A Multivariate Long Short-Term Memory Neural Network for Coalbed Methane Production Forecasting

Xijie Xu 1, Xiaoping Rui 2, Yonglei Fan 1, Tian Yu 1 and Yiwen Ju 3,*

1 College of Resources and Environment, University of Chinese Academy of Sciences, Beijing 100049, China; xuxijie18@mails.ucas.ac.cn (X.X.); fanyonglei17@mails.ucas.ac.cn (Y.F.); yutian17@mails.ucas.ac.cn (T.Y.)
2 School of Earth Sciences and Engineering, Hohai University, Nanjing 211000, China
3 Key Laboratory of Computational Geodynamics, College of Earth and Planetary Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
* Correspondence: ruixp@hhu.edu.cn (X.R.); juyw@ucas.ac.cn (Y.J.)

Received: 17 November 2020; Accepted: 6 December 2020; Published: 10 December 2020

Abstract: Owing to the importance of coalbed methane (CBM) as a source of energy, it is necessary to predict its future production. However, the production process of CBM is the result of the interaction of many factors, making it difficult to perform accurate simulations through mathematical models. We must therefore rely on the historical data of CBM production to understand its inherent features and predict its future performance. The objective of this paper is to establish a deep learning prediction method for coalbed methane production without considering complex geological factors. In this paper, we propose a multivariate long short-term memory neural network (M-LSTM NN) model to predict CBM production. We tested the performance of this model using the production data of CBM wells in the Panhe Demonstration Area in the Qinshui Basin of China. The production of different CBM wells has similar characteristics in time. We can use the symmetric similarity of the data to transfer the model to the production forecasting of different CBM wells. Our results demonstrate that the M-LSTM NN model, utilizing the historical yield data of CBM as well as other auxiliary information such as casing pressures, water production levels, and bottom hole temperatures (including the highest and lowest temperatures), can predict CBM production successfully while obtaining a mean absolute percentage error (MAPE) of 0.91%. This is an improvement when compared with the traditional LSTM NN model, which has an MAPE of 1.14%. In addition to this, we conducted multi-step predictions at a daily and monthly scale and obtained similar results. It should be noted that with an increase in time lag, the prediction performance became less accurate. At the daily level, the MAPE value increased from 0.24% to 2.09% over 10 successive days. The predictions on the monthly scale also saw an increase in the MAPE value from 2.68% to 5.95% over three months. This tendency suggests that long-term forecasts are more difficult than short-term ones, and more historical data are required to produce more accurate results.

Keywords: coalbed methane (CBM); production forecasting; long short-term memory neural networks (LSTM NN); multi-step predictions

1. Introduction

Coalbed methane (CBM) is an important unconventional clean energy source that has the potential to eventually replace natural gas. As such, it is attracting increasing attention from around the world. Accurate prediction of CBM production can not only forecast the economic benefits of CBM, but also provide advice for the establishment of mining plans, which play an important role in the production process of CBM. However, CBM is subject to a range of complex interactions relevant to the many factors involved in its creation [1]. These unique features have led to its classification...
as an unconventional gas resource [2]. Therefore, accurate production forecasts for CBM present certain challenges. At present, the prediction methods for CBM production mainly include the type curve and decline curve methods [3], numerical simulation methods [4], material balance methods [5], and machine learning methods, including neural networks [6] and support vector machines (SVMs) [7] among others.

The type curve and decline curve methods have been widely used for production analysis, achieving important results. Fetkovich [8] proposed the first-generation production type curve, which subsequent type curves have improved upon. Aminian et al. [2] proposed a series of curves that could predict CBM and water production, studying the effects of different parameters on these curves and discussing their applications and limitations, the most notable of which being the difficulty in representing a uniform model at the production cycle's different stages. Jang et al. [9] used the method of decline curve analysis and combined material balance and fluid state analysis to predict the production dynamics of CBM, subsequently establishing a comprehensive production data model. Decline curve analysis is currently the most commonly used and effective yield forecasting method, but it also has its limitations [10].

The numerical simulation method uses a complicated mathematical model. When using this method, it is necessary to obtain enough production data and measure geological parameters as accurately as possible [6]. These factors will have a very large impact on the production simulation. Therefore, if these influencing factors cannot be accurately obtained, the numerical simulation method will not be used. A detailed description of numerical simulation techniques can be found in [11]. Numerical simulation technology is increasingly used in unconventional gas reservoirs, and more and more geological factors are taken into consideration. Li et al. [12] used geological survey and experimental data to study the formation history of coalbed methane reservoirs and analyzed the role of various stages in the process of coalbed methane production. Zhou et al. [13] used numerical simulation technology to predict production and believed that the skin factor and coal shrinkage rate have important impacts on CBM production. Thararoop et al. [14] proposed a numerical simulation model that took into account the water in the coal matrix and the swelling and shrinkage of coal. In addition, numerical simulation software for coalbed methane reservoirs, whose development was based on the C++ programming language, was proposed in [15]. Numerical simulation technology requires sufficient geological data to provide support, but it is difficult to obtain these factors in actual production. Complex mathematical models also limit the application of this method.

Material balance is also an important method for estimating the reserves of coalbed methane, since this method can comprehensively consider the influence of many factors. In [5], two material balance methods were proposed, which were used to predict unconventional gas reservoirs and estimate the original gas, respectively. The difference between this method and the traditional material balance method is that the influence of adsorbed gas is taken into account. Shi et al. [16] established a material balance equation to estimate coalbed methane reserves, taking into account factors such as dissolved gas and free gas. Sun et al. [17] used a flow material balance equation, combined with the relationship between pressure and saturation, to analyze the production of low-permeability CBM wells and achieved good results. More and more material balance models that consider multiple factors have been developed. However, the actual production of coalbed methane is a dynamic process. The influencing factors are complex and diverse, and it is impossible to take all of them into consideration. Additionally, similar to the numerical simulation method, it is very difficult to obtain many factors, so the material balance equation is also restricted in the prediction of coalbed methane reserves.

The development of machine learning provides a new method for forecasting the production of CBM. Compared with the previous method, the advantage of machine learning is that it does not need to obtain the geological conditions of the coalbed methane reservoir and can make predictions only from the production data. For example, the back propagation (BP) neural network method can efficiently predict production without having to understand the conditions of coalbed methane reservoirs or deal with insufficient production data [6]. Xia et al. [18] achieved favorable results through their proposal of
a hybrid method to forecast CBM production capacity. This method takes both the rough set (RS) and least-square support vector machine models into consideration. Huang and Wang [7] optimized their own SVM using a genetic algorithm (GA), and their results showed that their GA-SVM model could also achieve high accuracy in CBM production capacity predictions. The machine learning method has the characteristics of being simple and convenient to implement, and it does not need to consider the complex factors in the actual process, so it is widely used in production forecasting. However, traditional machine learning methods, such as SVM, the Bayes method [19], multiple regression analysis [20,21], and neural networks do not consider time dependence when processing time series data. Therefore, the use of such methods to predict CBM production also has limitations.

The actual production process of CBM is very complicated, involving the interaction of geological factors and human factors. However, the production data of CBM has certain rules, which reflects the production process of CBM to a certain extent. As such, it is possible to predict future production by focusing on inherent laws and tendencies based on the available historical data rather than complex process research. This can be accomplished through long short-term memory neural networks (LSTM NNs). LSTM NNs are deep learning recurrent neural network structures with the ability to process long-term sequence data [22]. These networks can learn the inherent laws present in historical data through a time series without having to consider complex coal seam environments. LSTMs can also be applied to other fields such as environmental science, in which some scholars have applied LSTM models to predict PM$_{2.5}$ concentrations in air pollution, achieving valuable results [22–25].

In terms of public transportation, Chen et al., Tian et al., and Li and Cao [26–28] used the LSTM model to study traffic flow. In [29], LSTMs were used to forecast traffic speed in the Beijing area. Petersen et al. [30] used CNNs (Convolutional Neural Networks) and LSTMs to predict bus travel time. In the financial sector, Fischer and Krauss [31] and Kim and Won [32] applied the LSTM model to market forecasts and achieved superior results to those of random forests methods and deep neural networks (DNNs). Vochozka et al. [33] used the LSTM model to establish a method for predicting company bankruptcy, which provided a reference for the company’s future development. In the industrial and energy sectors, Wu et al. [34] used the LSTM network to estimate the remaining service life of an engineering system, while Peng et al. [35] used LSTMs and differential evolution to predict the price of electricity, and the prediction accuracy was superior to current models. Sagheer and Kotb [36] proposed a genetic algorithm-optimized deep LSTM method to predict oil production, which has proven to be more accurate than statistical and software calculation methods. Although the LSTM model has been widely used in research related to production and price prediction, it is rarely applied in the research of unconventional gas reservoirs. Xu et al. [37] used LSTM networks to predict the production of coalbed methane and achieved good results. However, this study did not consider the influence of multiple factors and only used coalbed methane production data.

In view of the complexity and limitations of the current methods, the objective of this article is to propose an artificial intelligence-based method for CBM production forecasting. This paper proposes the use of multivariate LSTM NNs as a prediction method for CBM production. Auxiliary data such as casing pressure, water production, and bottom hole temperatures are also inputted into the LSTM NN to improve prediction performance. This combined model was validated using production data from a CBM well; its results were compared with those obtained using a traditional LSTM NN model. The results demonstrated that auxiliary data can improve the prediction outcome. In addition, this paper proposes a multi-step prediction method more in line with the time process of CBM production, as forecasting performances will deteriorate as time lag increases.

2. Data and Methods

2.1. Data Description

The data used in the experiment were coalbed methane production data in the Panhe Demonstration Area in the Qinshui Basin of China, totaling 1945 days of recorded data. The Qinshui Basin was the
first CBM development base of high rank coal reservoirs in China. The daily log included the current CBM production, casing pressure, water production, and bottom hole temperature (including both highest and lowest temperatures). The data did not take into account geological factors such as rock formations, coal seam thickness, and burial depth, because these may not have been data that changed over time. The time series of these five variables is shown in Figure 1, where the abscissa represents the time and the ordinate represents the value recorded on that day. It can be easily noted that there exists a periodic correlation between the five aforementioned variables. As can be seen from the figure, the CBM production data of this well can be roughly divided into three phases: the drainage phase, the high yield phase, and the steady phase. The drainage phase was from 0 to 600 days, during which the CBM production was extremely unstable or very low. The high yield phase was from 600 to 1000 days, during which the continuous high yield of CBM was achieved, and the water yield was almost zero. After 1000 days was basically the steady phase of this well. The CBM production was basically stable and the CBM well almost no longer produced water. There are more complex divisions of the CBM production phases. Here, we simply divided the process into three phases for the purpose of subsequent prediction process analysis and cross-validation. In fact, in the early stage (0–600 days) of CBM production, due to the instability of the casing pressure and water production, the production of CBM was very unstable and difficult to predict. In addition, the actual production should also take into account the influence of human factors in the early stage, so we did not consider the early production stage forecast because it was very difficult. When the water production dropped to very little and remained stable, the production of CBM rose rapidly at this time and maintained a relatively stable state, showing a certain regularity.

Before a prediction could be made, the input data needed to be preprocessed. It was necessary that the input data be normalized, as the range of values for each variable was quite different. In this case, we utilized the min-max normalization method, as seen in Formula (1), to make all input variables take values between 0 and 1. It is important to note that the time series data were recorded daily with no missing values, thus making any data interpolation unnecessary. We could then convert
the normalized time series data into a supervised learning problem. Because of the phases of CBM production, it was necessary to have sufficient training data and cover each phase of the production process. To evaluate the accuracy of the model while also ensuring adequate orientation for the model, the first 1600 days of data were used as a training set, while the remaining 345 days were used as a test set:

$$ y_{\text{norm}} = \frac{y_i - \min(y_i)}{\max(y_i) - \min(y_i)} $$

(1)

where $y_{\text{norm}}$ is the normalized value, $\min(y_i)$ is the smallest value of the dataset, and $\max(y_i)$ is the maximum value of the dataset.

2.2. M-LSTM NN Model

In this study, the traditional LSTM NN was improved using multivariate inputs. The overall framework of the multivariate LSTM NN (i.e., multivariate long short-term memory (M-LSTM)) model is shown in Figure 2. LSTM NN is a kind of recurrent neural network (RNN), but it can solve the problem of gradient explosion well when traditional RNNs analyze long time sequences [38]. Details on RNNs and LSTM NNs are given in Appendices A and B.

In our prediction model, the input variables were divided into two parts (shown in the blue solid box in Figure 2): the first was CBM production data, and the second was auxiliary production record data (casing pressure, water production, and the bottom hole temperatures). Suppose we were to predict the CBM production at time $t$. The input data would be the historical data at time $t-1$. Again, note that the input data had been normalized and transformed into a supervised learning problem. Two or more hidden stacked LSTM layers are included in a green dashed box in Figure 2. The LSTM network structure was used to determine the inherent features from the historical input data to predict CBM production at time $t$. Both the number of LSTM layers and nodes in each layer were optimized through experimentation. We then used a fully connected dense layer to obtain the forecast result of the daily production of coalbed methane. After inputting the long-term series of production data into the LSTM layer, it took multiple iterations to obtain the corresponding features.
2.3. Evaluation Index

In order to optimize the model parameters and evaluate the model’s accuracy, three indicators were used in this paper: root mean square error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). The calculation formula of the three indexes is as follows:

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |P_i - A_i| 
\]

(2)

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|P_i - A_i|}{A_i}
\]

(3)

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (P_i - A_i)^2}
\]

(4)

where \(P_i\) represents the model predicted value, \(A_i\) represents the true value, and \(n\) is the number of days of testing.

2.4. Network Architecture

In order to achieve the predicted structure of CBM production shown in Figure 2, it was necessary to establish some initial parameters for the network, such as the number of LSTM layers, the number of nodes in each of these, the number of dense layers, the number of nodes in each of these, the learning rate, the batch size, and the gradient descent function among others. We explored the influence of each parameter on network performance under the condition that other parameters were fixed during this process. The optimum values were selected by the most favorable RMSE, MAE, and MAPE values. The initial parameters of the model were 2 LSTM hidden layers, 1 dense layer, and a batch size of 72. The computer specifications for the experiment were as follows. The CPU was an Intel(R) Core(TM) i7-7700; the RAM was 32.00 GB; and the GPU was an NVIDIA GeForce GTX 1060 3 GB.

In the process of model parameter selection and model accuracy validation, cross-validation was needed. However, when it came to the time series data, traditional methods, such as the leave-one-out method and K-fold cross-validation, were not applicable because of time dependence. For the time series, nested cross-validation was a very good choice. Therefore, rolling origin recalibration evaluation was adopted in this paper [39]. It is a method of nested cross-validation of a time series. The process of nested cross-validation of the time series is shown in Figure 3. In the figure, the blue boxes represent training data, the red boxes represent validation data, the yellow boxes represent test data, and the white boxes represent unused data. We split the data into 7 processes and averaged the results of the 7 processes to calculate the final error of the model. We used 100 days of data each time as the validation data and moved it to the training data of the next process in chronological order.

![Figure 3. The nested cross-validation process.](image-url)
We then examined the influence of the number of nodes in each hidden layer of the LSTM model. For simplification purposes, we selected the number of hidden layer nodes from an alternative set of \{32, 64, 128, 256, 512\}, which was determined from previous experience. The prediction performance (shown in Table 1) indicated that our M-LSTM model achieved optimal performance with 128 nodes. Increasing the number of nodes to 256 did not improve the prediction performance but did increase the training time of the network. Finally, when the number of nodes was increased to 512, the accuracy of the model decreased rapidly, as indicated by the MAE, MAPE, and RMSE.

### Table 1. The effects of different numbers of nodes in each LSTM hidden layer.

<table>
<thead>
<tr>
<th>Nodes</th>
<th>RMSE (m³)</th>
<th>MAE (m³)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>32</td>
<td>81.84</td>
<td>50.11</td>
<td>1.50</td>
</tr>
<tr>
<td>64</td>
<td>79.30</td>
<td>46.85</td>
<td>1.41</td>
</tr>
<tr>
<td>128</td>
<td>75.77</td>
<td>41.41</td>
<td>1.24</td>
</tr>
<tr>
<td>256</td>
<td>80.81</td>
<td>50.02</td>
<td>1.53</td>
</tr>
<tr>
<td>512</td>
<td>84.46</td>
<td>52.62</td>
<td>1.58</td>
</tr>
</tbody>
</table>

Having fixed the number of hidden layer nodes to 128, we next investigated the effects of different learning rates. The learning rate was also selected from an alternative set of \{0.005, 0.001, 0.0005, 0.0001, 0.00005\}. The results (shown in Table 2) demonstrated that the training results of the larger learning rate were poor, which was because the larger learning rate caused the model to miss the optimal parameters. As the learning rate gradually decreased, the accuracy of the model was significantly improved. When the learning rate continued to decrease from 0.0001, the accuracy was not significantly improved. Therefore, considering efficiency and accuracy, the learning rate was set to 0.0001. This was due to the value of the learning rate being crucial to the training of the model; in other words; having too large a learning rate led to shock in the training process, and having too small a learning rate made it difficult for the model to reach convergence.

### Table 2. The effects of different learning rates.

<table>
<thead>
<tr>
<th>Learning Rate</th>
<th>RMSE (m³)</th>
<th>MAE (m³)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.005</td>
<td>142.42</td>
<td>121.87</td>
<td>3.69</td>
</tr>
<tr>
<td>0.001</td>
<td>95.37</td>
<td>72.60</td>
<td>2.22</td>
</tr>
<tr>
<td>0.0005</td>
<td>82.27</td>
<td>54.08</td>
<td>1.66</td>
</tr>
<tr>
<td>0.0001</td>
<td>76.24</td>
<td>44.68</td>
<td>1.35</td>
</tr>
<tr>
<td>0.00005</td>
<td>76.59</td>
<td>43.18</td>
<td>1.30</td>
</tr>
</tbody>
</table>

### 3. Results and Discussion

#### 3.1. Prediction Performance

Once the initial network parameters were fixed, the M-LSTM NN model was instructed with the aforementioned training set until convergence was reached. The accuracy of the network was then evaluated using the test set, and the resulting predicted and actual CBM production values are shown in Figure 4. It is shown that the predicted production values were generally consistent with the actual values, as most of the points were near the \(y = x\) line. Obtaining an \(R^2\) value between the predicted and actual CBM production indicated that 94.4% of the variance was successfully described by the M-LSTM NN model.
3.2. Comparison with Traditional LSTM NN Model (without Multivariate Inputs)

To verify the importance of the auxiliary data (casing pressure, water production, and bottom hole temperatures) for CBM production prediction, we established an LSTM NN model without these inputs and compared its accuracy with that of the M-LSTM model. At the same time, and in order to avoid accidental errors in the experiment, after the optimal network structure was trained, 30 experiments were conducted on the LSTM NN and M-LSTM NN models, respectively, and the error distribution of both models was statistically analyzed. Thirty independent repeated experiments were sufficient to obtain a good error distribution. The boxplot distribution of the RMSE, MAE, and MAPE values is shown in Figure 5. The results demonstrate that the test errors obtained by the M-LSTM NN model were significantly smaller than those of the traditional LSTM NN model. Table 3 shows the average error over 30 experiments. Compared with the predicted results of the LSTM NN model (MAPE = 1.14%), the predicted results of the M-LSTM NN model (MAPE = 0.91%) were closer to the actual value. Thus, it can be inferred that using additional production data (casing pressure, water production, and bottom hole temperatures) as auxiliary inputs can significantly improve the prediction accuracy of CBM production.

We further explored the influence of each variable on the prediction results. Only one auxiliary variable was input into the M-LSTM network at a time to predict the production of coalbed methane. The results obtained are shown in Table 4. It can be found that casing pressure and water production as auxiliary variables could improve prediction accuracy, of which water production improved it the most (MAPE = 1.01%). However, the input of temperature variables had almost no effect on the prediction results.
Figure 5. Comparison of error distribution between the M-LSTM NN model and the traditional LSTM NN model. (a) RMSE, (b) MAE and (c) MAPE.

Table 3. Average error over 30 experiments.

<table>
<thead>
<tr>
<th>Model</th>
<th>RMSE (m³)</th>
<th>MAE (m³)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM NN</td>
<td>46.47</td>
<td>31.46</td>
<td>1.14</td>
</tr>
<tr>
<td>M-LSTM NN</td>
<td>42.46</td>
<td>24.98</td>
<td>0.91</td>
</tr>
</tbody>
</table>

Table 4. The influence of each variable on the prediction results.

<table>
<thead>
<tr>
<th>Variable</th>
<th>RMSE (m³)</th>
<th>MAE (m³)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Casing pressure</td>
<td>44.15</td>
<td>31.22</td>
<td>1.10</td>
</tr>
<tr>
<td>Water production</td>
<td>42.54</td>
<td>28.58</td>
<td>1.01</td>
</tr>
<tr>
<td>Lowest temperature</td>
<td>46.64</td>
<td>31.45</td>
<td>1.14</td>
</tr>
<tr>
<td>Highest temperature</td>
<td>46.63</td>
<td>31.08</td>
<td>1.15</td>
</tr>
<tr>
<td>No auxiliary variables</td>
<td>46.47</td>
<td>31.46</td>
<td>1.14</td>
</tr>
<tr>
<td>All variables</td>
<td>42.46</td>
<td>24.98</td>
<td>0.91</td>
</tr>
</tbody>
</table>

3.3. Multi-Step Predictions

Following the development of our M-LSTM model to predict daily CBM production, we continued our study through a multi-step forecast model. In the multi-step models, each time step prediction was compared with the actual production data for the next day. This actual data, taken from the test set, was then made available to the model for the next time step prediction. As an example, Figure 6 shows how a two-step model predicted the production over three days. In Step 1 (that is, from time \( t \)), we must forecast the time \( t + 1 \), \( t + 2 \), and \( t + 3 \). In Step 2, from time \( t + 1 \), we must forecast the time \( t + 2 \), \( t + 3 \), and \( t + 4 \), and so on. This type of model mimics real-world CBM production scenarios, as new CBM production data is obtained each day and used to predict future production.
Figure 6. A multi-step predictions model.

In our model, the daily production of CBM was predicted in five time steps, spanning a time period of ten days for each step. The network structure in our experiment included two LSTM layers and a fully connected dense layer. As shown in Table 5, the accuracy of the five initial time steps was noted for each of the next ten days. These results demonstrated that with an increase in each predicted time interval, the performance of the model grew worse, with the MAPE value increasing from 0.24% to 2.09% from the first to the tenth day. It can be inferred from these results that long-term predictions are more difficult than short-term predictions, requiring more historical data to establish sufficient training for the model.

Table 5. Prediction accuracy of multi-step predictions for CBM daily production.

<table>
<thead>
<tr>
<th>Time Lag (Day)</th>
<th>RMSE (m³)</th>
<th>MAE (m³)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t + 1</td>
<td>11.55</td>
<td>6.42</td>
<td>0.24</td>
</tr>
<tr>
<td>t + 2</td>
<td>17.39</td>
<td>12.82</td>
<td>0.48</td>
</tr>
<tr>
<td>t + 3</td>
<td>22.50</td>
<td>19.18</td>
<td>0.71</td>
</tr>
<tr>
<td>t + 4</td>
<td>23.77</td>
<td>20.66</td>
<td>0.77</td>
</tr>
<tr>
<td>t + 5</td>
<td>25.00</td>
<td>22.06</td>
<td>0.82</td>
</tr>
<tr>
<td>t + 6</td>
<td>26.16</td>
<td>23.37</td>
<td>0.87</td>
</tr>
<tr>
<td>t + 7</td>
<td>27.28</td>
<td>24.61</td>
<td>0.92</td>
</tr>
<tr>
<td>t + 8</td>
<td>38.50</td>
<td>35.39</td>
<td>1.31</td>
</tr>
<tr>
<td>t + 9</td>
<td>47.57</td>
<td>46.09</td>
<td>1.71</td>
</tr>
<tr>
<td>t + 10</td>
<td>59.51</td>
<td>56.74</td>
<td>2.09</td>
</tr>
</tbody>
</table>

In the actual production process of CBM, it is often more common to focus on the monthly production of CBM rather than production on any given day. With this in mind, we scaled the CBM production data and converted daily production into monthly production. After this scaling, we were able to obtain 65 months of production data. We made five time step forecasts for the previous eight months, with predictions for the next three months in each step. The multi-step predictions’ results of monthly coalbed methane production are shown in Table 6, where we can see that comparatively similar results were obtained from the multi-step prediction of daily output. The MAPE values for this iteration of the model increased from 2.68% to 5.95%. As the time interval increased again, the predicted production values began to deviate further from the actual values. Although the performance of the model worsened with an increase in each time interval, the error (MAPE = 5.95%) still fell into an acceptable range.

Table 6. Prediction accuracy of multi-step predictions for CBM monthly production.

<table>
<thead>
<tr>
<th>Time Lag (Month)</th>
<th>RMSE (m³)</th>
<th>MAE (m³)</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>t + 1</td>
<td>2410</td>
<td>2230</td>
<td>2.68</td>
</tr>
<tr>
<td>t + 2</td>
<td>3380</td>
<td>2980</td>
<td>3.65</td>
</tr>
<tr>
<td>t + 3</td>
<td>5140</td>
<td>4830</td>
<td>5.95</td>
</tr>
</tbody>
</table>
4. Conclusions

A multivariate long short-term memory neural network (M-LSTM NN) model for coalbed methane production forecasting was proposed in this paper. Due to the optimization of the parameters presented in the model (including the number of LSTM NN hidden layer nodes, learning rate, and others), as well as the subsequent results, we were able to reach the following conclusions from our study:

1. The M-LSTM NN model we developed was able to achieve favorable results in predicting CBM production. The results show that the use of deep neural networks to predict CBM production can achieve good results without considering complex geological factors and by only using historical production data. The M-LSTM network provides a fast artificial intelligence prediction method for CBM production.

2. Thirty independent and repeated experiments were conducted, comparing the results of LSTM NN models without additional auxiliary inputs and our own M-LSTM NN model, with MAE, MAPE, and RMSE values indicating that the M-LSTM NN model achieved better results than LSTM NN model. In addition, we analyzed the impact of each variable on the results and found that water production and casing pressure can improve the accuracy of the prediction, while inputting the temperature into the M-LSTM network did not improve the results. This shows that the bottom hole temperature has almost no effect on the production of CBM in the actual production process of CBM, and inputting it into the M-LSTM network cannot improve prediction accuracy. Since we have only obtained very little auxiliary production information, we cannot analyze the factors that have the highest impact on CBM production forecasts. Therefore, in future research, it is necessary to select those variables that are highly correlated with CBM production.

3. A multi-step predictions model was also developed that was more consistent with the actual production processes of CBM, utilizing historical as well as current data to predict future CBM production. During our experimenting process, it was found that prediction accuracy was inversely related to the increase in time lag, regardless of whether the CBM production in question was at daily or monthly intervals. This finding suggests that to successfully predict term CBM production, more historical data may be needed to calibrate and train future models.

Author Contributions: X.X. established the prediction model, reached the conclusion, and wrote the main part of the paper. X.R. proposed a research framework for the study. Y.F. and T.Y. modified the introduction and checked the grammar of the article. Finally, Y.J. provided data processing methods and provided suggestions on the structure of the article. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (Grant No. 2019YFC1804304), the National Science and Technology Major Project of China (Grant No. 2017ZX05064005; 2016ZX05066006), the Fundamental Research Funds for the Central Universities (Grant No. 2019B02514), the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDA05030100), and the National Natural Science Foundation of China (Grant No. 41771478).

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

Appendix A. RNN

RNNs are commonly used neural networks for dealing with time series problems. Different from general neural networks, RNNs have memory units so they can capture time series information. Figure A1 shows the network structure of RNNs. $x_t$, $O_t$, and $S_t$ represent the input, memory, and output, respectively, at time $t$. $W$, $U$, and $V$ represent the weights between the input layer, hidden layer, and output layer, respectively. Because the weights between different layers of RNNs are shared, the network parameters can be greatly reduced. RNNs may cause gradient disappearance or gradient explosion when calculating connections between nodes with a long time interval [35]. Therefore, RNNs can only deal with short-term problems and cannot solve long-term dependence problems. CBM production is related over a long period of time, so RNNs cannot be used directly to predict it.
Appendix B. LSTM NN

LSTM NNs can overcome the problem of gradient disappearance in RNNs in the back propagation process [38]. Based on RNNs, the LSTM method improves the memory unit of the hidden layer. The memory unit of an LSTM NN is shown in Figure A2.

LSTM NNs use three gates to solve the problem that RNNs cannot handle long-term sequences: input gates, forget gates, and output gates. The input, output and memory state of an LSTM NN at time $t$ are $x_t$, $h_t$, and $C_t$ respectively. The memory state $C_t$ is defined as follows:

$$c_t = f_t \cdot C_{t-1} + i_t \cdot \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c)$$ (A1)

where $f_t$ is the forget gate, $i_t$ is the input gate, and $W$ and $b$ are the corresponding weights and offsets.

The function of the input gate is to control the input of data. The calculation formula for the input gate is as follows:

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ci}C_{t-1} + b_i)$$ (A2)

where $\sigma$ is the sigmoid function.

The function of the forget gate is to decide which information to keep and discard the unnecessary information. The formula for the forget gate is as follows:

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}C_{t-1} + b_f)$$ (A3)

The function of the output gate is to control the output information. It can be calculated by the following formula:

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}C_t + b_o)$$ (A4)

The definition of $h_t$ is as follows:

$$h_t = o_t \tanh(c_t)$$ (A5)

where $\tanh$ is the hyperbolic tangent.
Figure A2. An LSTM network hidden layer cell structure.

References

19. Zhao, C.; Li, J. Equilibrium Selection under the Bayes-Based Strategy Updating Rules. *Symmetry* 2020, 12, 739. [CrossRef]
27. Tian, Y.; Zhang, K.; Li, J.; Lin, X.; Yang, B. LSTM-based traffic flow prediction with missing data. *Neurocomputing* 2018, 318, 297–305. [CrossRef]
33. Vochozka, M.; Vrbka, J.; Suler, P. Bankruptcy or Success? The Effective Prediction of a Company’s Financial Development Using LSTM. *Sustainability* 2020, 12, 7529. [CrossRef]
37. Xu, X.X.; Rui, X.P.; Fan, Y.L.; Yu, T.; Ju, Y.W. Forecasting of Coalbed Methane Daily Production Based on T-LSTM Neural Networks. *Symmetry* 2020, 12, 861. [CrossRef]


**Publisher’s Note**: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 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/).