (2013) established a model based on the ANN, which set pressure and temperature as input and oil rate as output. Although the performance of the model for offshore oilfield is good, the precise pressure and temperature data of every time step cannot be obtained for the onshore oilfield due to the less utilization of multiphase flow meters. So, it is unrealistic to apply this model to predict the oil rate of the onshore oilfield. Fulford et al. (2015) proposed a machine learning model which can give a time-series oil rate prediction. However, the model should specify a decline pattern in advance which cannot suit the whole production variation history and also limits the prediction range.

When it comes to time-series prediction, LSTM has attracted much attention recently. Zhang et al. (2018a) used LSTM to predict water table depth in agriculture areas which associates with complex and heterogeneous hydrogeological characteristics, boundary conditions, and human activities. The results proved that LSTM has a strong ability for time series data prediction. Well log information can be reconstructed by LSTM with the consideration of variation trend and context information with depth (Zhang et al., 2018b). The results from the LSTM were of higher accuracy than traditional Fully Connected Neural Network. The dynamic feature of wind field can be captured within five-percent error by LSTM (Qin et al., 2019). LSTM can also give an accurate large time scale of mid-to-long term prediction for photovoltaic power generation (Han et al., 2019). Time series tourism flow can also be inferred by LSTM (Li and Cao, 2018). LSTM based model is established to predict spatio-temporal PM 2.5 (Tong et al., 2019), which gave excellent results. The impressive application results of LSTM in different areas prove that it can capture not only the variation trend of data but also characterize the dependence relationship of time sequence data. Therefore, this study attempts to use this unique approach to forecast the time-series oil rate. The novelty of this proposed methodology lies in using the PSO to design structure of LSTM neural network to improve the process of production forecasting.

This paper is organized as follows. Section 2 explains the relevant theory of the proposed model. Section 3 gives the specific workflow to establish the LSTM model and provides two application case studies. Section 4 summarizes the main findings.

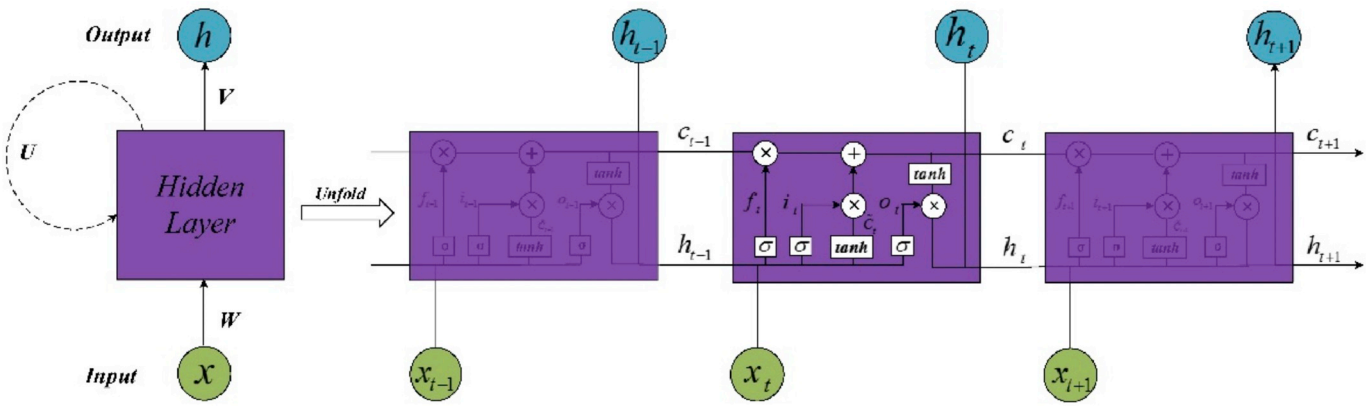


Fig. 2. The typical structure of LSTM.

2. Methodology

2.1. Basic theory of LSTM neural network

The precursor of LSTM neural network is Recurrent Neural Network (RNN). RNN, the evolution of Multi-layer Perception, is skilled in settling sequential data. A simple RNN structure whose recurrent layer is unfolded into a full network is shown in Fig. 1. x is input, h is the hidden state which gives the network memory ability. The subscripts $t-1$, t, and $t+1$ represent different time steps. W, U, and V are hyper-parameters of different layers. The distinctive characteristic lies in that the output of hidden layer with the present information is transferred to the hidden layer of the next time step as part of the input. This kind of loop can preserve the information of the previous step to keep the data dependency, which improves the ability of learning and abstracting from the sequential data. However, the vanishing gradient problem during the calculation of the back-propagation makes the influence of the input cannot be far transferred (Bengio et al., 1994); in other words, the performance of RNN is not satisfying for long-term dependence problem.

Therefore, LSTM neural network is proposed to improve RNN. The inner structure of the hidden layer in LSTM is more complicated comparing with RNN (Hochreiter and Schmidhuber, 1997). As shown in Fig. 2, a typical LSTM block is configured mainly by memory cell state, forget gate, input gate, and output gate. The crucial element, memory cell state, runs down through the entire chain to selectively add or remove information to the cell state with the help of the three gates.

The specific procedures of LSTM neural network at time step t are as follows:

(a) Decide what information should be discarded from the previous cell state C_{t-1} in the forget gate f_t.

$$f_t = \sigma(\mathrm{W_f} x_t + U_f h_{t-1} + \mathrm{b_f}) \tag{1}$$

(b) Identify what information of input x_t should be stored into the cell state C_t in the input gate, where the input information i_t and the candidate cell state $\tilde{C}_t$ should be updated.

$$i_t = \sigma(\mathrm{W}_i x_t + U_i h_{t-1} + \mathrm{b}_i) \tag{2}$$

$$\tilde{C}_t = \tanh(\mathrm{W}_c x_t + \mathrm{U_c h}_{t-1} + \mathrm{b}_c) \tag{3}$$

(c) Update the cell state of the present time step C_t, which combine the candidate memory $\tilde{C}_t$ and the long-term memory C_{t-1}.

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

(d) Confirm the outcome h_t in the output gate where the output information o_t and the cell state C_t is used.

$$o_t = \sigma(\mathrm{W}_o x_t + \mathrm{U_o h}_{t-1} + \mathrm{b}_o) \tag{5}$$

$$h_t = o_t * \tanh(C_t) \tag{6}$$

where W is the input weights, U is the recurrent weights, b is the bias, the subscripts f, i, o represent the forget gate, input gate, and output gate. The activation function of the three gates σ is the sigmoid function which enables the valuable between 0 and 1. The activation function tanh is the hyperbolic tangent function, which squeezes the value between -1 and 1. Both of the two activation functions are utilized to enhance the nonlinearity of the network, which are expressed as:

$$\sigma\left(x\right) = \frac{1}{1+e^{-x}} \tag{7}$$

$$\tanh\left(x\right) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{8}$$

The subtle design of the structure of LSTM, which involves memory cell state regulated by the three gates, solves the vanishing gradient problem of the RNN. Therefore, LSTM performs better in capturing and extracting the history information and predicting future development for long term dependencies issues of sequence data.

2.2. The proposed LSTM model framework

The productivity of wells changes with time and other influential factors. It is suitable to utilize LSTM model to explore the production mechanism from the relevant data. The architecture of the proposed LSTM model is depicted in Fig. 3. The dimension of input data is decided according to the number of influential factors of oil rate. The LSTM layer is embedded to memorize and extract containing information from input data, where the ADAM optimizer is used to update the weights and bias and the mean square error is set as the fitness function. The final output is given by a fully connected layer after the termination condition is reached.

2.3. Particle Swarm Optimization (PSO)

In the process of constructing LSTM model, the number of neurons in the LSTM layer and the number of previous time steps oil rate used as input variables to predict the next time oil rate called time window size

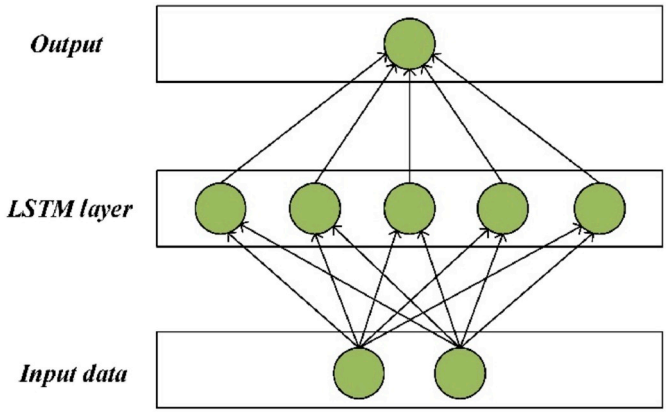


Fig. 3. Sketch map of the propsoed LSTM model.

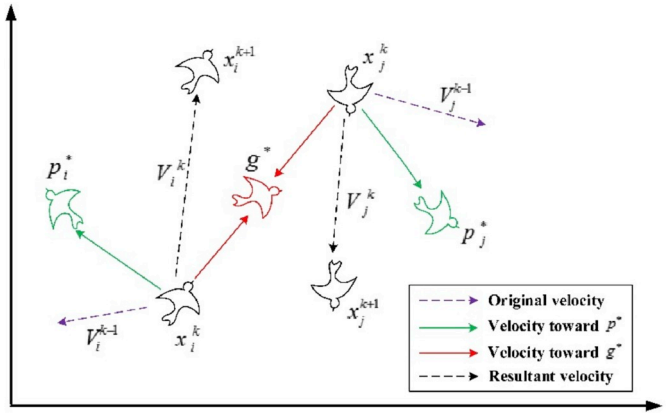


Fig. 4. The schematic of the optimization process of PSO.

are not fixed but crucial to prediction performance. To enhance the efficiency and accuracy of the model, the Particle Swarm Optimization (PSO) algorithm, widely used for optimization problems due to its easy realization, arithmetic stability and good convergence (Niknam et al., 2008, 2009; Kennedy and Eberhart, 1999; Niknam, 2006, 2010; Gaing, 2004), is utilized to search the optimal number of neurons and time window size for LSTM in this study.

PSO is a heuristic search algorithm inspired by the biological and sociological behavior of bird flock in searching for food (Kennedy and Eberhart, 1995). During the foraging of the bird swarm, birds are considered as particles and each of them is characterized by its position and velocity. The movement pattern of each particle is a result of incorporating the overall movement trend of the entire swarm and its own cognition, which are called global search and local search respectively. The global best position g* is shared among all particles while at the same time each particle also retains its own history of the best local position p*. As shown in Fig. 4, the value of g* and p* both guide the particles adjusting their flight directions and pointing to its destination. The iterative form of velocity and position are shown in Eq. (9) and Eq. (10). The particles keep on finding and improving the best position until they saturate or the maximum number of iterations is achieved.

$$v_i^{k+1} = \omega v_i^k + c_1 r_1 \left(p_i^* - x_i^k\right) + c_2 r_2 \left(g^* - x_i^k\right) \tag{9}$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \tag{10}$$

where x_i^k and v_i^k are the position and velocity vector for the particle i at the k th iteration respectively. ω is the inertia weight, p_i^* is the best individual particle position, and g^* is the global best position for all particles. c_1 and c_2 are two positive acceleration constants standing for personal and global nature of the swarm. r_1 and r_2 are two random values in the range of [0,1].

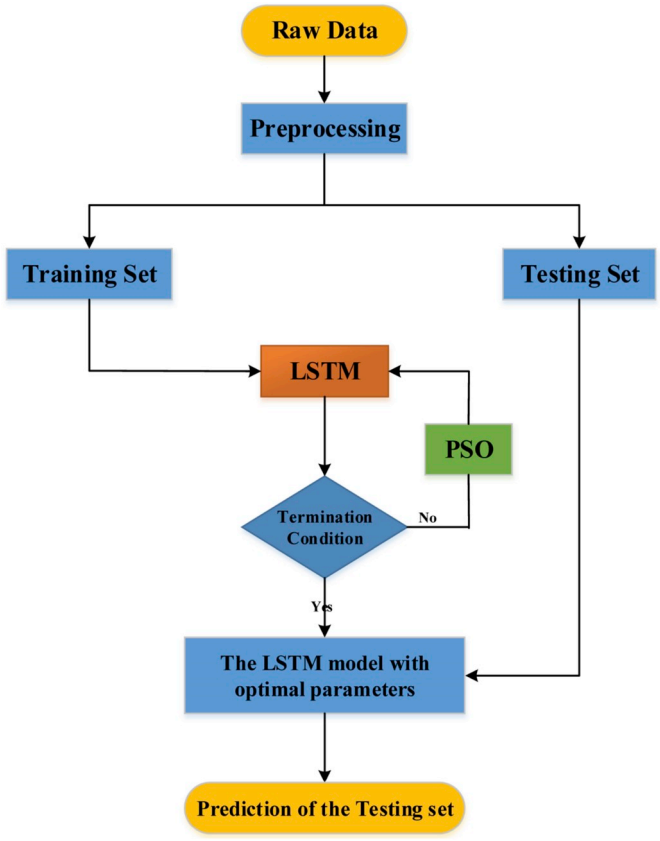


Fig. 5. Overall workflow of our suggested model.

2.4. Model evaluation criteria

Several indexes, mean absolute percentage error (MAPE), mean absolute error (MAE) and root mean square error (RMSE), were employed to evaluate the performance of models. The MAPE provides a useful measure of prediction accuracy in a forecasting method which usually expresses accuracy as a percentage. The MAE directly gives an average difference between the outcomes of the model and the actual data. The RMSE represents the standard deviation of the prediction results of the model. A smaller value indicates better model performance. The three criteria are defined below:

$$MAPE = \left(\frac{1}{n}\sum_{i=1}^{n}\frac{y_i - \hat{y}_i}{y_i}\right) \times 100 \tag{11}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} y_i - \hat{y}_i \tag{12}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{13}$$

where y_i and $\hat{y}_i$, are the i th actual value and the i th predicted value of n samples respectively.

3. Case studies

In this section, the overall workflow of the study is firstly introduced. Then we provide two example applications for the time-series production forecast. In the first case, ideal production variation data of a fractured horizontal well from a numerical simulator are employed to manifest the feasibility of the proposed model. In the second one, data from actual oilfield are collected to further prove the robustness of the model. Finally, a parametric analysis of the LSTM model is demonstrated.

3.1. The entire workflow of the case studies

The experiments are conducted for volcanic reservoir which cannot maintain the pressure or enhance the oil recovery by water injection due to early water breakthrough through natural fractures. Most of the volcanic reservoirs are with depletion development strategy whose liquid rate is only controlled by choke size. The proposed model is to search the oil rate variation mechanism of the volcanic reservoir under different production constraints, where the oil rate of time window size and the corresponding choke size are used as input and the next time oil rate as output. Therefore, if the future choke size is set, the future oil rate can be predicted by the model. The overall structure of this work is illustrated in Fig. 5. Each step in the flowchart is described in detail below:

Step 1. Data Pre-processing. The collected oil rate data and production constraints data need to be processed. Firstly, outliers should be detected and thrown away to reduce the noise. Then the data set is normalized to the interval [0, 1] to avoid model parameters being dominated by large or small data range since LSTM network is sensitive to the scale of the input data. The normalization formulas is shown as Eq. (14)

$$x_{new} = \frac{x_{old} - x_{\min}}{x_{\max} - x_{\min}} \tag{14}$$

where x_{new} represents the normalized data points inserted in the training process, while x_{old}, $x_{\min}$ and $x_{\max}$ are the real values of the sample data and the lower and upper constraints of these real data corresponding to design variables or targets, respectively.

Based on the characteristics of volcanic reservoir and LSTM neural

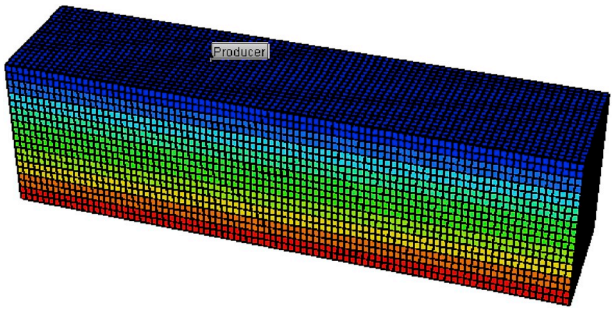


Fig. 6. Ideal model of volcanic reservoir.

network, the model can be described as

$$O(t) = F(C(t), O(t-1), O(t-2), \cdots, O(t-n)) \tag{15}$$

where $O(t)$ is the predicted oil rate, $C(t)$ is the choke size of at the t time step, $O(t-1), O(t-2), \cdots, O(t-n)$ is the previous oil rate and the number n is the time window size also called time lag. Therefore the processed data are reframed as the following formation where the time window size is assumed as 3:

$$\overbrace{\begin{Bmatrix} C(4) & O(1) & O(2) & O(3) \\ C(5) & O(2) & O(3) & O(4) \\ \vdots & \vdots & \vdots & \vdots \\ C(t) & O(t-3) & O(t-2) & O(t-1) \end{Bmatrix}}^{Input} \quad \overbrace{\begin{Bmatrix} O(4) \\ O(5) \\ \vdots \\ O(t) \end{Bmatrix}}^{Output}$$

Step 2. Training LSTM model. The reframed data are split into two parts; the first 80% is taken as a training set, and the last 20% as a testing set. The LSTM model is fed by the training set during which the time window size and number of neurons of LSTM layer are optimized by PSO algorithm until they satisfy the termination condition.

Step 3. Forecasting with LSTM model. The purpose is to use the past production dynamics to forecast a period of the future oil rate. The testing set directly obtained in step 2 contains the future information. Hence, the inputs of the testing should be iteratively updated with the predicted value. So the input formation is:

$$\overbrace{\begin{pmatrix} C(t+1) & O(t-2) & O(t-1) & O(t) \\ C(t+2) & O(t-1) & O(t) & O'(t+1) \\ \vdots & \vdots & \vdots & \vdots \\ C(t+n) & O'(t+n-3) & O'(t+n-2) & O'(t+n-1) \end{pmatrix}}^{Input} \quad \overbrace{\begin{pmatrix} O'(t+1) \\ O'(t+2) \\ \vdots \\ O'(t+n) \end{pmatrix}}^{Output}$$

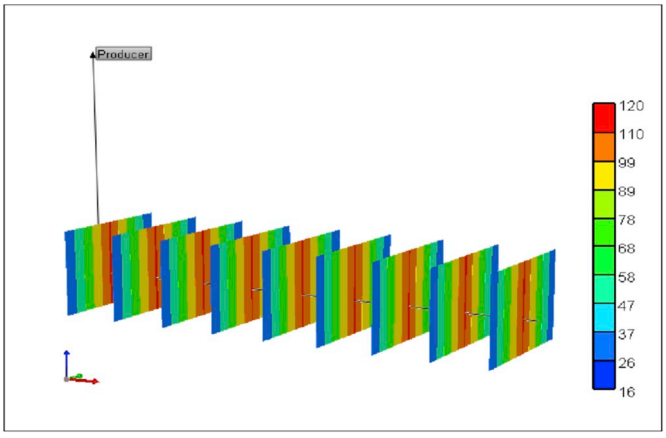


Fig. 7. The fracture model of horizontal well.

Table 1
The basic parameters of the ideal model.

Basic Parameters	Value
Initial pressure/Mpa	22
Temperature/°C	40
Permeability of matrix/md	0.4
Permeability of fracture/md	15
Porosity of matrix/%	8
Porosity of fracture/%	0.2
Well length/m	1000

where $O'(t+n)$ is the predictive value at the $t+n$ time step. The corresponding results of the testing set can then be get from the optimal LSTM model obtained above. We can gain the final results after denormalizing the output data.

The LSTM network is realized on the basis of Keras (Keras Documentation,), which is a deep learning library using Tensorflow (Abadi et al., 2016) as backend. The whole workflow is coded in python 3.7 and executed it on Intel ® Core™ i7-2600 3.40 GHz CPU.

3.2. Ideal production variation

To verify the feasibility of the proposed model, oil production rates with 351 data points from an ideal volcanic reservoir model shown in Fig. 6 are collected, because the data from simulators are with less noise and can better characterize the performance of the proposed model. The fracture model of the horizontal well for the volcanic reservoir is configured in Fig. 7. Other basic reservoir parameters of the ideal model are listed in Table 1, which based on the practical reservoir of Xinjiang oilfield.

The oil rate and the corresponding choke size are shown in Fig. 8. Due to the generated data without noise, the preprocessing process applied is just normalization. As can be seen from Fig. 8, the oil rate shows a decrease trend while the choke size keep the same. Because of the stationary choke size, the prediction of the oil rate can just consider the oil rate time sequence. In this condition, the propsoed LSTM model is evaluated by comparisons with the traditional decline curve analysis, the ordinary neural networks which are fully-connected ANN and RNN, the traditional time series forecasting methods which are Autoregressive Moving Average (ARMA) and Auto-Regression Integrated Moving Average (ARIMA).

The decline curve used to represent the production profile in this study can be expressed as (Arps, 1945):

$$Q = \frac{Q_i}{(1 + \mathrm{nD_i t})^{\frac{1}{n}}} \tag{16}$$

where the oil rate Q is a function of time t, Q_i is the initial oil rate, D_i is

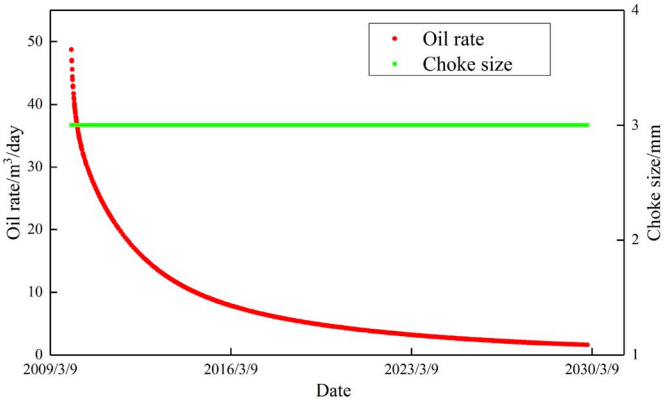
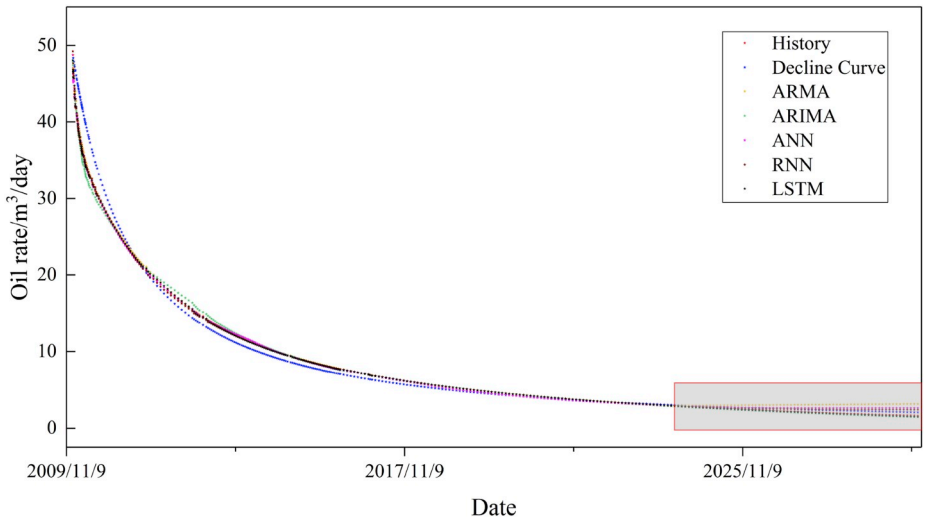
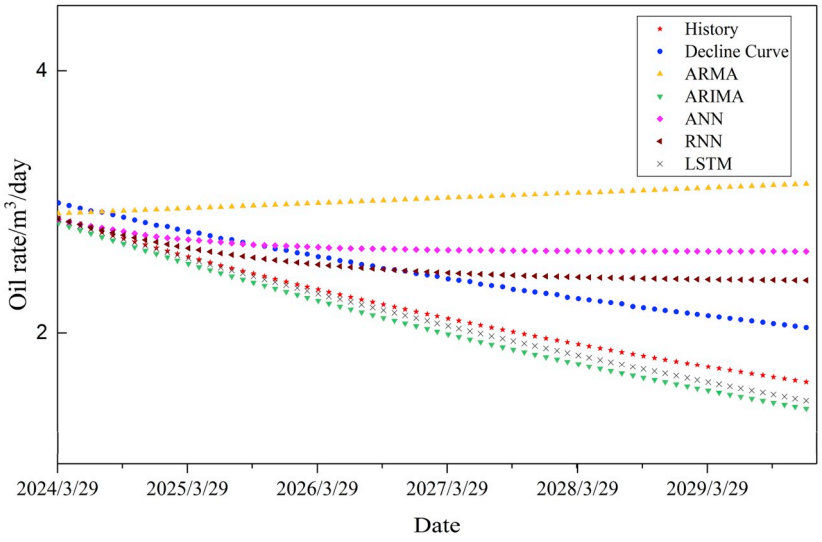


Fig. 8. The raw data visualization of the synthetic model.



(a) Experimental results of ideal model



(b) Zooming of the testing part of the results

Fig. 9. Prediction performance of the six models.

the initial decline rate and n is the hyperbolic-decline coefficient. The decline behavior of the production rate depends upon the D_i and n for a well.

The initial decline rate is 0.0025 and the decline index is 0.9 in this case. The LSTM model is established following the workflow explained in the previous section. The fundamentals of ARMA and ARIMA can refer to (Srivastava, 2015) and the models in this study are implemented using Statsmodels library (Seabold and Perktold, 2010). The ARMA parameters $p = 2$, $q = 2$ and the ARIMA parameters $p = 2$, $d = 1$, $q = 2$ are decided by Partial Auto Correlation Function (PACF) and Auto Correlation function (ACF) curves (Balaguer et al., 2008; Cryer and Chan, 2008; Mohammadi et al., 2005; Toth et al., 2000). According to the test results, the search ranges for neurons of the LSTM layer and the time window size are: neurons $\in [1, 10]$, time window size $\in [1, 10]$. The PSO finds that the best LSTM structure is with 4 neurons and time window size is 4. To compare the ANN model and RNN model fairly, we use the same parameter settings.

Fig. 9 shows the prediction results of the decline curve, ARMA, ARIMA, fully-connected ANN, RNN and the proposed LSTM model. As shown in Fig. 9 (a), the prediction of training set for the five models, ARMA, ARIMA, fully-connected ANN, RNN and the LSTM model fit very well. However, the decline curve prediction are deviated from the

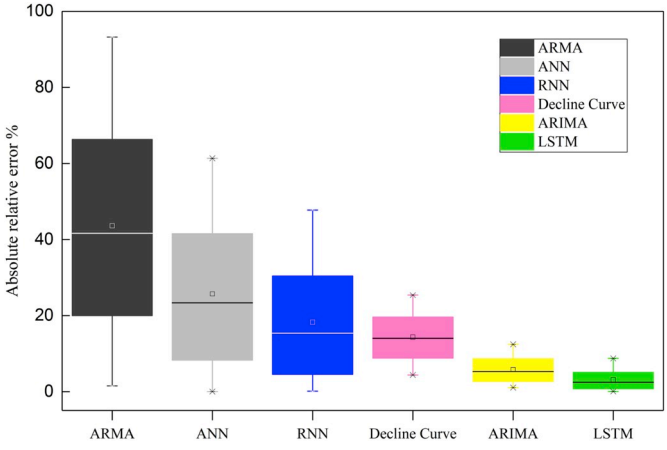


Fig. 10. Boxplots of relative error of the three models for testing data.

Table 2
Performance comparision for the three models.

Models	MAPE	MAE	RMSE
ARMA	43.63	0.85	0.95
ANN	25.68	0.49	0.58
RNN	18.26	0.35	0.42
Decline analysis	14.28	0.29	0.30
ARIMA	5.81	0.11	0.13
LSTM	3.12	0.06	0.07

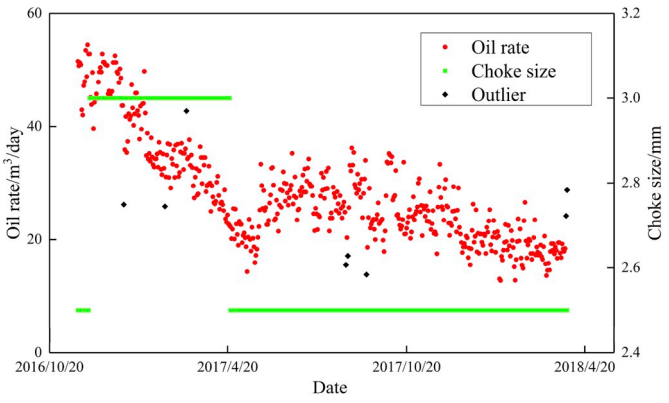


Fig. 11. The raw data visualization of the real field case.

history value. The testing part of the results are zoomed in Fig. 9 (b), from which we can clearly find that the LSTM model outperform others and the ARIMA follows. The other four model are far away from the history data. It indicates that the models, LSTM and ARIMA, with the ability to capture the dependencies of the data series can predict the oil rate variation more accurately.

The percentage relative error distribution of the testing set is depicted in Fig. 10. It can be found that the ARMA model to predict the values that are never appeared before have significant errors. The results of the decline curve better than fully-connected ANN and RNN model. The ARIMA model and LSTM model are both with high accuracy while the LSTM is more accurate than the ARIMA model. To further quantify the predictive performance of the models, the three criterias, MAPE, MAE, and RMSE of the testing set of the six models are listed in Table 2. It is clear that the values of the proposed LSTM model are the global minimum values among other reference models. It indicates that the method proposed can better extract the dependencies of the time series data to improve the accuracy of the oil rate forecasting.

3.3. Complex production variation

The production data of the actual oilfield is not as smooth as the data from the simulator. In contrast, they may significant discontinuous because of frequent shut-ins and choke size changing. The conventional method cannot give predictions under this circumstance. The data of this case is collected from a well of the Xinjiang volcanic reservoir located in northwestern China to certify the predictive ability of the proposed model. The fractured horizontal well has been producing for around 17 months with 501 data points which are shown in Fig. 11. As can be seen, the choke size initially was set as 2.5 mm, it was then changed to 3 mm. However, due to the big choke size, the formation pressure depleted rapidly which cause a sharply decrease of oil rate. Therefore, the choke size was changed to 2.5 mm to maintain a stable decrease of the formation pressure. The black diamond shaped data are detected outliers with Local Outlier Factor method (Breunig et al., 2000) and the outliers were replaced by the average of the previous and next time step oil rates. Then the data was normalized and feed the LSTM model. Because the future oil rate is decided by the choke size of the correspondent time and the previous time oil rate. The inputs for the predictive model have multiple input valuables. The ARMA and ARMA model cannot be used in this complex condition. The traditional decline curve, fully-connected ANN and RNN are the compared with the LSTM model.

The production prediction of the decline curve analysis must observe that the well should be under the same production constraint and the oil rate should show a stable decline trend. Therefore, the analysis of this problem is divided into two parts as shown in Fig. 12. The initial decline rate is 0.005 and the decline index is 0.4 for the first part. The initial decline rate and decline index are 0.002 and 0.5 respectively for the second part.

The PSO found that the best LSTM structure is with 15 neurons and the time window size is 6. For a fair comparison, the fully-connected ANN and RNN are set with the same structure. The results of the three neural network models are shown in Fig. 13. Both of the results of fully-connected ANN and RNN are acceptable for the training set, as shown in Fig. 13 (a) and (b). However, the results for the testing set are both with obvious errors, which implies that the generalizability is not strong enough to predict the data that did not appear during training. In contrast, the proposed LSTM model extracts the inner variation mechanism of this well, so it gives a good prediction for both training set and testing set, as shown in Fig. 13 (c).

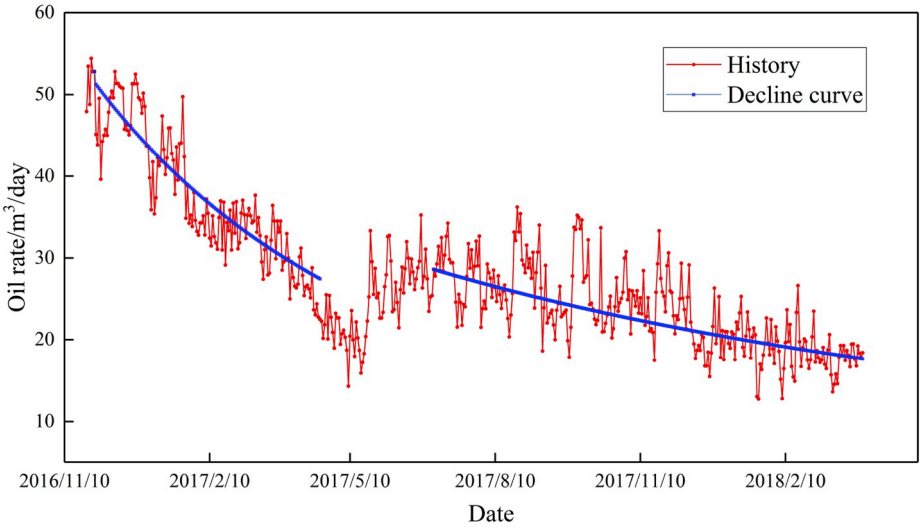
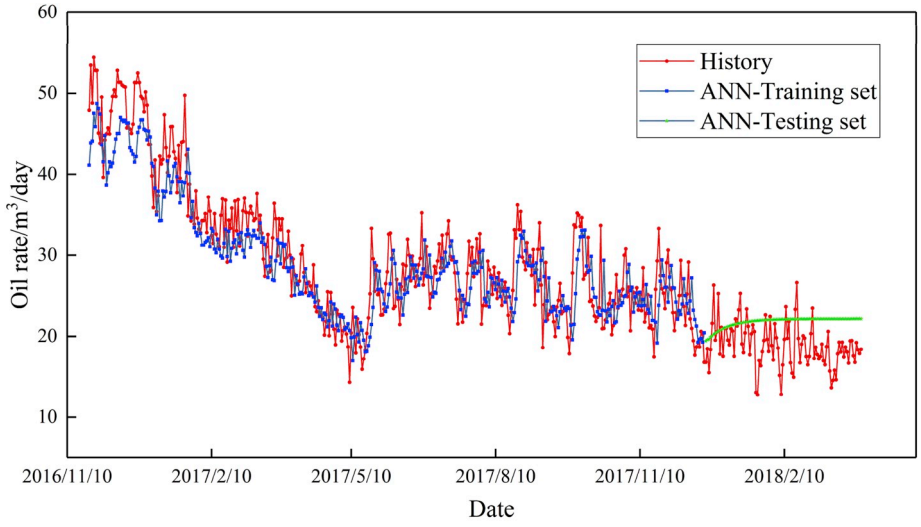
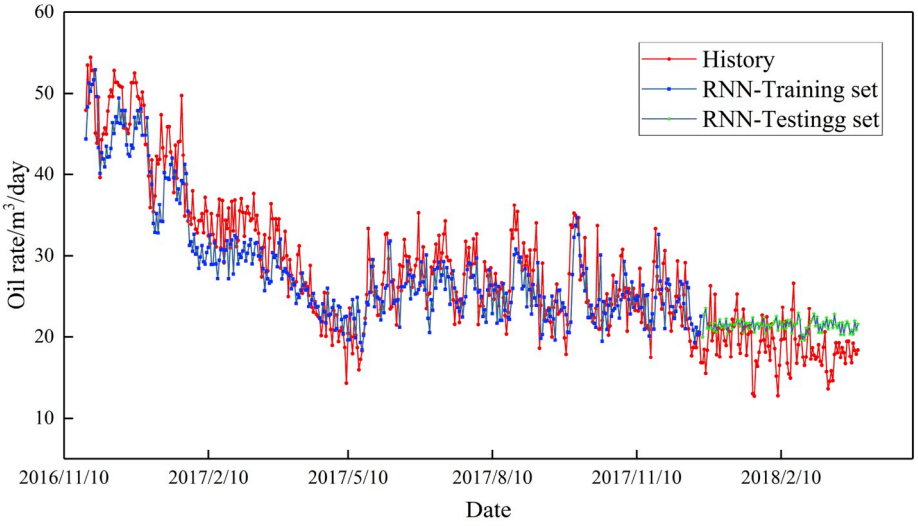


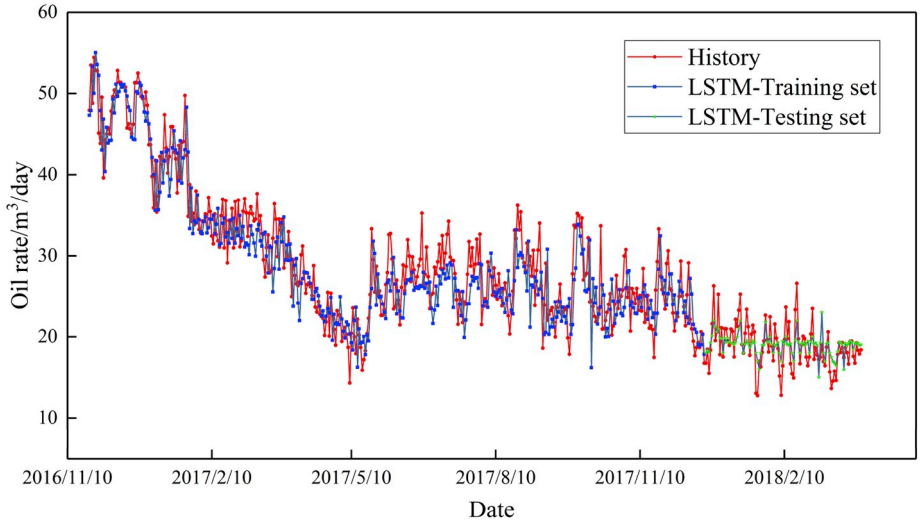
Fig. 12. The decline curve analysis for the real field case.



(a)The prediction results of the ANN model



(b) The prediction results of the RNN model



(c) The prediction results of the proposed LSTM model

Fig. 13. Comparison between history and prediction oil rate of the three models.

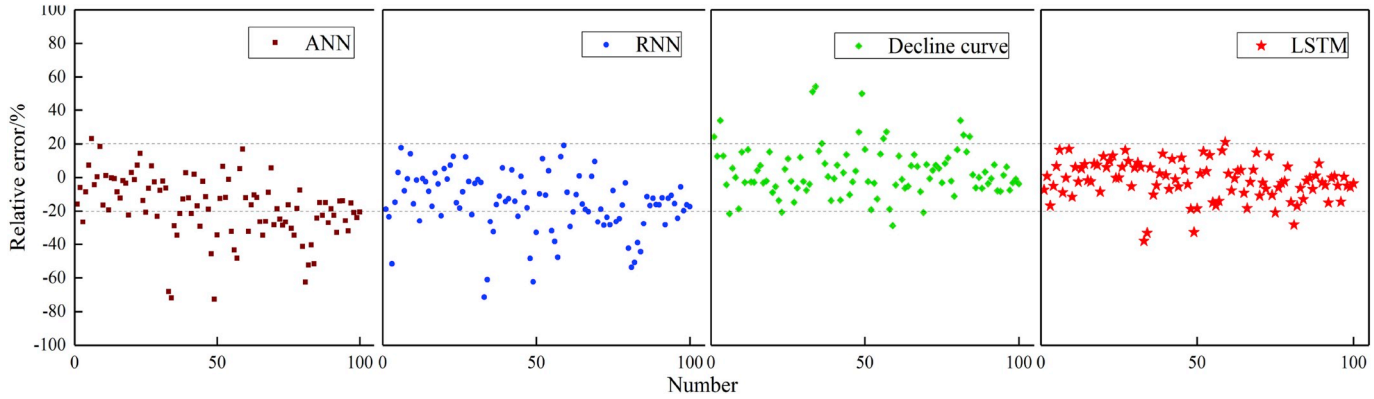


Fig. 14. The relative error distribution of the four methods.

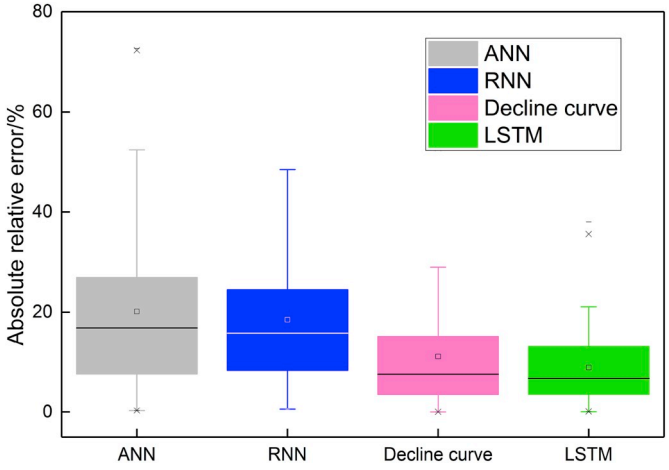


Fig. 15. Boxplots of absolute relative error of the four models.

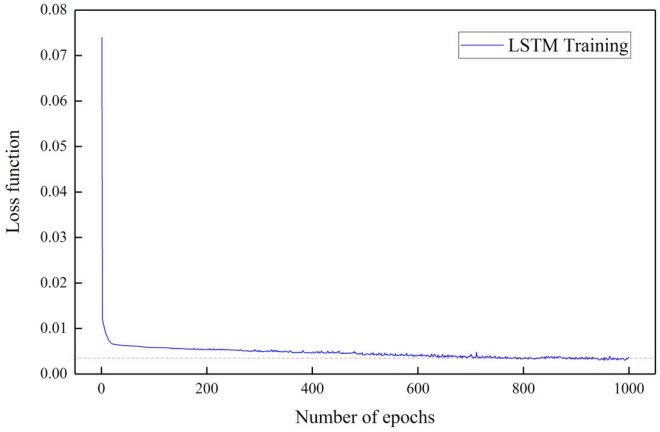


Fig. 16. The variation of the loss function with training epoch.

The testing set is the most critical set for validation purpose since the results of the testing set is what we really want. Thus, the results of the testing set are further analyzed. Fig. 14 graphically illuminates the relative error distribution characteristics of the four methods. It can be seen from Fig. 14 that many points of fully-connected ANN and RNN are beyond 20% and some of them even close to 80%. Although the decline curve is better than the previous two model, a certain of data are with large error and some of them close to 60%. However, the data obtained from LSTM most are concentrated within 20%, and only four points outside of the 20% but not larger than 40%. It can be concluded that the proposed model can not only give a global trend prediction but also with accurate single point prediction.

Fig. 15 presents the absolute relative error distribution range of the four models. As can be seen, nearly 50% of the predictions of fully connected-ANN and RNN are greater than 20%, so these two models are not suitable for this problem. Decline curve is better than the previous two models. For the LSTM model, more than 75% of the predictions are within 20%, and 50% of data less 10%, which is superior to the other three methods.

The statistic evaluation indexes are reported in Table 3. It shows that the fully-connected ANN and RNN are with larger error. The results of decline analysis are better than the two networks in terms of the MAPE, MAE, and RMSE. For the LSTM model, the prediction results outperform other three models. The mean absolute percentage error and the mean absolute error are 9.88% and 1.60 respectively, which indicates the proposed method with higher accuracy. Still, the smaller RMSE shows that the LSMT model is also stable than the other three methods. The oil rate forecasting accuracy in this work mainly relies on the generalization ability of the models which can be derived from the constructing of the testing set. Therefore, the generalization ability of LSTM model is superior to the tradition neural network fully-connected ANN and RNN according to the experimental results. As for the traditional decline analysis, although it can give a global variation tendency of oil rate, the prediction error of single points are not reliable. The reported values confirm that the LSTM model can not only give a variation trend of the oil rate but also gives more accurate predictions for oil rate of every single day.

Table 3
Performance comparision for the four models.

Model	MAPE	MAE	RMSE
ANN	20.14	3.46	4.10
RNN	18.49	3.18	3.81
Decline analysis	14.15	2.31	2.73
LSTM	9.88	1.60	2.02

3.4. Parametric analysis of the LSTM architecture

The number of epochs influences the performance of a LSTM model with fixed architecture since the loss function of the model varies with the training process. As shown in Fig. 16, the loss function of the proposed model for the real field case keeps decreasing before the 700th epoch which means the model is not well trained. Only when the loss funtion become stable can the model be used to predict. However, an excessive training may lead to the over-fitting problem which will decrease the generalization ability.

The number of neurons and the hidden layers determine the architecture of the LSTM model and it is believed that an increasing of hidden

Table 4
The prediction performance of LSTM model with different structure.

Stacked layers	Neurons	MAPE	MAE	RMSE
1	15	9.88	1.60	2.02
2	24	9.65	1.48	1.86
3	11	9.52	1.37	1.74

layers contributes to improve the model performance (LeCun et al., 2015). Therefore, the experiments using the LSTM neural network with 2 and 3 LSTM layers are implemented to investigate the effects of the architectures on the model performance based on the real field case. The experimental results are listed in Table 4, which shows that an increase of hidden layer is capable of slightly improving the prediction accuracy but the degree is small for this problem. However, the execution time is almost doubled when each addition of hidden layer.

4. Conclusion

In this study, a novel hybrid model which is LSTM neural network optimized by PSO is proposed for predicting the daily oil rate of volcanic reservoir. The model enables engineers and decision-makers to acquire the information of the well performance in advance, which can guide them better managing and adjusting the development plan of reservoirs. The ability of the proposed model to forecast daily oil rate of two cases is discussed. Meanwhile, the proposed model is also compared with the traditional methods.

The major conclusions are as follows:

(1) The proposed model provides a promising method for predicting the time-series oil rate, as evidenced by satisfactory performance on oil rate data from the simulator and the Xinjiang oilfield. It can serve as an alternative tool to give a reliable prediction.
(2) The proposed model outperforms the decline curve analysis, the traditional neural networks, and ARMA and ARIMA models. It can more accurately capture the complex variation pattern of the oil rate under multiple factors. Moreover, its generalization ability is higher when predict the data never appeared during training for the sequence problem.
(3) The parameter setting is crucial to the LSTM model performance. The training epoch should be decided carefully to avoid inadequate training and over-fitting problem. The addition of the hidden layer can improve the accuracy of the LSTM to some extent at the cost of increasing computation time.

In the future, we will investigate the coupling of LSTM neural network and Convolutional Neural Network to handle temporal-spatial dependent production prediction of different wells.

Acknowledgements

The paper is sponsored by the National Natural Science Foundation of China (No. 51374222), the National Basic Research Program of China (973 Program, No. 2015CB250905), CNPC Major Scientific Research Project (No. 2017E-0405) and SINOPEC Major Scientific Research Project (No. P18049-1).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.petrol.2019.106682.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al., 2016. Tensorflow: a system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16), pp. 265–283.

Ahmadi, M.A., Ebadi, M., Shokrollahi, A., Majidi, S.M.J., 2013. Evolving artificial neural network and imperialist competitive algorithm for prediction oil flow rate of the reservoir. Appl. Soft Comput. 13 (2), 1085–1098. https://doi.org/10.1016/j.asoc.2012.10.009.

Amirian, E., Dejam, M., Chen, Z., 2018. Performance forecasting for polymer flooding in heavy oil reservoirs. Fuel 216, 83–100. https://doi.org/10.1016/j.fuel.2017.11.110.

Arps, J., 1945. Analysis of decline curves. Trans. AIME 160 (1), 228–247. https://doi.org/10.2118/945228-G. ARMA.

Balaguer, E., Palomares, A., Sorie, E., Martin-Guerrero, J.D., 2008. Predicting service request in support centers based on nonlinear dynamics, ARMA modeling and neural networks. Expert Syst. Appl. 34 (1), 665–672.

Bengio, Y., Simard, P., Frasconi, P., 1994. Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. 5 (2), 157–166.

Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J., 2000. LOF: identifying density-based local outliers. In: ACM Sigmod Record, vol. 29. ACM, pp. 93–104. No. 2.

Cao, Q., Banerjee, R., Gupta, S., Li, J., Zhou, W., Jeyachandra, B., 2016. Data driven production forecasting using machine learning. In: SPE Argentina Exploration and Production of Unconventional Resources Symposium. Society of Petroleum Engineers. https://doi:10.2118/180984-MS.

Clarkson, C.R., Qanbari, F., 2016. A semi-analytical method for forecasting wells completed in low permeability, undersaturated CBM reservoirs. J. Nat. Gas Sci. Eng. 30, 19–27. https://doi.org/10.1016/j.jngse.2016.01.040.

Clarkson, C.R., Williams-Kovacs, J.D., Qanbari, F., Behmanesh, H., Sureshjani, M.H., 2015. History-matching and forecasting tight/shale gas condensate wells using combined analytical, semi-analytical, and empirical methods. J. Nat. Gas Sci. Eng. 26, 1620–1647. https://doi.org/10.1016/j.jngse.2015.03.025.

Cryer, J.D., Chan, K.S., 2008. Time Series Analysis: with Applications in R. Springer.

Du, D.F., Wang, Y.Y., Zhao, Y.W., Sui, P.S., Xia, X., 2017. A new mathematical model for horizontal wells with variable density perforation completion in bottom water reservoirs. Pet. Sci. 14 (2), 383–394. https://doi.org/10.1007/s12182-017-0159-0.

Fulford, D.S., Bowie, B., Berry, M.E., Bowen, B., Turk, D.W., 2015. Machine learning as a reliable technology for evaluating time-rate performance of unconventional wells. SPE Econ. Manag. 8 (01), 23–39. https://doi.org/10.2118/174784-MS.

Gaing, Z.L., 2004. A particle swarm optimization approach for optimum design of PID controller in AVR system. IEEE Trans. Energy Convers. 19 (2), 384–391.

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

Han, S., Qiao, Y.H., Yan, J., Liu, Y.Q., Li, L., Wang, Z., 2019. Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. Appl. Energy 239, 181–191. https://doi.org/10.1016/j.apenergy.2019.01.193.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Kalra, S., Tian, W., Wu, X., 2018. A numerical simulation study of CO 2 injection for enhancing hydrocarbon recovery and sequestration in liquid-rich shales. Pet. Sci. 15 (1), 103–115. https://doi.org/10.1007/s12182-017-0199-5.

Kennedy, J., Eberhart, R.C., 1995. Particle swarm optimization. In: Proceedings of the 1995 IEEE International Conference on Neural Networks, Perth, Australia.

Kennedy, J., Eberhart, R., 1999. Particle swarm optimization. In: Proceedings of the IEEE International Conference on Neural Networks, Piscataway, NJ, pp. 1942–1948.

Keras Documentation. Keras Documentation. https://keras.io.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553). https://doi.org/10.1038/nature14539, 436–4 4 4.

Li, Y., Cao, H., 2018. Prediction for tourism flow based on LSTM Neural Network. Procedia Comput. Sci. 129, 277–283. https://doi.org/10.1016/j.procs.2018.03.076.

Luo, G., Tian, Y., Bychina, M., Ehlig-Economides, C., 2018. Production optimization using machine learning in Bakken shale. In: Unconventional Resources Technology Conference, Houston, Texas, 23-25 July 2018. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers, pp. 2174–2197. https://doi.org/10.15530/URTEC-2018-2902505.

Mohammadi, K., Eslami, H.R., Dardashti, S.D., 2005. Comparison of regression ARIMA and ANN models for reservoir inflow forecasting using snowmelt equivalent (a case study of Karaj). J. Agric. Sci. Technol. 7, 17–30.

Niknam, T., 2006. An approach based on particle swarm optimization for optimal operation of distribution network considering distributed generators. In: IECON 2006-32nd Annual Conference on IEEE Industrial Electronics. IEEE, pp. 633–637.

Niknam, T., 2010. A new fuzzy adaptive hybrid particle swarm optimization algorithm for non-linear, non-smooth and non-convex economic dispatch problem. Appl. Energy 87 (1), 327–339. https://doi.org/10.1016/j.apenergy.2009.05.016.

Niknam, T., Firouzi, B.B., Nayeripour, M., 2008. An efficient hybrid evolutionary algorithm for cluster analysis. World Appl. Sci. J. 4 (2), 300–307.

Niknam, T., Amiri, B., Olamaie, J., Arefi, A., 2009. An efficient hybrid evolutionary optimization algorithm based on PSO and SA for clustering. J. Zhejiang Univ. - Sci. 10 (4), 512–519. https://doi.org/10.1631/jzus.A0820196.

Nwaobi, U., Anandarajah, G., 2018. Parameter determination for a numerical approach to undeveloped shale gas production estimation: the UK Bowland shale region application. J. Nat. Gas Sci. Eng. 58, 80–91. https://doi.org/10.1016/j.jngse.2018.07.024.

Panja, P., Velasco, R., Pathak, M., Deo, M., 2018. Application of artificial intelligence to forecast hydrocarbon production from shales. Petroleum 4 (1), 75–89. https://doi.org/10.1016/j.petlm.2017.11.003.

Qin, Y., Li, K., Liang, Z., Lee, B., Zhang, F., Gu, Y., et al., 2019. Hybrid forecasting model based on long short term memory network and deep learning neural network for wind signal. Appl. Energy 236, 262–272. https://doi.org/10.1016/j.apenergy.2018.11.063.

Schuetter, J., Mishra, S., Zhong, M., LaFollette, R., 2018. A data-analytics tutorial: building predictive models for oil production in an unconventional shale reservoir. SPE J. 23 (04), 1–075. https://doi.org/10.2118/189969-PA.

Seabold, S., Perktold, J., 2010. Statsmodels: econometric and statistical modeling with python. In: Proceedings of the 9th Python in Science Conference, vol. 57. Scipy, p. 61.

Srivastava, T., 2015. A Complete Tutorial on Time Series Modeling in R. Analytics Vid. https://www.analyticsvidhya.com/blog/2015/12/complete-tutorial-time-series-modeling/.

Tong, W., Li, L., Zhou, X., Hamilton, A., Zhang, K., 2019. Deep learning $PM_{2.5}$ concentrations with bidirectional LSTM RNN. Air Qual. Atmos. Health 12 (4), 411–423. https://doi.org/10.1007/s11869-018-0647-4.

Toth, E., Brath, A., Montanari, A., 2000. Comparison of short-term rainfall predication models for real-time flood forecasting. J. Hydrol. 239 (1–4), 132–147.

Wang, S., Chen, S., 2019. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. J. Pet. Sci. Eng. 174, 682–695. https://doi.org/10.1016/j.petrol.2018.11.076.

Zhang, J., Zhu, Y., Zhang, X., Ye, M., Yang, J., 2018. Developing a Long Short-Term Memory (LSTM) based model for predicting water table depth in agricultural areas. J. Hydrol. 561, 918–929. https://doi.org/10.1016/j.jhydrol.2018.04.065.

Zhang, D., Chen, Y., Meng, J., 2018. Synthetic well logs generation via recurrent neural networks. Pet. Explor. Dev. 45 (4), 598–607. https://doi.org/10.11698/PED.2018.04.06.

Zhang, R.H., Zhang, L.H., Tang, H.Y., Chen, S.N., Zhao, Y.L., Wu, J.F., Wang, K.R., 2019. A simulator for production prediction of multistage fractured horizontal well in shale gas reservoir considering complex fracture geometry. J. Nat. Gas Sci. Eng. 67, 14–29. https://doi.org/10.1016/j.jngse.2019.04.011.

Zhong, M., Schuetter, J., Mishra, S., Lafollette, R.F., 2015. Do data mining methods matter?: a Wolfcamp Shale case study. In: SPE Hydraulic Fracturing Technology Conference. Society of Petroleum Engineers. https://doi.org/10.2118/173334-MS.