Potential for Prediction of Water Saturation Distribution in Reservoirs Utilizing Machine Learning Methods

Qitao Zhang 1,2, Chenji Wei 3, Yuhe Wang 4, Shuyi Du 1,2, Yuanchun Zhou 2,5 and Hongqing Song 1,2, *

1 School of Civil and Resource Engineering, University of Science and Technology Beijing, Beijing 100083, China; romainzhang1128@gmail.com (Q.Z.); 15702449699@163.com (S.D.)
2 National & Local Joint Engineering Lab for Big Data Analysis and Computer Technology, Beijing 100190, China; zyc@cnic.cn
3 Research Institute of Petroleum Exploration and Development, PetroChina, Beijing 100083, China; weichenji@petrochina.com.cn
4 Department of Petroleum Engineering, Texas A&M University at Qatar, Doha 999043, Qatar; yuhe.wang@qatar.tamu.edu
5 Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China
*
Correspondence: songhongqing@ustb.edu.cn

Received: 12 August 2019; Accepted: 16 September 2019; Published: 20 September 2019

Abstract: Machine learning technology is becoming increasingly prevalent in the petroleum industry, especially for reservoir characterization and drilling problems. The aim of this study is to present an alternative way to predict water saturation distribution in reservoirs with a machine learning method. In this study, we utilized Long Short-Term Memory (LSTM) to build a prediction model for forecast of water saturation distribution. The dataset deriving from monitoring and simulating of an actual reservoir was utilized for model training and testing. The data model after training was validated and utilized to forecast water saturation distribution, pressure distribution and oil production. We also compared standard Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU) which are popular machine learning methods with LSTM for better water saturation prediction. The results show that the LSTM method has a good performance on the water saturation prediction with overall AARD below 14.82%. Compared with other machine learning methods such as GRU and standard RNN, LSTM has better performance in calculation accuracy. This study presented an alternative way for quick and robust prediction of water saturation distribution in reservoir.

Keywords: machine learning; water saturation; long short-term memory; artificial neural network; computational time

1. Introduction

In the oilfield development process, due to reservoir heterogeneity and different production modes, there often exist great amounts of remaining oil in reservoirs. As a result, the prediction of remaining oil or water saturation distribution is of great significance for oilfield future development [1]. Accurate prediction of water saturation distribution in a reservoir facilitate to tap the remaining oil in the reservoir.

There are mainly two traditional methods to analyze and predict water saturation at the current time. On the one hand, some researchers calculate the water saturation in the reservoir based on petrophysical models. Archie [2] used the formation resistivity and porosity to calculate water saturation $S_w$ in the reservoir. Some other researchers adopted saturation–height function like Leverett...
J function [3] and Heseldin method [4] to formulate capillary pressure data and then water saturation could be obtained. These methods are relatively time-consuming, and their sample range is hard to reflect the reservoir heterogeneity. On the other hand, numerical simulation has been widely used in the petroleum industry. A variety of simulation techniques have evolved for different development conditions and reservoir types [5–9]. Reservoir simulation plays a dominant role in formulating the reservoir development plan and improving oil recovery at the current stage [10]. In this background, many researchers studied water saturation distribution utilizing numerical simulation for different type of problem, such as water imbibition [11–13], performance of EOR (enhance oil recovery) methods [14,15], unconventional reservoir exploitation [16,17], etc. However, it will take a long time for history matching and forecast calculation, and the prediction cost is relatively higher.

In recent years, a data driven approach and artificial intelligence technology have been advantageous for quick exploitation determination with or without physical model. Take remaining oil saturation prediction in a reservoir as example, machine learning methods can show good ability to solve this problem. Reservoir engineers could propose rational scenarios for tapping of remaining oil and long-term healthy development of oilfields combining machine learning results with numerical simulation. In previous study, numbers of machine learning methods have already been applied into petroleum industry to analyze data, find patterns and predict target variables [18]. The utilization of a machine learning method in petroleum industry often focus on two purposes. For one purpose, try to figure out the relationship between petrophysics and reservoir properties. Siplngarmers et al. [19] used the BP neural network method to learn different relative permeability curve data from a certain amount of papers and experiments, so as to develop a liquid/liquid and liquid/gas two-phase relative permeability predictors. Talebi et al. [20] utilized two improved algorithms: Multilayer Perceptron (MLP) neural network and Radial Basis Function (RBF) for efficient estimation of saturation pressure of reservoir oil. Nouri-Taleghani et al. [21] used three machine learning methods separately to predict the fracture density with full set log data as inputs. Masoudi et al. [22] adopted Bayesian Network and K2 algorithm to find interrelationships between petrophysical parameters and optimum production feature. Hegde et al. [23] conducted real time rate of penetration (ROP) optimization in drilling by utilizing Data-Driven model. Tian and Horne [24] applied three machine learning methods as linear regression, convolution kernel and ridge regression to interpret flow-rate, pressure and temperature data from permanent downhole gauge. For another, try to forecast future well production performance. Gupta et al. [25] forecasted gas production in unconventional resources using data mining and time series analysis. Schuetter et al. [26] adopted simple regression random forest (RF), support-vector regression (SVR), gradient-boosting machine (GBM) and multidimensional Kriging to predict oil production in an unconventional shale reservoir focusing on establishment of robust predictive models. Ma et al. [27] predicted the oil production using the novel multivariate nonlinear model based on traditional Arps decline model and a kernel method. Kamari et al. [28] used least square support machine vector method (LS-SVM) to develop a robust model for predicting surfactant-polymer flooding performance. The prediction on well performance and reservoir characteristics including water saturation distribution tried by Gomez [29] and Mohaghegh [30] with a data-driven reservoir modeling was successful while without detailed implementation description. This attempt indicated the data driven method is feasible and effective to characterize reservoir and facilitate production. Actually, typical machine learning approach should be modified or selected properly in terms of practical production to produce best results. The specific process introduction on machine learning implementation must benefit reservoir engineers in the petroleum industry.

Combined with a specific machine learning approach, the aim of this study is to predict water saturation distribution in future production, which is a typically time-series problem. Recurrent Neural Network (RNN) was found to be good at solving these problems. The output of the previous moment is partial input of next time step of RNN which is appropriate for short-term prediction. However, long-term dependencies problem of RNN limits its further development. To overcome this drawback, a Long-Short Term Memory (LSTM) method was presented. As a variant of RNN, LSTM has gradually
become a research hotspot in the field of machine learning in recent years. Significant results have been achieved in language translation, speech recognition and machine reading [31–35]. As an improved form of RNN, LSTM solves the long-term dependencies due to its unique data preservation mechanism which could remember long-term data properties. This characteristic has great advantages in dealing with large-scale multi-dimensional data and time-series problems. In the petroleum industry, LSTM was used to perform secondary generation of well logging data [36], forecast production decline of multiple wells [37] and predict rate of penetration (ROP) based on recorded drilling data [38]. However, this method has not been used for the prediction of water saturation distribution in reservoir yet. In terms of the huge scale of data amount and consistence between water saturation and time, LSTM is a comparatively appropriate machine learning method for water saturation prediction.

This paper was organized in four aspects. Firstly, a water saturation prediction model was established utilizing neural network LSTM with data processing, model training and testing. Then, the model was calibrated and validated quantitatively for further prediction. After that, the prediction and analysis of water saturation distribution were carried out to tap remaining oil. Finally, a comparison of accuracy and computational time between different machine learning methods was presented. In this study, we provide an alternative way for quick and robust prediction of water saturation distribution in a reservoir without empirical and numerical simulation.

2. Methodology

2.1. Data Model for Water Saturation Prediction with LSTM

Neural networks include many different forms. Most neural network forms are utilized to process static data, that is, the samples are single and have no connection with each other. However, for practical case involving time-series problems, neural network requires not only calculation to a single sample, but also ability to remember characteristics like a human. The problem of serialization can be well solved by the Recurrent Neural Network (RNN) method. However, standard RNN does have obvious problems, especially the Long-Term dependencies problem [39]. To solve this defect, researchers developed the Long-Short Term Memory (LSTM) neural network.

LSTM is an RNN variant which is optimized based on standard RNN. It is able to learn long-term information. LSTM was first introduced by Hochreiter and Schmidhuber in 1997 [40]. It has been widely used in different study fields in recent years and has achieved good performance in different research issues. The core idea of LSTM is to add three functional gates into the standard structure. The repeating network structure of LSTM is shown in Figure 1, where $x_t$ is the input data in one single step, $h_{t-1}$ is the output data of last step which is also inputted again into the network.

![Figure 1. Illustration of the repeating network structure of Long-Short Term Memory (LSTM).](image)

Compared with standard RNN structure, LSTM has an extra subline called cell state $C$ in the top of Figure 1, which could store information in the past like production information and water saturation.
distribution history for months. It runs straight down entire chains, just like the outline or storyline of a book, and after each calculation, the information in cell state will be updated.

The first key functional gate is “forget gate layer”. This gate determines which part of information should be thrown away from the cell state. After analyzing the input parameter $x_t$ and output results in last time step $h_{t-1}$, the forget gate layer will produce a value $f_t$ between 0–1 into the cell state $C_{t-1}$, using a sigma function $\sigma$ in Equation (1). Value 1 means “completely keep this” and 0 means “completely discarded” as shown in Equation (2) [41]:

$$\sigma(x) = \frac{1 - e^{-x}}{1 + e^{-x}}$$  \hspace{1cm} (1)

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$  \hspace{1cm} (2)

Another functional gate is namely “input gate layer” which determines what new information in input data of this time step needs to be stored in the cell state, as follows [41]:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$  \hspace{1cm} (3)

$$\tilde{C}_t = AF(W_C \cdot [h_{t-1}, x_t] + b_C)$$  \hspace{1cm} (4)

Equation (3) plays a similar role to Equation (2) and $i_t$ is a value between 0–1 to determine the information to be added into the cell state. An activation function (AF) is used to deal with input data $x_t$ and $h_{t-1}$ in Equation (4). Combine these two factors $i_t$ and $\tilde{C}_t$, specific information will be stored in cell state.

The old state $C_{t-1}$ can be multiplied by $f_t$ to forget the things computer decided to forget earlier then add $i_t$ with $\tilde{C}_t$. The result is the updated cell state in this time step $C_t$, as follows [41]:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$  \hspace{1cm} (5)

Finally, the last gate which controls the output results every step is namely “output gate layer”. It generates the output information $h_t$ which could include water saturation than we expected, based on the updated cell state $C_t$ and input data $x_t$ and $h_{t-1}$, as follows [41]:

$$h_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \cdot AF(C_t)$$  \hspace{1cm} (6)

2.2. Training Procedure

2.2.1. Data Acquisition and Processing

This paper selects a block in the Middle East as the research object with buried depth about 2400 m, average permeability of 57.4 mD, average porosity of 17.1% and initial formation pressure of 26.5 MPa. The exploitation conditions are relative ideal. There are 12 functional wells in this block and the well type is vertical well. Nine of these wells are production wells, numbered as: R-025, R-074, R-205, R-278, R-370, R-408, R-418, R-420, R-526. The other are injection (water) wells, numbered as: R-213, R-287, R-366. The well location in the block is shown in Figure 2. The block started production from August 1, 2004. In the beginning, wells production relied on formation energy. Then, from the third year of production, water injection for oil well production. The daily oil production per well ranged from 100 m$^3$/day to 1000 m$^3$/day.
The monitoring data can be obtained in wells, however data of water saturation distribution in a reservoir only could be acquired by numerical simulation according to well history data. As a result, the dataset for machine learning must be obtained including monitoring and simulating data together which also reflects a combination of data science and physical law. Then we can utilize the trained data model to predict water saturation in the same block faster and more accurately than numerical simulation in terms of a physical model. Based on the block geological model and well history information, the entire dataset was obtained after numerical simulation with 148-month production in this block, starting from August 1, 2004. In the process of history matching, the reservoir was divided into $25 \times 24$ grids in the X-Y plane, and was divided into 34 layers in the longitudinal direction. Permeability distribution, porosity distribution and initial water saturation distribution of the 22nd layer, one of main oil producing layers in the reservoir, are shown in Figure 2.

In order to make the model predict the distribution of water saturation efficiently, reasonable input parameters and output parameters should be assigned in advance. Input/output parameters in this study and their data formats for one month are shown in Tables 1 and 2:

### Table 1. Input parameters information in the model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Data Format for One Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Porosity distribution</td>
<td>$25 \times 24 \times 34$</td>
</tr>
<tr>
<td>Permeability distribution</td>
<td>$25 \times 24 \times 34$</td>
</tr>
<tr>
<td>Pressure distribution</td>
<td>$25 \times 24 \times 34$</td>
</tr>
<tr>
<td>Water saturation distribution</td>
<td>$25 \times 24 \times 34$</td>
</tr>
<tr>
<td>Oil production of 9 wells</td>
<td>9</td>
</tr>
<tr>
<td>Water injection of 3 wells</td>
<td>3</td>
</tr>
</tbody>
</table>
Table 2. Output parameters information in the model.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Data Format for One Month</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pressure distribution</td>
<td>$25 \times 24 \times 34$</td>
</tr>
<tr>
<td>Water saturation distribution</td>
<td>$25 \times 24 \times 34$</td>
</tr>
<tr>
<td>Oil production of 9 wells</td>
<td>9</td>
</tr>
</tbody>
</table>

Data assimilation is the basic work of data processing. For artificial neural networks with multiple input parameters, different input parameters could have different units and different values. If the values between different parameters are distinctive, the interferences between the parameters will be obvious. In order to eliminate this influence and improve computational efficiency of the model, data was assimilated in advance as follows [42]:

$$X' = \frac{X - \mu}{\delta},$$  \hspace{1cm} (7)

where $X$ is the initial value, $X'$ is its normalized value, $\mu$ and $\delta$ are mean and standard deviation, respectively.

2.2.2. Machine Learning Training Design

After obtaining the dataset, it is necessary to consider the division of the dataset for training, validation and testing [43]. A previous study shows different dataset partition has a certain influence on the training results and reasonable dataset partitioning can effectively improve the accuracy of the model [28]. The dataset of the model data in this paper was divided as shown in Table 3:

Table 3. Dataset partition for machine learning.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Size</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set</td>
<td>85</td>
<td>month</td>
</tr>
<tr>
<td>Validation set</td>
<td>15</td>
<td>month</td>
</tr>
<tr>
<td>Test set</td>
<td>48</td>
<td>month</td>
</tr>
</tbody>
</table>

The flow-process diagram for the usage of datasets is shown in Figure 3. At first, a 148-month history dataset can be obtained after data processing. Then the dataset was divided as: 1st–100th month for training set and validation set, 101st–148th months for test set. The training set is used for LSTM model training. By learning the information in the training set, the LSTM model could predict 101st–148th month’s results. Finally, we can compare the data between LSTM prediction results and test set data. Prediction evaluation could be carried out through this comparison.

In this paper, a deep learning framework Keras was utilized to build up a LSTM model. Before a LSTM model is built, length of a sequence of input data should be defined. For better description, it can be called “Moving Time Window” or “Sild Window”. In this model, the length of the moving time window was 10 months. Here, the principle of LSTM model calculation process is illustrated using Figure 4. At the beginning of calculation, 10 months of input data are obtained and then the model can predict the 11th month’s data. After that, the 11th month’s prediction results are added back into the moving time window (with other input data at 11th month). In that way, the 12th prediction results can be figured out as well, using the updated window data (2nd month to 11th month). By continuously cycling this step, the whole calculation could be done eventually.
In the training process, to prevent the overfitting and gradient vanishing problems, a regularization process is needed. Random dropout is a method to solve overfitting problem. In the learning process, weight of some nodes in the hidden layer is randomly reduced to zero, in other words. In this way, the model will not rely too much on some local features, and structural risk can be reduced [44–46].

In addition to dropout, activation function is also a key factor in the model. The purpose of the activation function is to add some nonlinear factors between the input and output of the neural network. Common activation functions include Sigmoid, Tanh and ReLU. The function curve and mathematic equation of these three activation functions are shown in Figure 5.

**Figure 3.** Flow-process diagram for usage of datasets.

**Figure 4.** Illustration of the LSTM calculation process.

In the training process, to prevent the overfitting and gradient vanishing problems, a regularization process is needed. Random dropout is a method to solve overfitting problem. In the learning process, weight of some nodes in the hidden layer is randomly reduced to zero, in other words. In this way, the model will not rely too much on some local features, and structural risk can be reduced [44–46].

In addition to dropout, activation function is also a key factor in the model. The purpose of the activation function is to add some nonlinear factors between the input and output of the neural network. Common activation functions include Sigmoid, Tanh and ReLU. The function curve and mathematic equation of these three activation functions are shown in Figure 5.

**Figure 5.** Diagram for different activation functions and its characteristics: (a) Sigmoid; (b) Tanh; (c) ReLU.
According to Figure 5, it can be seen that Sigmoid can compress the value between 0,1. However, because of its soft saturation characteristic, it is easy for Sigmoid to produce gradient vanishing problem, leading to training problems. Tanh compresses the value between $-1,1$. This characteristic solves the problem that Sigmoid’s output is not zero-centered and it turns out that Tanh has a higher convergence time than Sigmoid, but saturation problem has not been solved by Tanh as well. To sum up, while Sigmoid neurons are more biologically plausible than Tanh neurons, the latter work better for training multi-layer neural networks [47].

In this paper, the rectified linear unit (ReLU) was used for the activation function. The advantage is that its intrinsic function form is relatively simple (Sigmoid and Tanh contain exponential operations when deriving, and ReLU has almost no computation in derivation), and the convergence time is much faster than the function Sigmoid and Tanh [47,48].

2.2.3. Model Evaluation Criteria

When the calculation is completed, the model needs to be evaluated by statistical evaluation indicators, so as to verify the model. This paper used some error evaluation criteria to study the accuracy of model results. These methods include average relative deviation (ARD), average absolute relative deviation (AARD), coefficient of determination ($R^2$) and root mean square error (RMSE) as follows [49,50]:

$$ARD\% = \frac{100}{N} \sum_{i=1}^{N} \left( \frac{X_{i,\text{data}} - X_{i,\text{model}}}{X_{i,\text{data}}} \right)$$

$$AARD\% = \frac{100}{N} \sum_{i=1}^{N} \left( \frac{|X_{i,\text{data}} - X_{i,\text{model}}|}{X_{i,\text{data}}} \right)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N} \left( X_{i,\text{data}} - X_{i,\text{model}} \right)^2}{\sum_{i=1}^{N} \left( X_{i,\text{data}} - \text{average}(X_{i,\text{model}}) \right)^2}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left( X_{i,\text{data}} - X_{i,\text{model}} \right)^2}$$

where $N$ represents total number of data in each set, for example, in terms of evaluation of water saturation prediction, that is 48 (months) $\times$ 25 $\times$ 24 $\times$ 34 (water saturation data points for a single month). $X_{i,\text{data}}$ is the data value from each set, and $X_{i,\text{model}}$ is the corresponding data value calculated by the LSTM model.

3. Results and Discussion

3.1. LSTM Model Calibration

In this section, experiments of LSTM model calibration were carried out to ensure optimal model parameter values. Three model parameters including Length of moving time window (default value is 10), dropout rate (default value is 0.1) and number of hidden layers (default value is 1) were selected for further study and discussion.

First, we trained the same network architecture with different length of moving time window. Window length varied from 5 to 25 months. Fixing other parameters, the AARD of water saturation distribution prediction (48 months) for different window size could be presented as shown in Figure 6:
It can be seen from Figure 6 that, as the window length increases, the AARD curve decreases dramatically at first. When the window length reaches 10 months, AARD curve becomes flat and the value is around 12%. Then, as the window length continuously increases, the AARD increases slightly. Window length is significant to a LSTM model. If the length is too small, data information in the window could not be enough to support a good prediction performance. On the contrary, if the length of the window is too large, some important information in specific time node could be weakened. For different problems and different kind of data, the optimal window length could be obtained only through continuously experiments.

Then, the same network architecture was trained with different dropout rate. Dropout rate in these experiments varied from 0.1 to 0.9. Fixing other parameters, the AARD of prediction results (48 months) for different dropout rate could be presented as shown in Figure 7:

Figure 6. Effect of moving window length on prediction performance: (a) error curve of water saturation; (b) error curve of pressure; (c) error curve of oil rate.

Figure 7. Cont.
Figure 7. Effect of dropout rate on prediction performance: (a) error curve of water saturation; (b) error curve of pressure; (c) error curve of oil rate.

It can be seen from Figure 7 that, as dropout rate increase, the prediction error goes down dramatically at first. The AARD value becomes stable when the dropout rate ranges from 0.3 to 0.7 and then increases as dropout rate close to 0.9. In other words, dropout rate in this model exists optimal value which ranges from 0.3–0.7.

Number of hidden layers was also chosen to be a study object. Same network architecture was trained with different number of hidden layers which varied from 1 to 4. Fixing other parameters, the relevant AARD of prediction results (48 months) for different number of hidden layers could be presented as shown in Figure 8:

Figure 8. Effect of hidden layer number on prediction performance: (a) error curve of water saturation; (b) error curve of pressure; (c) error curve of oil rate.

From Figure 8, it can be known that the number of hidden layers has a positive relationship with the model prediction performance in this study. As the number of hidden layer increases, the prediction error goes down. This decrease is apparent when the layer number changes from 1 to 2 (i.e., AARD
of water saturation decreases from 20.22% to 12.46%) and relatively slight when the layer number changes from 2 to 4 (i.e., AARD of water saturation decreases from 12.46% to 11.07%). In theory, we can continuously increase the number of hidden layers to obtain the better prediction results. However, more hidden layers in the model mean longer computing time and could lead to overfitting problem. Hence, 2 hidden layers in this model is recommended.

### 3.2. Model Validation

According to the described model training procedure in the Section 2 and model calibration results in Section 3.1, a LSTM model can be developed. The values for all the necessary parameters are shown in Table 4.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Value</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning rate</td>
<td>0.01</td>
<td>#</td>
</tr>
<tr>
<td>Window length</td>
<td>10</td>
<td>month</td>
</tr>
<tr>
<td>Batch size</td>
<td>128</td>
<td>#</td>
</tr>
<tr>
<td>Number of hidden layers</td>
<td>2</td>
<td>#</td>
</tr>
<tr>
<td>Number of hidden dim</td>
<td>300</td>
<td>#</td>
</tr>
<tr>
<td>Dropout rate</td>
<td>0.3</td>
<td>#</td>
</tr>
<tr>
<td>Input layer neurons</td>
<td>81,611</td>
<td>#</td>
</tr>
<tr>
<td>Output layer neurons</td>
<td>9 for oil rate, 20,400 for pressure distribution, 20,400 for water saturation distribution</td>
<td>#</td>
</tr>
<tr>
<td>Total prediction time</td>
<td>48</td>
<td>month</td>
</tr>
<tr>
<td>Epoch</td>
<td>500</td>
<td>#</td>
</tr>
<tr>
<td>Sequence of layers</td>
<td>Input-dropout-hidden layer-dropout-hidden layer-output</td>
<td></td>
</tr>
</tbody>
</table>

The symbol # means dimensionless.

In the training process, a loss function is usually adopted to measure the model ability for fitting the training data. By monitoring the decreasing trend of the loss function curve, it can be figured out whether the model has achieved its best performance. Figure 9 shows loss decline curve in the training process for pressure distribution and water saturation distribution, respectively. It could be seen from the figures that model converges on pressure when calculating to 104 epochs, on water saturation when calculating to 179 epochs and on oil rate when calculating to 32 epochs. The calculation continued until the maximum epochs (500) is reached.

![Figure 9. Cont.](image)
To verify the model, it is necessary to analyze the accuracy of model calculation results. A variety of evaluation equations in Section 2.2.3 were selected to quantitatively evaluate the calculation results of the LSTM model. The analysis of pressure and water saturation calculation results are shown in the Table 5:

**Table 5.** The results of statistical error analysis performed in this study for the developed LSTM model.

<table>
<thead>
<tr>
<th></th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pressure distribution</strong></td>
<td>Training set &amp; Validation set</td>
<td>( R^2 ) 0.954</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average relative deviation, % 3.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average absolute relative deviation, % 10.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE, kPa 2636.49</td>
</tr>
<tr>
<td></td>
<td>Testing set</td>
<td>( R^2 ) 0.913</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average relative deviation, % 4.11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average absolute relative deviation, % 14.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE, kPa 3524.98</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>( R^2 ) 0.941</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average relative deviation, % 3.46</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average absolute relative deviation, % 11.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE, kPa 2947.28</td>
</tr>
<tr>
<td><strong>Water saturation distribution</strong></td>
<td>Training set &amp; Validation set</td>
<td>( R^2 ) 0.939</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average relative deviation, % 5.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average absolute relative deviation, % 11.35</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE, # 0.0313</td>
</tr>
<tr>
<td></td>
<td>Testing set</td>
<td>( R^2 ) 0.923</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average relative deviation, % 5.89</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average absolute relative deviation, % 12.41</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE, # 0.0343</td>
</tr>
<tr>
<td></td>
<td>Total</td>
<td>( R^2 ) 0.933</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average relative deviation, % 5.75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Average absolute relative deviation, % 11.81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RMSE, # 0.323</td>
</tr>
</tbody>
</table>

**Figure 9.** Loss decline curves in training process for pressure and water saturation: (a) Loss decline curve for pressure; (b) Loss decline curve for water saturation; (c) Loss decline curve for oil product.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training set &amp; Validation set</td>
<td></td>
</tr>
</tbody>
</table>
$R^2$ | 0.972 |
| Average relative deviation, % | 4.72 |
| Average absolute relative deviation, % | 5.89 |
| RMSE, # | 14.6 |
| Testing set |  
$R^2$ | 0.961 |
| Average relative deviation, % | 4.99 |
| Average absolute relative deviation, % | 6.28 |
| RMSE, # | 16.8 |
| Total |  
$R^2$ | 0.965 |
| Average relative deviation, % | 4.79 |
| Average absolute relative deviation, % | 6.01 |
| RMSE, # | 15.2 |

The symbol # means dimensionless.

Results in Table 5 indicate that the model has a good performance in training and testing process, respectively. Considering the great amount of data, the model overall AARD can still be controlled below 15% and the ARD is below 6.5% and the accuracy of the model is over 80%. In other words, Table 5 indicates that acceptable results are established and the prediction model is dependable. Moreover, the results obtained in this paper clearly show that the LSTM recurrent neural network is capable for efficient prediction of water saturation distribution in a reservoir with multi-layer.

Oil production is a significant output parameter. The comparison between production curve from well record and LSTM prediction model for the production well R-025 in the duration of prediction is as shown in Figure 10. According to the comparison results in Figure 10, it can be known that prediction results of oil production are acceptable with good precision.

![Figure 10](image-url)  
**Figure 10.** Comparison between actual well record and LSTM prediction in the duration of prediction (101st–148th month in whole time sequence) for oil production.

To intuitively display the prediction results of the model, the output results were split for parameter separation and then reprocessed into the initial data format. After that, they were imported into MATLAB for post-processing. Images of pressure and water saturation distribution for different layers at different time can be produced. The 22nd layer in the 48th month of prediction results (148th month in whole time sequence) was selected for an example. The comparison results of pressure distribution between Test set data and LSTM prediction results are shown in Figure 11:
Figure 10. Comparison between actual well record and LSTM prediction in the duration of prediction (101st–148th month in whole time sequence) for oil production.

To intuitively display the prediction results of the model, the output results were split for parameter separation and then reprocessed into the initial data format. After that, they were imported into MATLAB for post-processing. Images of pressure and water saturation distribution for different layers at different time can be produced. The 22nd layer in the 48th month of prediction results (148th month in whole time sequence) was selected for an example. The comparison results of pressure distribution between Test set data and LSTM prediction results are shown in Figure 11:

Figure 11. Comparison between Test set data and LSTM results at 48th month of prediction (148th month in whole time sequence) for pressure distribution: (a) Pressure distribution from Test set; (b) Pressure distribution from LSTM prediction; (c) Cross plot of pressure distribution for 22nd layer; (d) Cross plot of pressure distribution for all layers.

It can be seen from Figure 11 that, predicted pressure distribution from the LSTM model has obvious similarities with that from Test set. The comparison results indicate that the LSTM model has a good learning performance on the pressure fluctuation near the wellbore. To quantitatively study the prediction accuracy, cross plots of pressure data are produced as shown in Figure 11c. The abscissa represents the data in Figure 11a and the ordinate is the results predicted by the LSTM network model in Figure 11b. The red line in the figure stands for the line $y = x$, which means the ideal condition that the prediction results is completely consistent with the test set data. The blue point represents the data points from LSTM and there are 600 data points ($25 \times 24$) dispersed in this cross plot. The accuracy of the model prediction can be judged by the degree of coincidence between the blue data point and the red standard line. According to the Figure 11c, it can be known that the model has a good calculation accuracy for a single layer at each time node in terms of pressure distribution.

The comparison results of water saturation are shown in Figure 12. The results turn out that the LSTM model has a good predictive effect on the change of the water saturation distribution in the reservoir under water-injection condition. Prediction results show great response to the saturation changes near injection wells. In other words, the prediction results of water saturation distribution are credible and the LSTM have been proved to be capable in dealing with water saturation prediction problem.
The LSTM model has a good calculation accuracy for a single layer at each time node in terms of pressure distribution. The comparison results of water saturation are shown in Figure 12. The results turn out that the LSTM model has a good predictive effect on the change of the water saturation distribution in the reservoir under water-injection condition. Prediction results show great response to the saturation changes near injection wells. In other words, the prediction results of water saturation distribution are credible and the LSTM have been proved to be capable in dealing with water saturation prediction problem.

![Figure 12. Comparison between Test set data and LSTM results at 48th month of prediction (148th month in whole time sequence) for water saturation distribution: (a) Water saturation distribution from Test set; (b) water saturation distribution from LSTM prediction; (c) Cross plot of water saturation distribution for 22nd layer; (d) Cross plot of water saturation distribution for all layers.](image)

### 3.3. Water Saturation Prediction Analysis

After validation, the detailed analysis of calculation results was conducted. First, compared with numerical simulation methods, data-driven methods have advantages in terms of computing time. The computational time of LSTM and numerical history matching which is the part of source of the training set and test set in this study is shown in Table 6:

<table>
<thead>
<tr>
<th>Methods</th>
<th>Calculation Time with 100 Months</th>
<th>Prediction Time with 48 Months</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>4 min 26 s</td>
<td>3 s</td>
</tr>
<tr>
<td>Numerical simulation</td>
<td>12 min 11 s</td>
<td>3 min 24 s</td>
</tr>
</tbody>
</table>

It could be seen from the Table 6 that LSTM needs quite a period of time for model training and validation (4 min 26 s). However, once the model has been built, the prediction is tremendously fast which takes only few seconds. Although it seems like using machine learning method could be much faster than numerical simulation, numerical simulation still needs to be conducted for dataset before we begin the machine learning work and this advantage in computational time will be weakened. In other words, if the dataset is all from monitoring data rather than the numerical simulation, this advantage could completely manifest.
Besides computing time, prediction results of water saturation distribution are also necessary for further analysis. By analyzing the distribution of water saturation, the location of remaining oil in the reservoir and the future exploitation potential of the reservoir could be identified. The water saturation distribution of three main oil producing layers including 12th layer, 14th layer, 27th layer in the reservoir at 48th months of prediction (148th month in whole time sequence) are as shown in Figure 13:

![Figure 13](image_url)

**Figure 13.** Remaining oil distribution in different layers at 48th month of prediction (148th month in whole time sequence): (a) Water saturation distribution in 12th layer; (b) Water saturation distribution in 14th layer; (c) Water saturation distribution in 27th layer.

It could be seen from Figure 13 that the 12th and 14th layer of the reservoir has a higher water saturation level and a greater development degree. In contrast, there are still large areas in the 27th layer which have not been exploited thoroughly. In general, the reserves used in the block are not large, and each layer has a great amount of area to be developed as the red circle shown in Figure 13. The oil is greatly saturated near R-418, R-074 and R-025. Therefore, it is necessary to formulate an exploitation plan in these regions in future development. For example, in the regions inside the red circles, well pattern infilling construction could be carried out. It is reasonable to shorten the well spacing from 1000 m (at present) to 200 m. Besides, R-420 can be a good candidate for injection well because there exists a great amount of oil saturated near the R-420 according to Figure 13c.
3.4. Comparison Between Standard RNN, LSTM and GRU

To further study the performance of LSTM neural network, this paper selected the prediction results of three different machine learning methods for comparison, including standard Recurrent Neural Network (standard RNN), Gated Recurrent Unit (GRU) and LSTM. GRU originated in 2014 [51]. It is a variant of the LSTM recurrent network [52], and they have many similarities in internal structure. GRU combines the forget and input gates into a single “update gate” and merges the cell state and hidden state. This method can save computational resources and reduce the training difficulty. At the same time, its training effect is close to LSTM. Therefore, GRU is also gradually being widely used [53–55].

The AARD of prediction results for three different machine learning methods are shown in Figure 14:

![Figure 14. Comparison of average absolute relative deviation (AARD) results between standard Recurrent Neural Network (RNN), LSTM, Gated Recurrent Unit (GRU): (a) Comparison of AARD results between standard RNN, LSTM GRU in terms of pressure; (b) Comparison of AARD results between standard RNN, LSTM, GRU in terms of water saturation.](image)

It could be seen from Figure 14 that the prediction performances of different neural network methods are distinct. In general, LSTM has the best prediction performance, then the GRU, and standard RNN performs worst in this problem. Under the calculation condition of three hidden layers, the LSTM’s pressure AARD is 14.05%, and the water saturation AARD is 11.26% which are acceptable. The accuracy
of GRU is slightly lower than LSTM. In the condition of three hidden layers, the pressure and water saturation AARD of GRU prediction results are 15.32% and 18.23%, respectively. GRU could be another method choice beside LSTM in terms of this problem. In contrast, the standard RNN perform relatively badly. Its AARD of pressure and water saturation are both over 24%, which indicates that the standard RNN is not appropriate for processing the water saturation distribution prediction problem in this paper.

4. Conclusions

This study presented an alternative way for quick and robust prediction of water saturation distribution in a reservoir besides of numerical simulation. A reservoir water saturation distribution predictor was developed utilizing a Long Short-Term Memory recurrent neural network. An actual reservoir dataset deriving from monitoring and simulating was utilized for model training and test. Through the trained model, it is accessible to forecast water saturation distribution, pressure distribution and oil production faster than with numerical simulation. After analyzing the calculation results, several conclusions can be drawn as follows:

The LSTM method has a good performance on the water saturation prediction. In the case of large data volume, the overall AARD can be controlled below 14.82%. It is proven that the model is valid and reliable. On the basis of model validation, the model can predict water saturation distribution and present suggestions for tapping remaining oil in future production. Compared with numerical simulation, machine learning methods have great advantages in computation time which could be exploited in the future. Different types of machine learning methods have distinct performance in water saturation prediction. The accuracy of LSTM is better than GRU and standard RNN. GRU is another choice, although the accuracy is slightly lower than LSTM.

Author Contributions: Conceptualization, H.S. and Q.Z.; methodology, S.D.; software, Y.Z.; validation, Q.Z., Y.W.; formal analysis, S.D.; investigation, Q.Z.; resources, C.W.; writing—original draft preparation, Q.Z.; writing—review and editing, Q.Z.; supervision, H.S.; project administration, H.S.

Funding: This research was funded by the Fundamental Research Funds for the Central Universities of China (Grant No. FRF-TP-19-005B1) and National Natural Science Foundation of China (Grant No.51974357).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Yue, M.; Zhu, W.; Han, H.; Song, H.; Long, Y.; Lou, Y. Experimental research on remaining oil distribution and recovery performances after nano-micron polymer particles injection by direct visualization. Fuel 2018, 212, 506–514. [CrossRef]
3. Leverett, M. Capillary behavior in porous solids. Trans. AIME 1941, 142, 152–169. [CrossRef]


22. Masoudi, P.; Asgarinezhad, Y.; Tokhmechi, B. Feature selection for reservoir characterisation by Bayesian network. Arab. J. Geosci. 2015, 8, 3031–3043. [CrossRef]


© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).