

Article

# A Hybrid Oil Production Prediction Model Based on Artificial Intelligence Technology

Xiangming Kong \*, Yuetian Liu \*, Liang Xue, Guanlin Li and Dongdong Zhu

State Key Laboratory of Petroleum Resources and Prospecting, China University of Petroleum (Beijing), Beijing 102249, China

\* Correspondence: kongxm@student.cup.edu.cn (X.K.); lyt51x@163.com (Y.L.)

**Abstract:** Oil production prediction plays a significant role in designing programs for hydrocarbon reservoir development, adjusting production operations and making decisions. The prediction accuracy of oil production based on single methods is limited since more and more unconventional reservoirs are being exploited. Artificial intelligence technology and data decomposition are widely implemented in multi-step forecasting strategies. In this study, a hybrid prediction model was proposed based on two-stage decomposition, sample entropy reconstruction and long short-term memory neural network (LSTM) forecasts. The original oil production data were decomposed into several intrinsic mode functions (IMFs) by complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN); then these IMFs with different sample entropy (SE) values were reconstructed based on subsequence reconstruction rules that determine the appropriate reconstruction numbers and modes. Following that, the highest-frequency reconstructed IMF was preferred to be decomposed again by variational mode decomposition (VMD), and subsequences of the secondary decomposition and the remaining reconstructed IMFs were fed into the corresponding LSTM predictors based on a hybrid architecture for forecasting. Finally, the prediction values of each subseries were integrated to achieve the result. The proposed model makes predictions for the well production rate of the JinLong volcanic reservoir, and comparative experiments show that it has higher forecasting accuracy than other methods, making it recognized as a potential approach for evaluating reservoirs and guiding oilfield management.



**Citation:** Kong, X.; Liu, Y.; Xue, L.; Li, G.; Zhu, D. A Hybrid Oil Production Prediction Model Based on Artificial Intelligence Technology. *Energies* **2023**, *16*, 1027. <https://doi.org/10.3390/en16031027>

Academic Editors: Daoyi Zhu, Yiqun Zhang, Xiankang Xin, Shuda Zhao and Hongbin Yang

Received: 17 December 2022

Revised: 13 January 2023

Accepted: 14 January 2023

Published: 17 January 2023



**Copyright:** © 2023 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 (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** two-stage decomposition; sample entropy; hybrid model; time series forecasting; oil production forecast

## 1. Introduction

Well production is one of the most important indicators of oilfield development and management. Acquainting well production performance in advance can help engineers adapt development countermeasures and optimize development effects timely. Decline curve analysis has been widely utilized and achieves a good performance in conventional reservoirs [1]. The Arps model, however, may not be suitable due to the intricacy of flow dynamics in unconventional reservoirs. Under certain assumptions, the formation parameters are simplified, and the analytical or semi-analytical model is proposed and solved [2,3], which can simplify the complex formation seepage issue, but also limit the model's application. Numerical reservoir simulation techniques make production forecasts based on history matching by building a geological model of the actual reservoir [4]; however, establishing a model that is virtually identical to the actual reservoir requires reservoir engineers to have considerable experience. The complex geological characteristics of unconventional reservoirs, on the other hand, exacerbate the non-linear variance of oil production over time, making production prediction extremely challenging.

Various artificial intelligence algorithms have been implemented in the field of petroleum engineering with the growth of machine learning theory, paving a new route for the investigation of the production prediction issue [5,6]. Wang constructed a deep neural network

(DNN) model to forecast cumulative oil production of Bakken shale reservoirs [7]. On the basis of a long short-term memory (LSTM) structure, Huang conducted the development prediction task in a water-flooding reservoir [8]. Sagheer and Kotb established a deep long-short term memory (DLSTM) framework to enhance the oil production forecast performance and employed the genetic algorithm to optimize the hyperparameters [9]. Cheng used the long short-term memory (LSTM) network and gated recurrent unit (GRU) method to predict the oil production of actual oilfields in China and India [10]; the results indicate that LSTM and GRU have respective advantages under different circumstances.

Whereas single models are not sufficiently applicable for complicated issues, hybrid structures have become a research trend in the time series forecasting area [11,12], including well performance forecasting [13]. Fan developed a hybrid model that incorporated the autoregressive integrated moving average (ARIMA) with the long short-term memory (LSTM) network to predict the production of three actual wells under the influence of manual operations [14]. Li [15] used the PSO algorithm to optimize the proposed CNN-LSTM production forecast model, which has higher prediction accuracy than a single model. To enhance prediction validity, a current trend in time series forecasting is to incorporate the artificial intelligence algorithm with decomposition pre-processing strategies [16–18]. Liu proposed a hybrid model that combines ensemble empirical mode decomposition (EEMD) with an LSTM network [19], with the appropriate intrinsic mode functions (IMFs) of EEMD chosen using dynamic time warping (DTW). The method achieved a higher accuracy than other models in two reservoirs. Wang constructed a hybrid method with variational mode decomposition (VMD) and gated recurrent unit (GRU) [20], which was implemented in the Tahe oilfield and demonstrated an outstanding performance.

Oil production data of unconventional reservoirs are complicated and nonstationary; even a decomposition-forecasting method could not obtain excellent accuracy, so further data processing [21] is necessary. By considering both improving the prediction performance and reducing the accumulation errors, the secondary decomposition method and subsequence reconstruction approach are implemented comprehensively in this study. By analyzing the prediction performance of various reconstruction numbers and modes, guidelines for component reconstruction are generated, and the threshold is set for sample entropy values of first-stage decomposition IMFs. After evaluating the forecast efficiency, the most complex subsequence is further decomposed by VMD. As the core part of the structure, an optimum predictor is selected among four artificial intelligence algorithms to utilize their excellent data-learning skills. Then, the multi-type subseries data is inputted to the corresponding LSTM predictor based on a hybrid architecture. The multi-stage prediction structure is proposed and applied to forecast the well production rate of the Jinlong volcanic reservoir. The contributions of the proposed method are as follows:

- (1) A novel multi-step decomposition-integration framework is established for oil production forecasting;
- (2) Intrinsic mode functions (IMFs) are reconstructed to re-IMFs according to the rules for subsequence reconstruction numbers and modes, which reduces accumulation errors and calculation complexity;
- (3) The highest-frequency reconstructed IMF is preferred to be further decomposed to enhance prediction accuracy;
- (4) The hybrid model combines the advantages of both integral and corresponding architectures, maintaining both prediction accuracy and computing efficiency.

## 2. Methods

### 2.1. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise

Empirical Mode Decomposition (EMD) has been popularly utilized for decomposing sequential data in several time series forecasting fields. It also has disadvantages, such as weak stability. Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) is presented as an improved practice of EMD [22]. The raw time series data can also be decomposed into several intrinsic mode functions (IMFs) and a residue with

different frequencies by the CEEMDAN, which incorporates adaptive noise into the EMD process, and the decomposition process is complete so the least reconstruction error is obtained, helping resolve the issue of modal aliasing and residual noise in the sequence.

### 2.2. Variational Mode Decomposition

Variational Mode Decomposition (VMD) is a novel non-recursive data decomposition algorithm defined by Dragomiretskiy and Zosso [23] to solve the limitations of sensitivity to noise and sampling, which can decompose nonlinear and nonstationary original data into specific amounts of intrinsic mode functions (IMFs). The VMD method searches for the optimal solution to a variational problem to accomplish adaptive decomposition. The promotion points of VMD include minimizing the sum of evaluated bandwidth and inhibiting noise. It has been utilized to further decompose high-frequency subseries data from previous data processing, which can effectively decrease its complexity.

### 2.3. Long Short-Term Memory Network

Recurrent neural network (RNN) is widely utilized in Natural Language Processing (NLP) and Time Series Forecasting (TSF) areas; because of the gradient disappearance and explosion problem, improved methods have been proposed, especially the long short-term memory neural network (LSTM) which demonstrates excellent performance in dealing with many issues [24]. Figure 1 depicts the cell structure of the LSTM, which consists of the cell state, forget gate, input gate and output gate. As the core section of LSTM structure, a cell state contains information about all previous states, and at each new time step, operations are carried out to identify which old information to discard and which new information to add.

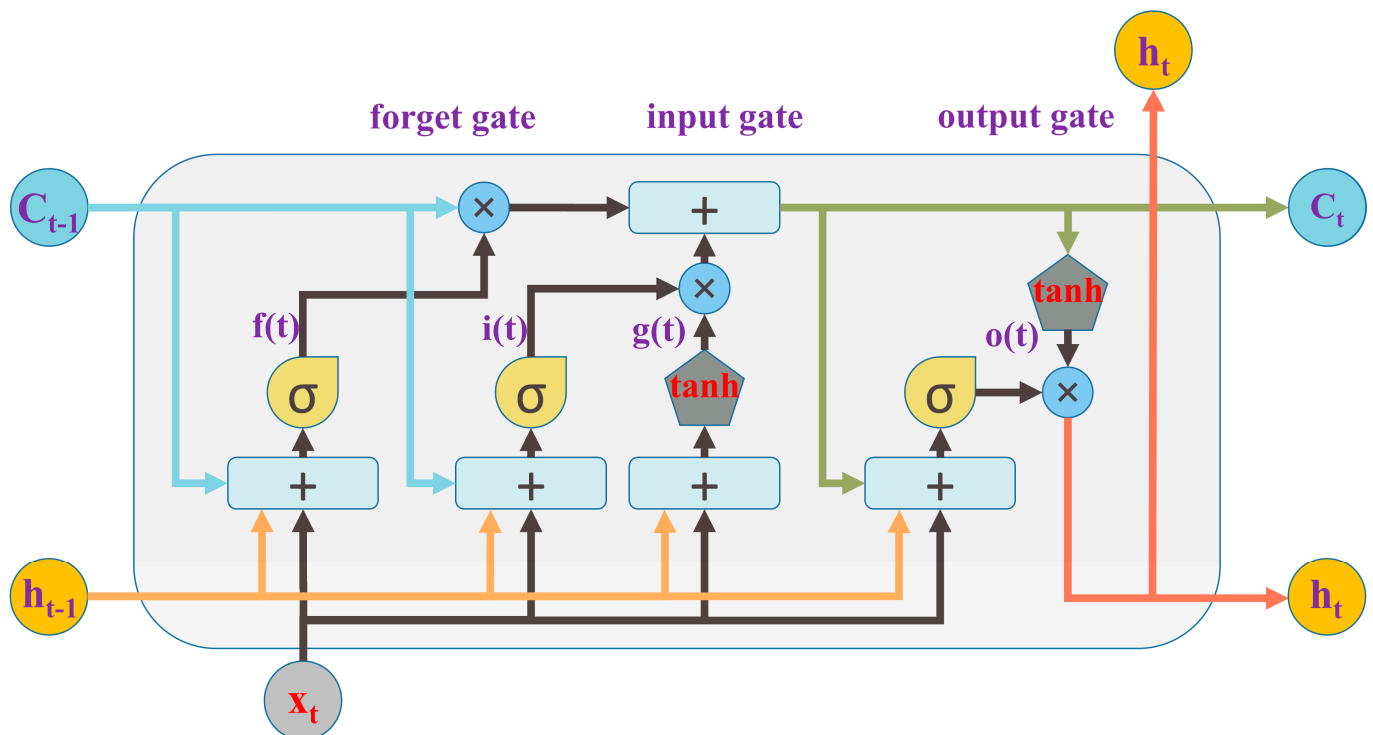


Figure 1. Structure of the LSTM.

#### 2.4. Sample Entropy

Sample entropy (SE), proposed by Richman and Moorman [25], can be employed to evaluate the complexity of time series; the higher the sample entropy value, the more complicated the sequence is. Although the series data is decomposed, IMFs still have several high-frequency subsequences. Based on the rules of reconstruction numbers and modes we defined in Section 2.5, IMFs decomposed from different source data could be integrated in a general flow, which can properly decrease computing workload and promote model efficiency.

#### 2.5. Rules of Subsequence Reconstruction and Secondary Decomposition

To reduce the accumulation errors, simplify the complexity of computation, and further process high-complexity components, this research implements a procedure after initial decomposition with the production data of actual oil wells: reconstruction and secondary decomposition. We determine the optimum number of reconstruction subsequences and prefer the most appropriate mode for the method's wide adaptability and migration. Details of the rules' definition process for this flow are depicted by comparable experiments in Section 3.2. The conclusions of the experiments are summarized as follows:

- (1) The proper number of reconstructed IMFs is set to three according to the prediction performance comparison of multiple hybrid models with various reconstructed IMF counts;
- (2) Based on the first decomposition process results of oil well production data in the JinLong volcanic reservoir, the corresponding most appropriate reconstruction modes are preferred. The optimum reconstruction modes show the threshold of the sample entropy value of IMFs to conduct the integration. The high-frequency subseries whose sample entropy values are over 1.0 and the low-complexity IMFs with values under 0.2 should be reconstructed, while the rest of the sequences comprise a re-IMF, regardless of whether the initial decomposition's component number is 8 or 9.
- (3) To consider both improving the prediction accuracy and simplifying the complexity of calculation, the secondary decomposition is applied to process only the highest frequency subsequence among the three reconstructed IMFs.

#### 2.6. Architecture of the Proposed Hybrid Model

The architecture of the hybrid model proposed in this paper is depicted in Figure 2. The main progress can be described as follows:

Step 1 Collect the actual oil production data.

Step 2 Decompose time series data into several IMFs by CEEMDAN.

Step 3 Calculate the sample entropy values of all IMFs and reconstruct them into fewer re-IMFs based on the rules for component reconstruction numbers and modes.

Step 4 Decompose the highest-frequency re-IMF0 from Step 3 by VMD to obtain new subsequences, and feed them into an integral LSTM architecture in the form of a matrix for prediction.

Step 5 Build the same number of LSTM models as reconstructed IMFs without secondary decomposition, input each IMF vectors and forecast them correspondingly.

Step 6 Integrate all the forecasting values of each re-IMF from Step 4 and Step 5 to obtain the final prediction result and evaluate it.

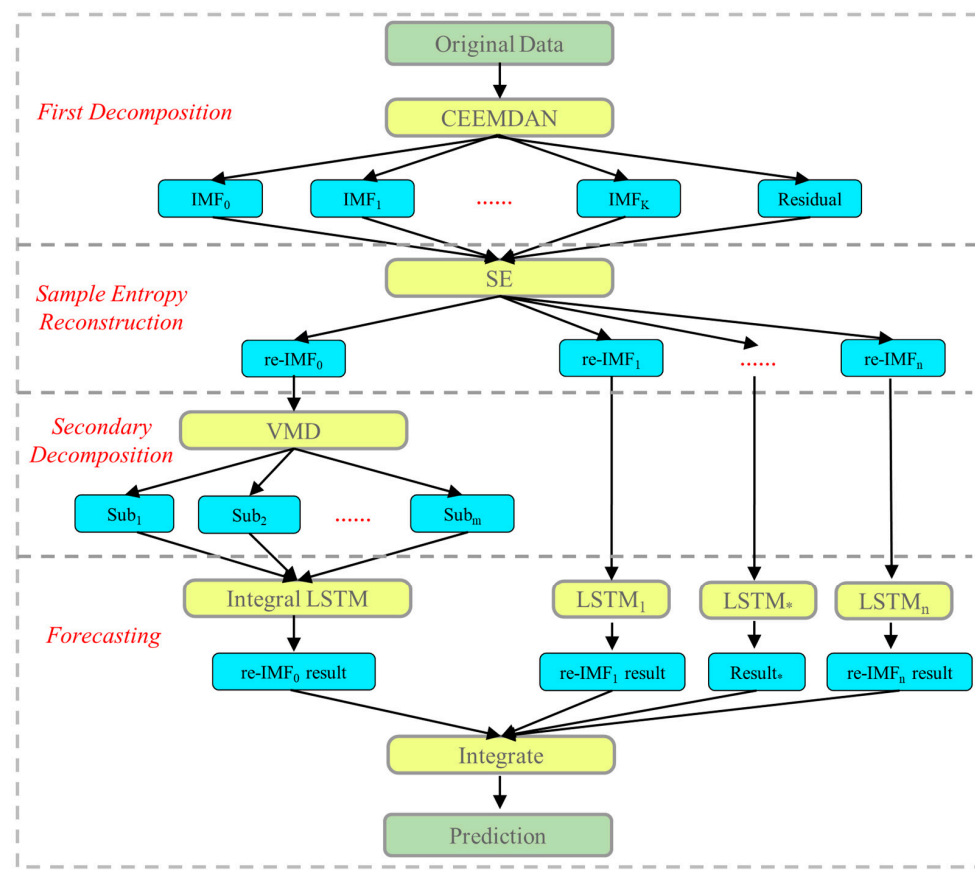


Figure 2. Framework of the proposed hybrid model.

### 2.7. Model Evaluation Index

For evaluating the forecasting performance of every model, this study selects the following four common performance measurement indices: root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and determination coefficient ( $R^2$ ). The calculation formulas for the indices are shown as follows:

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

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \frac{|y_i - \hat{y}_i|}{y_i} \times 100 \quad (3)$$

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4)$$

where  $y_i$ ,  $\hat{y}_i$ ,  $\bar{y}_i$  are the actual, forecast and mean values of the time series sample data, respectively,  $n$  is the sample size. The values of the four indexes represent the prediction accuracy of the proposed hybrid model; if the values of RMSE, MAE, MAPE are closer to 0, or  $R^2$  is closer to 1, the model is more excellent.

### 3. Experiments

Aiming to validate the established hybrid forecasting structure, oil production data from the JinLong volcanic reservoir are processed, and seven comparative experiments are conducted.

#### 3.1. Data Preparation

The hybrid model proposed above is implemented to forecast actual oil production in the JinLong (JL) volcanic reservoir, which is located on the east slope of the Zhongguai Uplift in the southwest margin of the Junggar Basin, as shown in Figure 3. The Jiamuhe Formation of the Permian is the primary oil-bearing layer to develop, which has an average depth of 4000 m. The thickness of the volcanic rock of the Jiamuhe Formation ranges from 22 m to 286 m, averaging 145.3 m. As a naturally fractured reservoir, oblique fractures, straight split fractures, reticular fractures, microfractures and partially filled fractures develop in the JL volcanic reservoir. The permeability of the formation varies from 0.01 to 68 mD and the average value is 0.56 mD; the porosity ranges from 8% to 22.3% with an average of 12.35%. Oil production data of the production wells in the JL volcanic reservoir are applied in multiple comparative experiments from Sections 3.3–3.8. After processing zero values, missing and abnormal values, the volume of oil production data for each single well varies from 916 to 1309. For the purpose of a high-quality modeling effect, the last 100 samples of each series of data are chosen as test sets; the rest is for training.



Figure 3. Location of the research area.

### 3.2. Sample Entropy Reconstruction and Secondary Decomposition

After the CEEMDAN process, oil production data could be decomposed into several IMFs, including a residual series. There are two kinds of first-decomposition results determined by the source data in this study: 8 and 9. The sample entropy values of each subsequence are calculated and listed in Table 1. In order to decrease the computational workload on the premise of ensuring prediction accuracy, we integrate the first-stage IMFs into fewer ones and implement secondary decomposition. The principles of reconstruction and secondary decomposition are defined by the following trials:

**Table 1.** First-stage decomposition results and sample entropy values of each subsequence.

Subsequence	Sample Entropy Value			
	Well-1	Well-2	Well-3	Well-4
IMF0	1.6896	1.6283	1.1927	1.3437
IMF1	1.3168	1.8620	1.4169	1.3095
IMF2	0.7788	1.2472	1.0695	0.6152
IMF3	0.5569	0.6540	0.6434	0.5737
IMF4	0.3745	0.2106	0.2698	0.4166
IMF5	0.1186	0.1255	0.1168	0.1666
IMF6	0.0442	0.0334	0.0572	0.0835
IMF7	0.0040	0.0277	0.0260	0.0084
IMF8	\	0.0022	0.0001	\

- (1) Determine the proper number of reconstruction IMFs.

To optimize the best reconstruction number, comparison experiments with different re-IMF quantities are carried out. All the experiments are based on the decomposition-reconstruction-integral LSTM structure. Table 2 shows the specific programs and evaluation index of these models' forecasting performance.

**Table 2.** Comparison of prediction models with different reconstructed IMF numbers.

Number of Re-IMFs	Well-1		Well-2		Well-3		Well-4	
	R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE
1 (no decomposition)	0.8030	3.4270	0.8044	3.1330	0.7658	0.8804	0.7955	4.3385
2	0.8035	3.1414	0.8064	2.5684	0.7801	0.8532	0.8119	1.9584
3	0.8529	2.2888	0.8656	2.1305	0.8549	0.6930	0.8603	1.3145
4	0.8135	2.7132	0.8064	2.4053	0.8134	0.7404	0.7853	2.4741
8 or 9 (no reconstruction)	0.8229	2.5888	0.8346	2.7063	0.7697	0.8731	0.8146	2.3529

When the number of re-IMFs is three, the prediction errors are the lowest and the R<sup>2</sup> scores are the highest. Therefore, in this research, the number of reconstruction IMFs is defined as three to obtain the best prediction accuracy.

- (2) Identify the optimum reconstruction modes.

To obtain three reconstruction subsequences, several optional integration modes of IMFs are listed in Tables 3 and 4. The forecasting results of these modes are compared with the same predictor, and the evaluation results are also shown in Tables 3 and 4.

**Table 3.** Reconstruction modes (first-stage decomposition IMF's number: 8).

Reconstruction Mode	Component of Re-IMF0	Component of Re-IMF1	Component of Re-IMF2	Well-1		Well-4	
				R <sup>2</sup>	RMSE	R <sup>2</sup>	RMSE
A	IMF0, IMF1	IMF2, IMF3, IMF4	IMF5, IMF6, IMF7	0.8529	2.2888	0.8603	1.3145
B	IMF0, IMF1, IMF2	IMF3, IMF4	IMF5, IMF6, IMF7	0.8056	3.1018	0.8167	1.8762
C	IMF0, IMF1	IMF2, IMF3	IMF4, IMF5, IMF6, IMF7	0.8115	2.6132	0.8186	1.6709

**Table 4.** Reconstruction modes (first-stage decomposition IMF's number: 9).

Reconstruction Mode	Component of Re-IMF0	Component of Re-IMF1	Component of Re-IMF2	Well-2		Well-3	
				$R^2$	RMSE	$R^2$	RMSE
A	IMF0, IMF1, IMF2	IMF3, IMF4, IMF5	IMF6, IMF7, IMF8	0.8171	2.3069	0.8278	0.7495
B	IMF0, IMF1	IMF2, IMF3, IMF4	IMF5, IMF6, IMF7, IMF8	0.8153	2.3210	0.8389	0.7302
C	IMF0, IMF1, IMF2	IMF3, IMF4	IMF5, IMF6, IMF7, IMF8	0.8656	2.1305	0.8549	0.6930

Mode A and Mode C for subsequence reconstruction, corresponding to an IMF number of 8 and 9, respectively, achieve the best prediction performance. Considering the sample entropy values of each IMF listed in Table 1, we could summarize the threshold of the sample entropy value to obtain the best reconstruction mode.

It can be inferred that subsequences whose sample entropy values are higher than 1.0 or lower than 0.2 should be reconstructed into a new re-IMF; the others are the third re-IMF, setting the principle of subsequence reconstruction in this study.

(3) Select the appropriate subsequence for secondary decomposition.

Secondary decomposition could improve model performance but consume more time; which reconstructed IMF from the subsequence reconstruction process should be re-decomposed is investigated by the comparison experiments listed in Table 5.

**Table 5.** Comparison of prediction models with different re-decomposition programs.

Components to be Re-Decomposed	Well-1		Well-2		Well-3		Well-4	
	$R^2$	Time (s)	$R^2$	Time (s)	$R^2$	Time (s)	$R^2$	Time (s)
re-IMF2	0.8282	865.777	0.8800	858.695	0.8472	801.331	0.8017	901.619
re-IMF1	0.8394	902.950	0.8826	867.678	0.8641	870.178	0.8334	909.153
re-IMF0	0.9235	928.145	0.9603	1001.435	0.9364	944.565	0.9483	926.474
re-IMF0, re-IMF1	0.8802	1206.589	0.9615	1267.169	0.8811	1282.281	0.9151	1205.483
re-IMF0, re-IMF1, re-IMF2	0.8546	1299.403	0.9621	1270.926	0.8781	1283.385	0.8313	1249.192

While decomposing only one subsequence, the prediction performance of the model gradually decreases from re-IMF0 to re-IMF2. On the other hand, with the increase in the number of sequences that are secondarily decomposed, the prediction accuracy of the model improves very slightly or doesn't improve; however, the computing time increases significantly. Re-IMF0 includes much complex information about the source data; further processing could capture the sufficient features to forecast more accurately. Furthermore, while low-frequency data is initially easy to forecast, additional processing leads to increased cumulative errors and calculation workload. Therefore, only the highest-frequency subsequence should be applied to the secondary decomposition procedure.

### 3.3. Experiment I: Comparison of Single Models

The forecasting performance of four single artificial intelligence models without data decomposition is compared in this section, including Support Vector Regression (SVR), Back Propagation (BP) Neural Network, Recurrent Neural Network (RNN) and LSTM. Based on the oil production prediction values of Well-1, the evaluation metrics are calculated and shown in Figure 4.

Evidently, LSTM has the smallest error (RMSE, MAE, MAPE) and highest accuracy ( $R^2$ ) among single models. It implies that, when compared to SVR, BP and RNN, LSTM can grasp the high-sophistication features of the oil production rate more effectively, making it more suitable as a prediction method for production dynamic analysis in oilfields. However, single models still cannot satisfy the requirements of high forecast accuracy for their low-level  $R^2$  score.



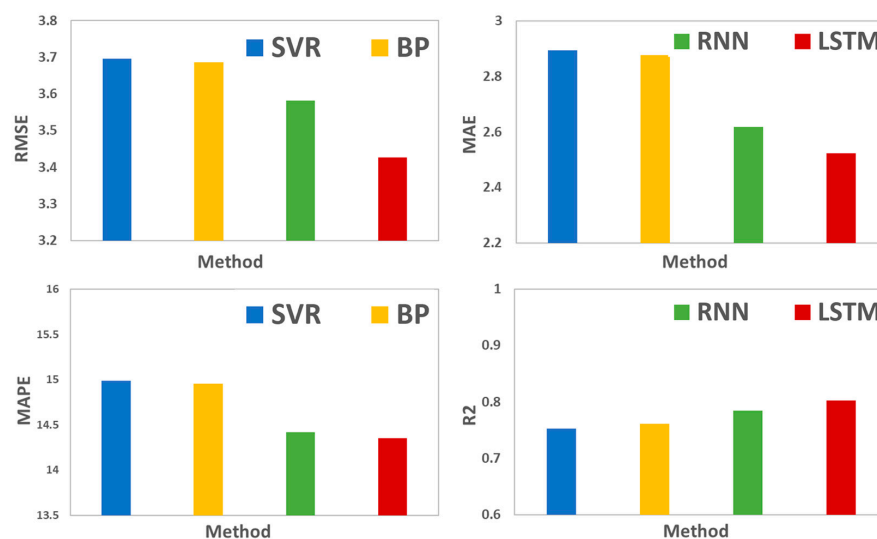


Figure 4. Evaluation results of single models.

### 3.4. Experiment II: Comparison of First Decomposition Methods before Sample Entropy Reconstruction

The primary objective of this scenario is to compare the performance of popular decomposition methods based on the decomposition-sample entropy reconstruction-ensemble forecasting framework. The error index values obtained by different decomposition methods, including EMD, Ensemble Empirical Mode Decomposition (EEMD) and CEEMDAN, are shown in Figure 5, while simple LSTM is the baseline.

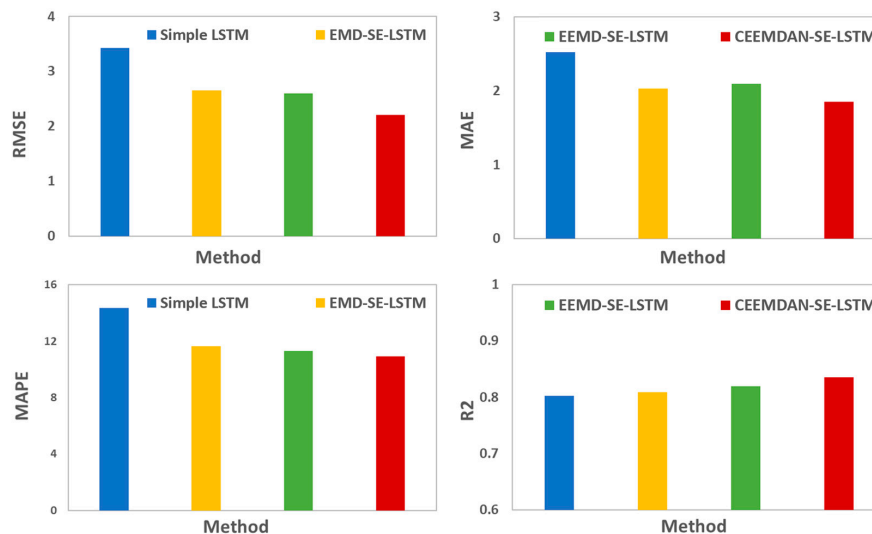


Figure 5. Performance of models with one-step decomposition compared with LSTM.

After adopting EMD, EEMD and CEEMDAN as first-stage decomposition methods, the prediction accuracy of the model has been significantly improved; even the EMD-SE-LSTM model with low accuracy (RMSE = 2.6570, MAE = 2.0306, MAPE = 11.6345,  $R^2 = 0.8093$ ) is much better than the single LSTM. In contrast with EMD and EEMD, CEEMDAN is more suitable for processing well production data.

### 3.5. Experiment III: Comparison of Different Predictors Based on the Hybrid Structure with Primary Decomposition and Sample Entropy Reconstruction

This section investigated the performance of the hybrid structure via primary decomposition and sample entropy reconstruction with a classical predictor. SVR, BP, RNN and

LSTM were introduced as predictors in the ensemble forecasting framework, and their prediction performances are shown in Figure 6.

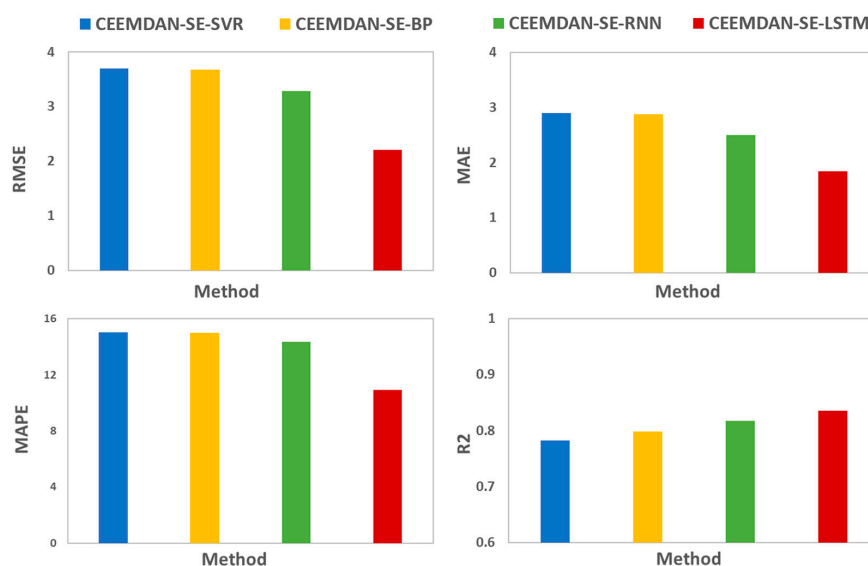


Figure 6. Evaluation indices of the hybrid forecasting structure with different predictors.

Based on the hybrid model, the performance of multiple predictors appeared to follow the following order: deep learning techniques (RNN, LSTM) > machine learning methods (SVR, BP). It could be inferred that the forecasting efficiency of LSTM was infinitely superior to traditional machine learning algorithms and RNN because of its particular structure and ability to process time series.

### 3.6. Experiment IV: Comparison of Different Forecasting Architectures

There are two basic structures during the time series forecast process, which are named integral architecture and corresponding architecture in this issue. Integral architecture means applying all series data to an individual model, so the input data should be a matrix. On the contrary, the corresponding structure is more complicated and accurate because it predicts all series data separately. The forecasting model’s quantities are dependent on the counts of input vectors, which increases the prediction procedure’s calculation time. Considering the evident advantages and disadvantages of two structures comprehensively, a hybrid architecture is established. First, it uses the corresponding structure to forecast each IMF or re-IMF, then uses the integral architecture to integrate the results of the previous step. Figure 7 exhibits the evaluation indicators for three forecasting architectures compared with a simple LSTM baseline. The calculation time of each structure is shown in Table 6.

Although the corresponding architecture has the smallest error, it requires a significant amount of calculation time. The integral architecture runs fastest but sacrifices accuracy. The hybrid structure demonstrates nearly identical performance to the corresponding architecture and saves lots of time. It combines the advantages of two basic architectures, maintaining prediction accuracy while promoting computing speed.

Table 6. Calculation time of different architectures.

Architecture	Calculation Time (s)
Simple LSTM	191.518
CEEMDAN-SE-Integral LSTM	267.187
CEEMDAN-SE-Corresponding LSTM	953.668
CEEMDAN-SE-Hybrid LSTM	554.536

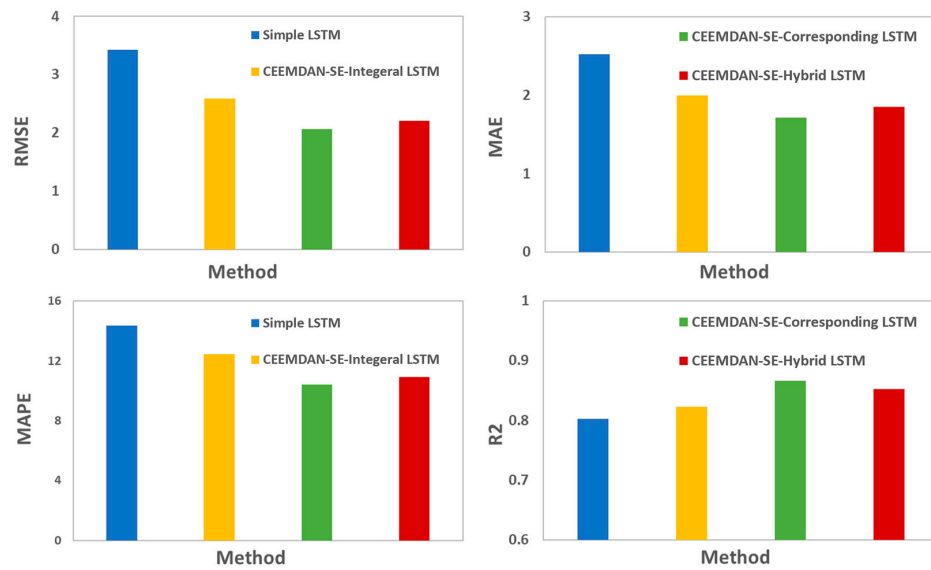


Figure 7. Performance comparison of different forecasting architectures based on LSTM.

### 3.7. Experiment V: Comparison of Second-Stage Decomposition Methods

After first-step processing, the raw data is decomposed into several IMFs, which consist of high-frequency sequences and low-frequency sequences; LSTM could forecast the latter more effectively. Further processing for high-frequency series data could enhance the model’s performance; multi-stage decomposition is suggested. In this experiment, IMF0 or re-IMF0 is decomposed secondarily by different decomposition methods based on the hybrid forecasting architecture, including EMD, EEMD, CEEMDAN and VMD. The comparative result of these models is shown in Figure 8.

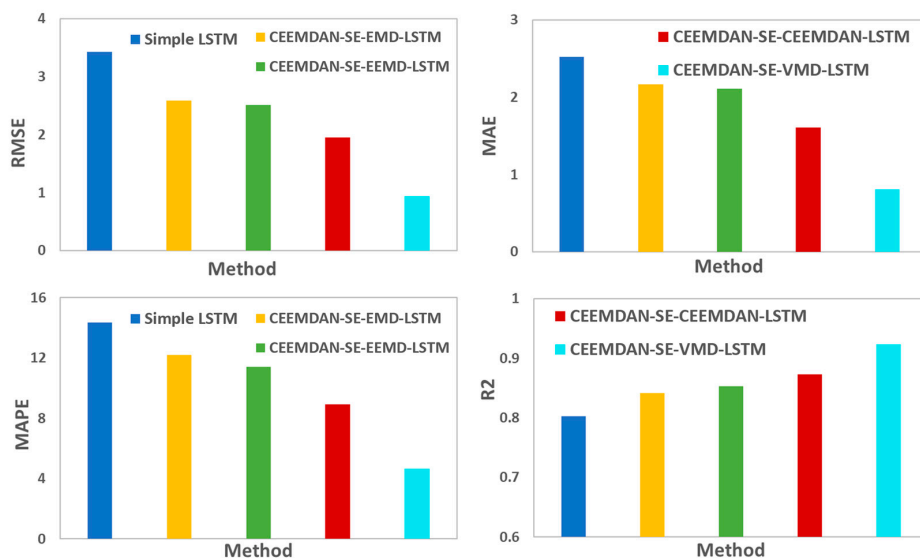


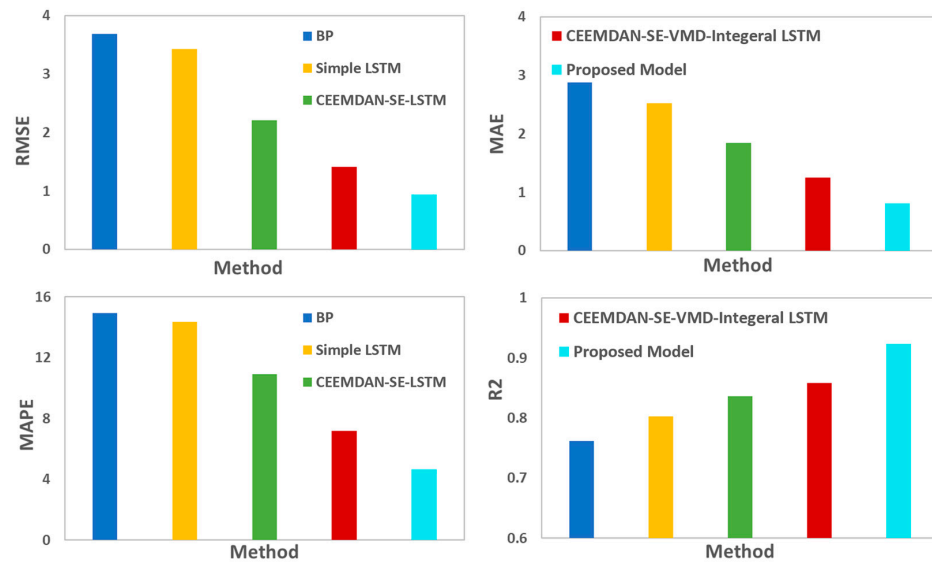
Figure 8. Results of model evaluation for the different secondary decomposition methods.

These three EMD-based methods perform better than simple LSTM; however, applying VMD to the second-stage decomposition process decreases the error evidently (RMSE decreased by 72.74%, MAE by 67.82% and MAPE by 67.66%) and promotes the  $R^2$  score by 15.01%.

### 3.8. Experiment VI: Comparison of Proposed Hybrid Model with Other Forecasting Methods

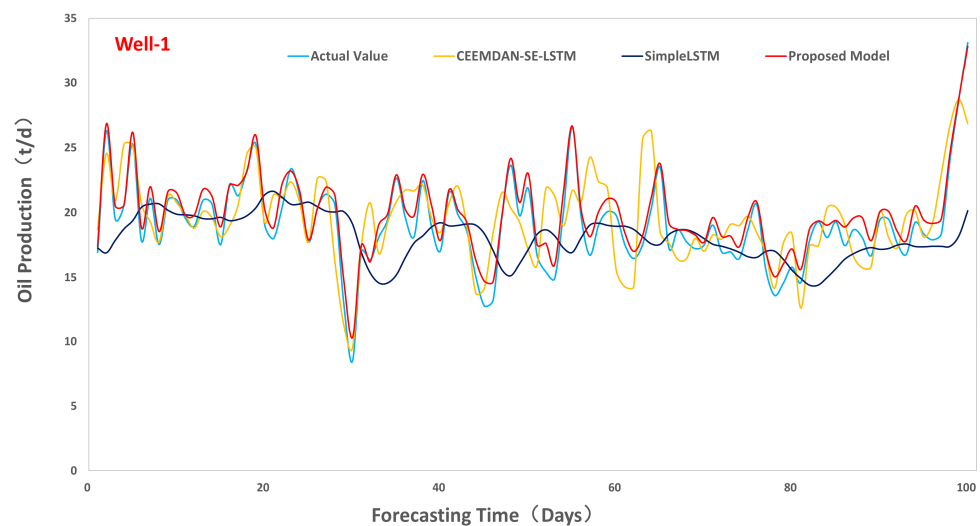
To verify the proposed model’s progression and creativeness, it is essential to compare it to other models that are usually utilized for time series forecasting, including BP, single

LSTM, CEEMDAN-SE-LSTM, and integral forecasting architecture based on VMD second-decomposition. The comparison with these models illustrates the value of the proposed model. The evaluation indices of these models are depicted in Figure 9.



**Figure 9.** Comparison of the proposed model with other methods.

The proposed model has the smallest error and the highest  $R^2$  value among the commonly used methods; the results confirmed the validity of the proposed hybrid approach for oil production forecasting. Figure 10 shows the performance of the hybrid model in predicting Well-1's oil production.



**Figure 10.** Forecasting performance of the proposed model for Well-1.

The hybrid structure of decomposition-reconstruction-secondary decomposition contributes to the ability to distinguish information of different frequencies and capture deeper features of oil production data, achieving more accurate prediction outcomes. In fact, frequently and abruptly changing values, as illustrated in Figures 11 and 12, remain a difficult issue in forecasting. More engineering parameters should be considered when using the model in the future.

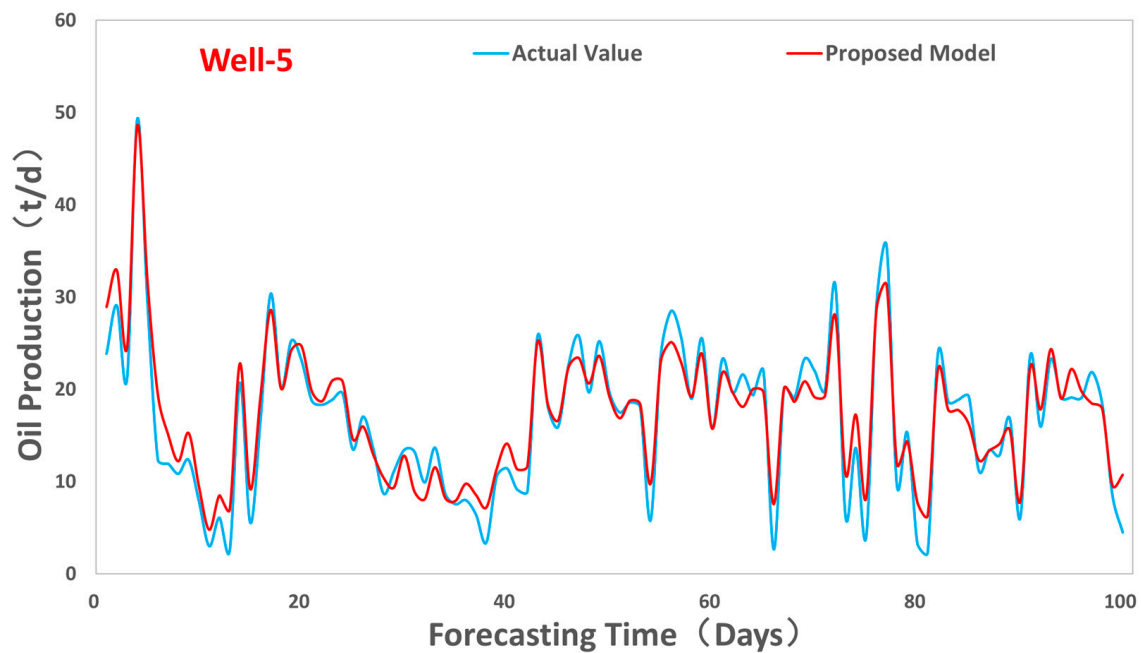


Figure 11. Forecasting performance of the proposed model for Well-5.

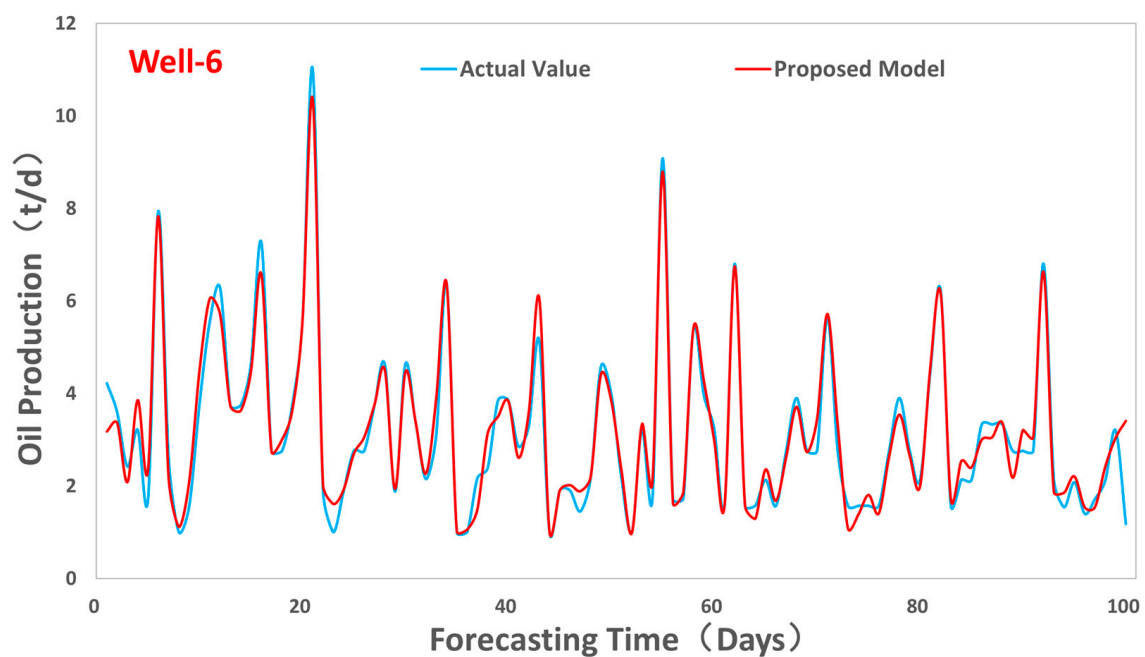


Figure 12. Forecasting performance of the proposed model for Well-6.

### 3.9. Experiment VII: Validations in Other Production Wells

The proposed hybrid prediction framework achieved outstanding performance in forecasting Well-1's production dynamics. We implement it for other wells' production predictions to validate the hybrid model. Figures 13–15 demonstrate the production forecasting results of three wells in the JL volcanic reservoir.

The results validate that the hybrid model proposed in this study also achieves good accuracy in forecasting other wells' production. The proposed model could obtain the characteristics and trends of production history data and make accurate predictions, providing an applicable method for reservoir production forecasting.

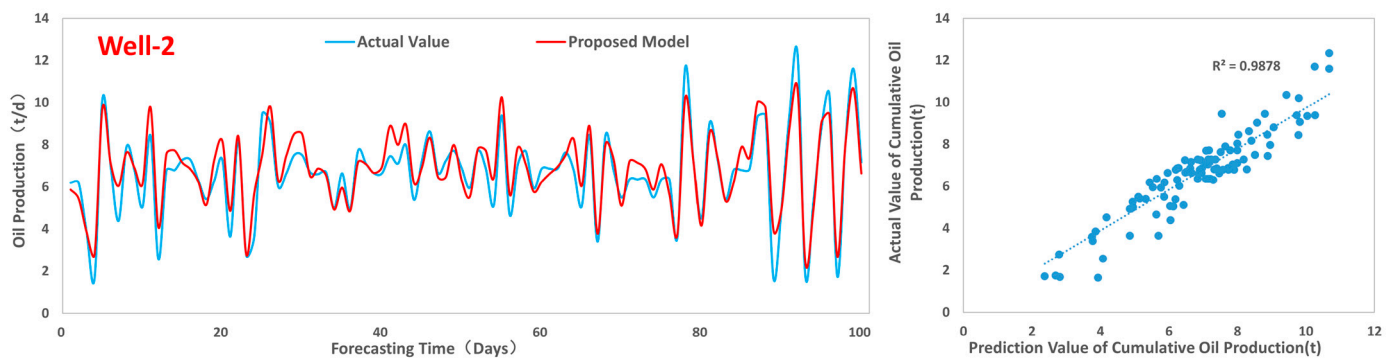


Figure 13. Production forecasting with proposed hybrid model for Well-2.

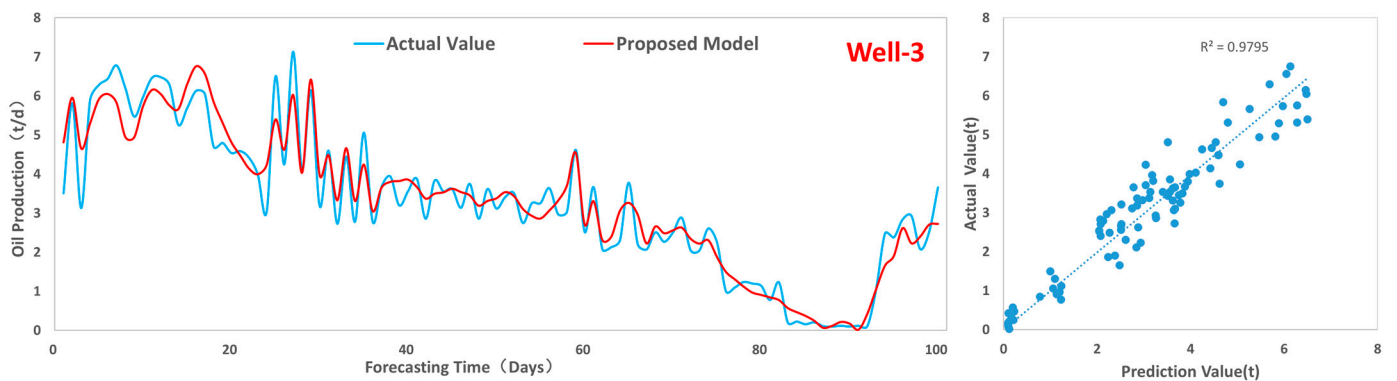


Figure 14. Forecasting result for Well-3.

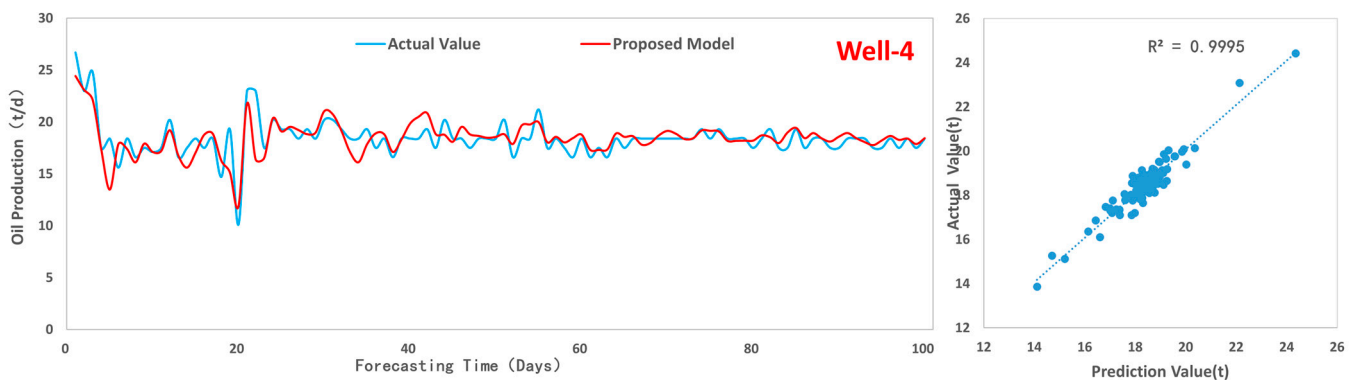


Figure 15. Prediction performance of the proposed model for Well-4.

#### 4. Discussion

This study aims to establish a hybrid framework to analyze the production dynamics and improve the prediction accuracy of oil reservoirs. Simple models cannot capture all features of reservoir history information, making inaccurate predictions. Decomposition methods could process nonstationary, complex and low-quality data in the petroleum industry. The raw data can be transformed into subsequences, facilitating feature engineering; however, the complexity of some subsequences is still high. To enhance the model's efficiency and avoid unnecessary computing, IMFs' integration is conducted under the rules of sample entropy reconstruction determined by evaluation experiments. The remaining highest-frequency component contains the main irregular parts of the data, which inhibits the model's performance. A secondary decomposition could improve the model's prediction performance by addressing nonstationary and nonlinear issues in the highest-frequency data and fully extracting time series features. The LSTM structure, which is able to capture and store valuable features from time series data, is appropriate for the hybrid

forecasting structure with high prediction accuracy. The proposed model is applied for single-well production prediction in the JL volcanic reservoir and performs outstandingly, achieving more accurate forecast results than single models and one-step decomposition structures. Moreover, the rules for reconstruction and re-decomposition make the process more automated and intelligent.

Although principles of reconstruction and secondary decomposition are specified for the proposed framework, we advise adopting the method in other reservoirs with similar geological or production characteristics. For different types of oilfields, the threshold in the rules should be adjusted according to the actual production dynamics of oil wells. Furthermore, more geological and engineering features should be considered to enhance the forecast of suddenly changing values, and the predictor in the final stage of the structure should be determined among multiple methods by comparable evaluations since there is no perfect technique for every task.

## 5. Conclusions

Oil production forecasting is extremely significant during oilfield development, particularly in unconventional reservoirs such as volcanic reservoirs. Traditional methods and simple machine learning algorithms cannot achieve sufficiently high accuracy in oil production forecasting results. High-frequency data from the decomposition-prediction strategy also limits the forecasting efficiency; thus, a multi-stage decomposition model is proposed in this study. The hybrid model consists of CEEMDAN, sample entropy reconstruction, secondary decomposition based on VMD and the LSTM forecasting process. Based on data-driven theory, two-stage decomposition could aid in extracting features and increasing prediction accuracy. The structure synthesizes the advantages of the integral and corresponding architecture, making it outstanding among comparative experiments. The rules of reconstruction and secondary decomposition are defined, and which subsequence should be decomposed again is also proposed, improving the workflow's generalizability. The proposed hybrid model was validated in several actual production wells, illustrating its wide applicability. We can apply the model to other reservoirs with similar geological features or similar oil production patterns, such as other volcanic reservoirs, naturally fractured reservoirs and low permeable reservoirs developed by artificial fracturing. The common property of these formations is that their oil production capacity is determined by fracture development. Besides predicting production, the method helps to understand the reservoir thoroughly and adjust the subsequent development scheme.

**Author Contributions:** Conceptualization, X.K.; methodology, X.K.; software, X.K.; validation, X.K.; formal analysis, X.K.; investigation, X.K.; resources, Y.L.; data curation, X.K. and D.Z.; writing—original draft preparation, X.K.; writing—review and editing, Y.L., L.X. and G.L.; visualization, X.K.; supervision, Y.L.; project administration, Y.L.; funding acquisition, Y.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China (No. 51374222, No.52274048), the National Basic Research Program of China (973 Program, No. 2015CB250905), the National Major Science and Technology Projects of China (No. 2017ZX05032004-002), CNPC Major Scientific Research Project (No. 2017E-0405), SINOPEC Major Scientific Research Project (No. P18049-1), Beijing Natural Science Foundation (No. 3222037), the PetroChina Innovation Foundation (No. 2020D-5007-0203), the Science Foundation of China University of Petroleum, Beijing (No. 2462021YXZZ010) and the PetroChina perspective fundamental research project (No. 2021DJ2104).

**Institutional Review Board Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

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

## References

1. Arps, J.J. Analysis of decline curves. *Trans. AIME* **1945**, *160*, 228–247. [[CrossRef](#)]
2. Ji, J.; Yao, Y.; Huang, S.; Ma, X.; Zhang, S.; Zhang, F. Analytical model for production performance analysis of multi-fractured horizontal well in tight oil reservoirs. *J. Pet. Sci. Eng.* **2017**, *158*, 380–397. [[CrossRef](#)]
3. Sun, R.; Hu, J.; Zhang, Y.; Li, Z. A semi-analytical model for investigating the productivity of fractured horizontal wells in tight oil reservoirs with micro-fractures. *J. Pet. Sci. Eng.* **2020**, *186*, 106781. [[CrossRef](#)]
4. Alfi, M.; Hosseini, S.A. Integration of reservoir simulation, history matching, and 4D seismic for CO<sub>2</sub>-EOR and storage at Cranfield, Mississippi, USA. *Fuel* **2016**, *175*, 116–128. [[CrossRef](#)]
5. Ning, Y.; Kazemi, H.; Tahmasebi, P. A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet. *Comput. Geosci.* **2022**, *164*, 105126. [[CrossRef](#)]
6. Song, X.; Liu, Y.; Xue, L.; Wang, J.; Zhang, J.; Wang, J.; Jiang, L.; Cheng, Z. Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *J. Pet. Sci. Eng.* **2020**, *186*, 106682. [[CrossRef](#)]
7. Wang, S.; Chen, Z.; Chen, S. Applicability of deep neural networks on production forecasting in Bakken shale reservoirs. *J. Pet. Sci. Eng.* **2019**, *179*, 112–125. [[CrossRef](#)]
8. Huang, R.; Wei, C.; Wang, B.; Yang, J.; Xu, X.; Wu, S.; Huang, S. Well performance prediction based on Long Short-Term Memory (LSTM) neural network. *J. Pet. Sci. Eng.* **2022**, *208*, 109686. [[CrossRef](#)]
9. Sagheer, A.; Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **2019**, *323*, 203–213. [[CrossRef](#)]
10. Cheng, Y.; Yang, Y. Prediction of oil well production based on the time series model of optimized recursive neural network. *Pet. Sci. Technol.* **2021**, *39*, 303–312. [[CrossRef](#)]
11. Lu, W.; Rui, H.; Liang, C.; Jiang, L.; Zhao, S.; Li, K. A Method Based on GA-CNN-LSTM for Daily Tourist Flow Prediction at Scenic Spots. *Entropy* **2020**, *22*, 261. [[CrossRef](#)] [[PubMed](#)]
12. Yang, W.; Wang, J.; Wang, R. Research and Application of a Novel Hybrid Model Based on Data Selection and Artificial Intelligence Algorithm for Short Term Load Forecasting. *Entropy* **2017**, *19*, 52. [[CrossRef](#)]
13. Zha, W.; Liu, Y.; Wan, Y.; Luo, R.; Li, D.; Yang, S.; Xu, Y. Forecasting monthly gas field production based on the CNN-LSTM model. *Energy* **2022**, *260*, 124889. [[CrossRef](#)]
14. Fan, D.; Sun, H.; Yao, J.; Zhang, K.; Yan, X.; Sun, Z. Well production forecasting based on ARIMA-LSTM model considering manual operations. *Energy* **2021**, *220*, 119708. [[CrossRef](#)]
15. Li, W.; Wang, L.; Dong, Z.; Wang, R.; Qu, B. Reservoir production prediction with optimized artificial neural network and time series approaches. *J. Pet. Sci. Eng.* **2022**, *215*, 110586. [[CrossRef](#)]
16. Meng, F.; Xu, D.; Song, T. ATDNNS: An adaptive time–frequency decomposition neural network-based system for tropical cyclone wave height real-time forecasting. *Future Gener. Comput. Syst.* **2022**, *133*, 297–306. [[CrossRef](#)]
17. Lv, P.; Wu, Q.; Xu, J.; Shu, Y. Stock Index Prediction Based on Time Series Decomposition and Hybrid Model. *Entropy* **2022**, *24*, 146. [[CrossRef](#)]
18. Zhou, F.; Huang, Z.; Zhang, C. Carbon price forecasting based on CEEMDAN and LSTM. *Appl. Energy* **2022**, *311*, 118601. [[CrossRef](#)]
19. Liu, W.; Liu, W.D.; Gu, J. Forecasting oil production using ensemble empirical model decomposition based Long Short-Term Memory neural network. *J. Pet. Sci. Eng.* **2020**, *189*, 107013. [[CrossRef](#)]
20. Wang, F.; Zhang, D.; Min, G.; Li, J. Reservoir Production Prediction Based on Variational Mode Decomposition and Gated Recurrent Unit Networks. *IEEE Access* **2021**, *9*, 53317–53325. [[CrossRef](#)]
21. Guo, Z.H.; Zhao, W.G.; Lu, H.Y.; Wang, J.Z. Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model. *Renew. Energy* **2012**, *37*, 241–249. [[CrossRef](#)]
22. Torres, M.E.; Colominas, M.A.; Schlotthauer, G.; Flandrin, P. A Complete Ensemble Empirical Mode Decomposition with Adaptive Noise. In Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech Republic, 22–27 May 2011; pp. 4144–4147.
23. Dragomiretskiy, K.; Zosso, D. Variational mode decomposition. *IEEE Trans. Signal Process.* **2013**, *62*, 531–544. [[CrossRef](#)]
24. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [[CrossRef](#)] [[PubMed](#)]
25. Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.