

Research Article

Time Series Analysis of Production Decline in Carbonate Reservoirs with Machine Learning

Liqiang Wang ¹, Mingji Shao ², Gen Kou ³, Maoxian Wang ², Ruichao Zhang ¹, Zhengzheng Wei ¹ and Xiao Sun ¹

¹Department of Petroleum Engineering, Shengli College, China Petroleum University, Dongying, 257061 Shandong, China

²Exploration and Development Research Institute of Tuha Oilfield Company, CNPC, Hami, 839009 Xinjiang, China

³Experimental Research Institute of Xinjiang Oilfield Company, CNPC, Karamay, 834000 Xinjiang, China

Correspondence should be addressed to Liqiang Wang; 281035925@qq.com

Received 21 October 2020; Revised 31 December 2020; Accepted 5 January 2021; Published 23 January 2021

Academic Editor: Zhiming Chen

Copyright © 2021 Liqiang Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Classical decline methods, such as Arps yield decline curve analysis, have advantages of simple principles and convenient applications, and they are widely used for yield decline analysis. However, for carbonate reservoirs with high initial production, rapid decline, and large production fluctuations, with most wells having no stable production period, the adaptability of traditional decline methods is inadequate. Hence, there is an urgent need to develop a new decline analysis method. Although machine learning methods based on multiple regression and deep learning have been applied to unconventional oil reservoirs in recent years, their application effects have been unsatisfactory. For example, prediction errors based on multiple regression machine learning methods are relatively large, and deep learning sample requirements and the actual conditions of reservoir management do not match. In this study, a new equal probability gene expression programming (EP-GEP) method was developed to overcome the shortcomings of the conventional Arps decline model in the production decline analysis of carbonate reservoirs. Through model validation and comparative analysis of prediction effects, it was proven that the EP-GEP model exhibited good prediction accuracy, and the average relative error was significantly smaller than those of the traditional Arps model and existing machine learning methods. The successful application of the proposed method in the production decline analysis of carbonate reservoirs is expected to provide a new decline analysis tool for field reservoir engineers.

1. Introduction

There are three main stages in the complete production cycle of oil and gas wells: production rise, stability, and decline. During the production decline stage, the selection of the decline model has a significant impact on the prediction of production dynamics and the evaluation of the final recovery factor [1, 2]. The advantages of classical methods, such as the Arps production decline curve analysis, are that their principle is simple and easy to apply, and various explicit expressions can be derived. The derived expressions can predict future dynamic production and recoverable reserves of oil and gas reservoirs in a pseudo-steady state. The disadvantage of the traditional methods [3–8] is that the selection of the decline model depends on experience; the dependent variable is

single, and it is difficult to describe the nonlinear relationship of the production change precisely. For example, the storage and control model of carbonate reservoirs is different from that of clastic rocks. Reservoirs have developed matrix, fractures, caverns, and other storage spaces, poor connectivity, strong heterogeneity, high initial production wells, rapid decline, and large production fluctuations. Most reservoir wells have no stable production period and need to be evaluated to determine a set of production prediction methods suitable for carbonate reservoirs. Therefore, new modeling methods need to be developed and applied to predict the production of carbonate reservoirs accurately. Currently, machine learning is increasingly applied in several industries, and some exploratory application cases in the petroleum industry have been investigated [9–11].

In recent years, the production decline methods based on multiple regression machine learning, such as artificial neural networks, support vector regression, random forests, and gradient enhancement, are gradually replacing traditional data analysis methods [12–17]. However, existing machine learning methods based on multiple regression produce large prediction errors in oil well production decline analysis.

In addition, deep learning methods, such as recurrent neural networks, have been applied for production decline analysis [18, 19]. However, the deep learning method is most suitable for high-frequency (such as daily) production data, owing to the characteristics of its network structure. This significantly limits the application of the deep learning method because most production data exist in the form of monthly records. Although deep learning methods can be designed to process monthly data, they also require complex network structures for processing temporal and nontemporal data. Compared to regression-based machine learning methods, reservoir engineers have difficulty using deep learning methods [20]. Because of the problems mentioned above and the lack of application of machine learning in the analysis of carbonate reservoir decline, the machine learning method of gene expression programming (GEP) was used in this study to analyze the production decline of carbonate reservoirs.

GEP is based on the genetic algorithm (GA) and genetic programming (GP). It exhibits excellent performance in knowledge mining, function discovery, optimization, and prediction [21]. GEP is a machine learning modeling tool that can be used to establish an explicit model with a simple structure and high prediction accuracy through evolution without knowing the structure and parameters of the model in advance and without having to have the domain background knowledge and thus avoids the mechanism analysis of the system. There is difficulty in establishing a predictive model and the preset model structure based on regression methods, and the subjectivity of parameters is then determined using statistical methods [22]. The GEP method has been successfully applied in many disciplines and fields [23–26]. However, the use of GEP to predict the production of carbonate reservoirs has not been reported. Therefore, it is necessary to model the GEP machine learning method and predictive effects to conduct more in-depth research.

2. Equal Probability GEP Algorithm

GEP combines the advantages of GA and GP. In terms of expression, it inherits the simple and rapid characteristics of the fixed-length linear coding of GA, and in terms of gene expression (semantic expression), it inherits the flexible tree structure of GP. The change characteristics, i.e., simple coding to solve complex problems, are 2–4 times faster than traditional machine learning evolutionary algorithms [22].

However, the knowledge mining process of GEP is passive and can easily fall into a local optimum. Undirected evolution and premature convergence reduce the efficiency and quality of the solution. Hence, it is necessary to guide the evolution process of the gene population and adopt specific methods to prevent the solution process from falling into

the local optimum. The equal probability GEP (EP-GEP) method developed in this study can effectively solve the problem of evolutionary undirected and premature local convergence and improve the convergence efficiency and solution quality of the algorithm.

The EP-GEP optimization calculation process is performed as follows. Randomly generate a specific number of chromosomal individuals to form the initial population. Produce the candidate set from the outstanding individuals in the initial population. Select the best adaptation for the individuals in the population according to the decline analysis of the carbonate oil well production. Next, based on the fitness function, evaluate the responsiveness of each individual in the population. Select, mutate, insert, recombine, and perform other genetic operations on the individuals in the population to produce new offspring and form a new population. The newly generated population continues to enter the next round of the optimization process. If premature local convergence occurs in this process summary, enter the calculation process of the equal probability gene expression optimization and perform genetic operations, such as equal probability selection, mutation, string insertion, and recombination, on individuals in the population (the three with equal probability in Figure 1, for example). New offspring are produced to form a new population, and the newly generated population and candidate set continue to enter the next round of optimization calculations. Subsequently, repeat the optimization calculation process until the iteration termination condition is satisfied. A flowchart of the optimization process is depicted in Figure 1.

Because the EP-GEP algorithm is based on traditional GEP, the gene structure, genetic operator, and fitness function are the same as those of the GEP algorithm described in Sections 2.1–2.3.

2.1. Gene Structure. The object of EP-GEP processing is a chromosome (genome) composed of a single gene or multiple genes. The gene in EP-GEP is based on a simplification of the principle of genes in biology. It consists of a linear, fixed-length string of symbols. Although the chromosome length is fixed, expression trees (ETs) of different sizes and shapes can be expressed to generate diverse individuals. An example is the following algebraic expression:

$$\sqrt{(a+b) \times (c+d)}. \quad (1)$$

The corresponding expression tree, i.e., the individual's phenotype, is shown in Figure 2, where Q is the square root function. From top to bottom and from left to right, the expression tree can be traversed to obtain the corresponding K expression.

From Figure 2, the K expression expressed in equation (2) can be obtained, which is the genotype in GEP.

$$\begin{array}{cccccccc} 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ q & * & + & + & a & b & c & d \end{array}. \quad (2)$$

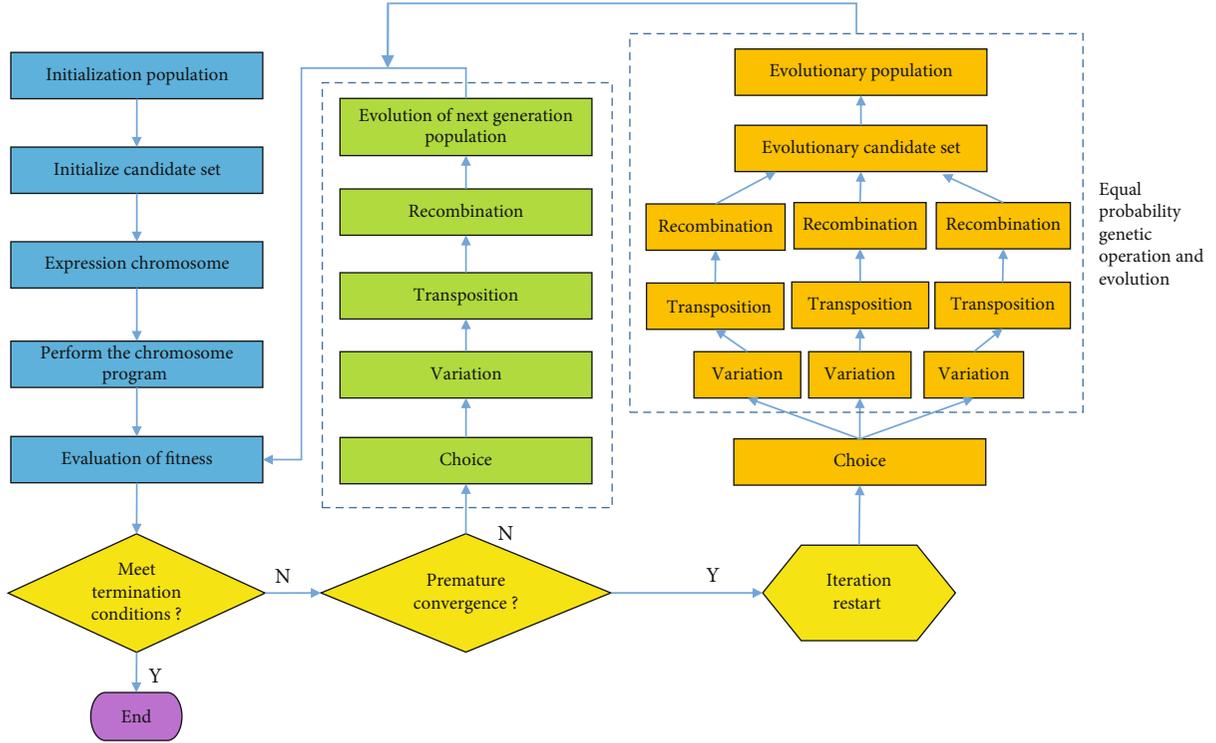


FIGURE 1: Flowchart of EP-GEP.

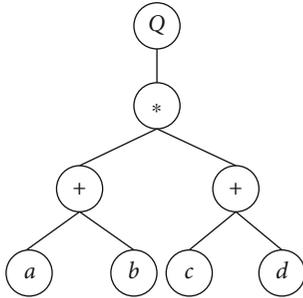


FIGURE 2: Expression tree corresponding to expression (1).

EP-GEP divides genes into the head and the tail. The gene head can be composed of variables or functions, but the gene tail can only be composed of variables. The head length and tail length t satisfy the following equation:

$$t = h \times (n - 1) + 1, \quad (3)$$

where n denotes all functions in the function set.

2.2. Genetic Operator. EP-GEP creates an initial population in the algorithm, and each chromosome in the population represents a solution to the problem. Subsequently, a series of genetic operations are performed to generate new individuals with high adaptability to obtain a better solution. The basic genetic operators of GEP include nine types, i.e., selection, mutation, inverted string, string insertion, root string insertion, gene transformation, single-point

recombination, two-point recombination, and gene recombination [21].

2.3. Fitness Function. The environmental adaptability of newly generated chromosomes should be evaluated to obtain the best solution. Similar to other machine learning evolutionary algorithms, the size of the fitness function value (i.e., fitness) is used in EP-GEP to evaluate chromosome quality. Sometimes, a suitable fitness function can be customized according to the problem to be solved.

The selection of the fitness function must be combined with specific practical problems. Choosing different fitness functions may cause the range of fitness functions such as variance, standard deviation, and root mean square error (RMSE) to vary significantly.

Combined with the problem of analysis of production decline in carbonate oil wells, the aim of the analysis is to solve sign regression. Fitness functions widely used to solve this problem are the mean square error, RMSE, and mean absolute error [27]. A minimum value was required. When the difference between the predicted and actual values is zero, then the ideal minimum value is zero [18]. In this study, the RMSE was obtained using the fitness function expressed in the following equation:

$$RMSE = \sqrt{\frac{1}{m} \sum_{j=1}^m (y_j - \hat{y}_j)^2}. \quad (4)$$

TABLE 1: Experimental parameters of EP-GEP.

Parameter	Value	Parameter	Value
Population size	50	Length of head	8
Gene number	5	Mutation rate	0.00546
Recombination rate	0.00277	One-point recombination rate	0.00277
Two-point recombination rate	0.00277	Transposition rate	0.00277
Root insertion sequence (RIS) transposition rate	0.00546	Insertion sequence (IS) transposition rate	0.00546
Link function	Avg2	Fitness function	RMSE

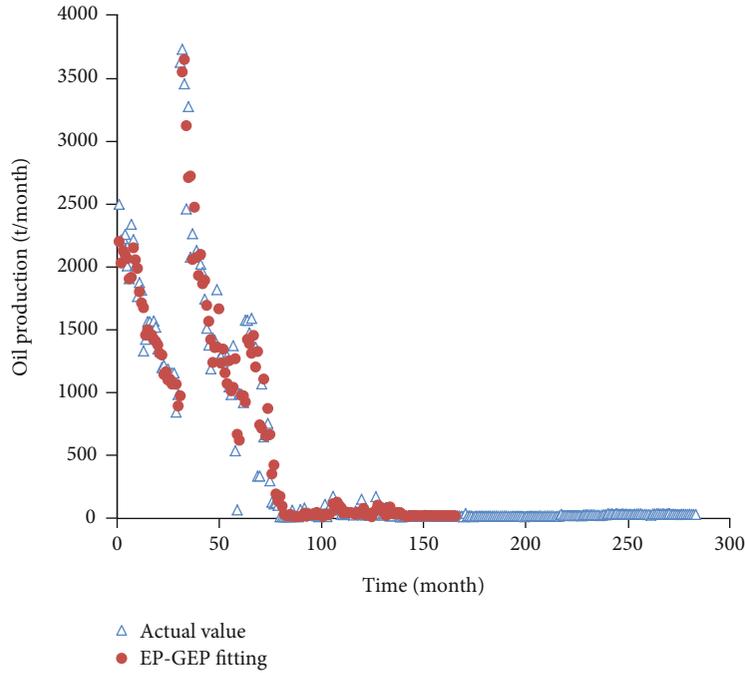


FIGURE 4: Comparison of actual oil production and EP-GEP fitting value.

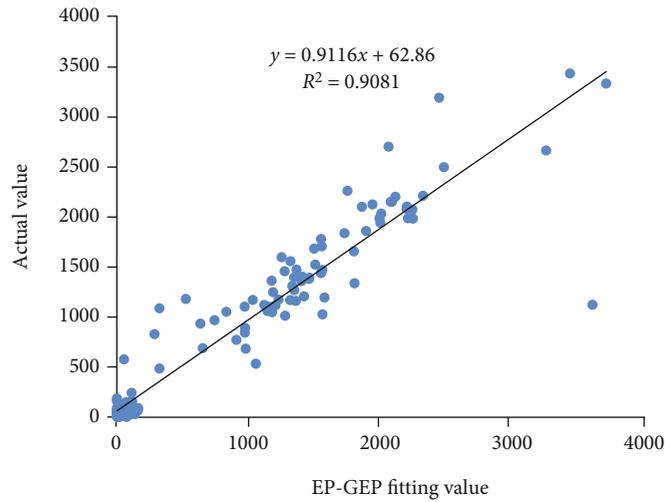


FIGURE 3: Regression graph of actual oil production and EP-GEP fitting value.

2.4. Decline Analysis Data Set. The North Akar Oilfield in Kazakhstan is a carbonate oil reservoir with reservoir spaces, such as matrix, fractures, and karst caves. The reservoirs have

weak connectivity and strong heterogeneity. Production wells rely on natural energy extraction and exhibit high initial production, rapid decline, and large production

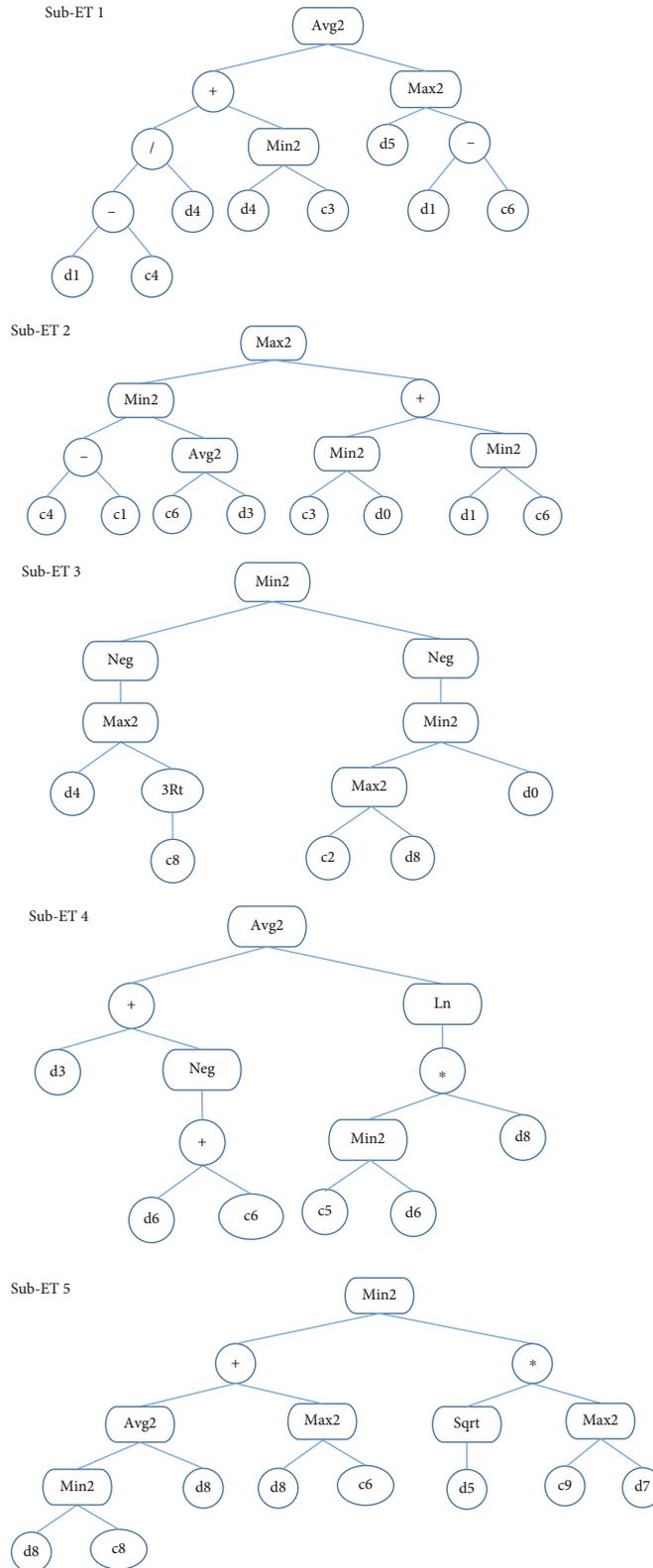


FIGURE 5: Phenotype frame diagram after training optimization.

fluctuations, and most wells have no stable production period. In this study, Well A2 with typical production characteristics was selected for the GEP method adaptability analysis.

3. Results and Analysis

3.1. EP-GEP Time Series Model Training. The first 175 data sets from 284 sets of Well A2 were used for the EP-GEP time

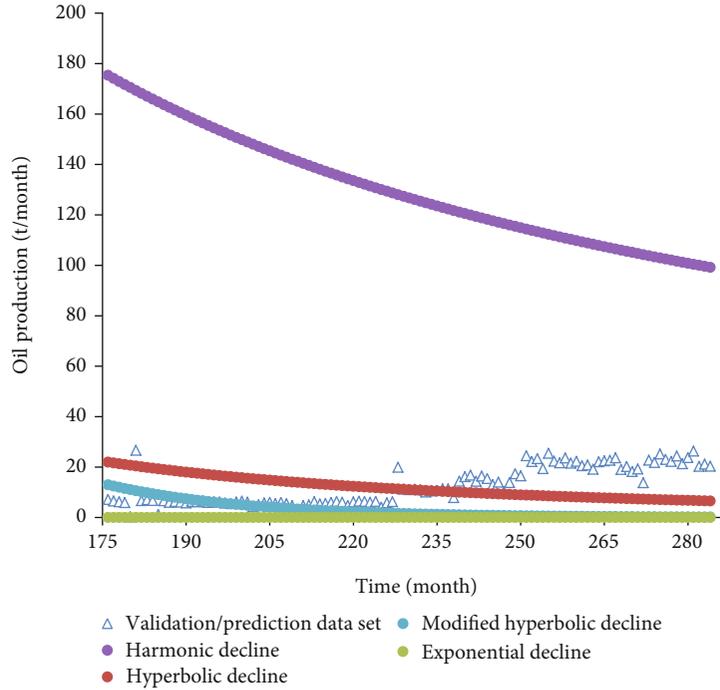


FIGURE 6: Comparison of prediction results of the Arps decline model and verification/prediction data set.

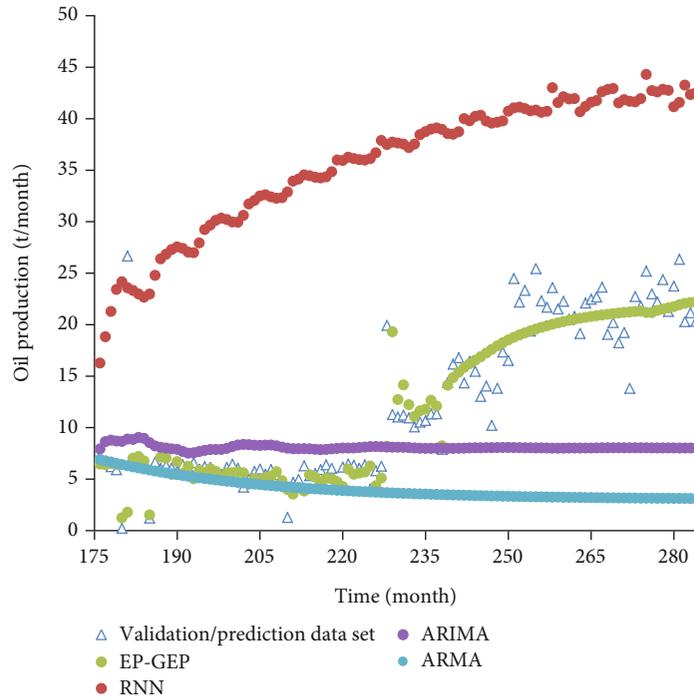


FIGURE 7: Comparison chart of prediction results of different time series analysis algorithms.

series training. The experimental parameters are listed in Table 1. Groups 176–284 were used as verification and prediction data sets.

The R^2 value of the model after training was 0.9084 (Figure 3). After the EP-GEP model training, a comparison between the fitted values and the actual oil production values was performed (Figure 4).

The optimized phenotype frame is composed of five sub-ETs (Figure 5).

3.2. EP-GEP Model Verification and Prediction. The trained EP-GEP model was used to predict the verification/prediction data set. The model was compared with other decline methods, such as hyperbolic decline, exponential decline,

TABLE 2: Comparison of the average relative error of machine learning model predictions for different time series of oil wells.

Well ID	Average relative error (%)			
	EP-GEP	ARMA	ARIMA	RNN
A20	4.37	-12.88	42.48	324.53
A35	6.79	-15.24	38.59	375.22
A7	4.56	-11.91	31.23	289.55
A12	3.27	-11.13	29.02	247.36
A18	7.63	-22.54	45.77	346.31

harmonic decline, and modified hyperbolic decline, to analyze the decline of the Well A2 production curve. The decline equation was used to predict the verification/prediction data set. In addition, the forecasting results obtained using other time series forecasting methods, e.g., the autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), and those using neural network machine learning methods, e.g., the recurrent neural network (RNN), were compared.

Among the four Arps declining prediction models, the harmonic declining model showed the largest error, followed by the exponential and modified hyperbolic declining models (Figure 6). The hyperbolic declining model had the smallest error, but there was a large error with the verification/prediction data set.

The R^2 value after the ARMA model training was 0.8910 (Figure 7), which was more consistent with the change in the actual production data from 175 to 230 months. The trend continued from 230 to 280 months but was slightly different from the verification/prediction data set. The R^2 value after the ARIMA model training was 0.8963, but the prediction result did not exhibit a significant upward or downward trend, and the error with the verification/prediction data set was also large. The RNN model showed a good training effect on the training set; its R^2 value was 0.9081, but the prediction result showed a large error, indicating that the method had poor adaptability to small sample data sets.

Overall, the model established using the EP-GEP method performed better than the traditional decline analysis, ARMA/ARIMA time series, RNN, and other neural network machine learning models. The average error of the EP-GEP prediction model was 3.69%.

3.3. Verification of EP-GEP Machine Learning Validity. The description above only showed the validity of the EP-GEP model for the analysis of oil well production decline. The validity of the EP-GEP algorithm was verified for the other five wells in the study area (Table 2). The results showed that the overall prediction effect of the EP-GEP algorithm was better than that of other time series machine learning methods.

4. Conclusion

- (1) In this study, GEP methods were developed to address the shortcomings of conventional Arps

decline models in analyzing the production decline of carbonate reservoirs. The comparative analysis of model validity and prediction demonstrated that the EP-GEP model exhibited good prediction accuracy, and the average relative error was smaller than those of the traditional Arps models

- (2) The results of multiple oil well production decline analyses showed that the EP-GEP machine learning algorithm yielded higher prediction accuracy and provided better stability than those of other time series machine learning methods. The proposed algorithm can provide on-site reservoir engineers with new reservoir management analysis tools

Nomenclature

- Avg2: Average of two inputs, $\text{avg}(x, y)$
 Max2: Maximum of two inputs, $\max(x, y)$
 Min2: Minimum of two inputs, $\min(x, y)$
 Neg: Negative value of one input, $-x$
 3Rt: Cube root of one input, $x^{1/3}$
 Ln: Natural logarithm of one input, $\ln(x)$
 Sqrt: Square root of one input, $\text{sqrt}(x)$.

Data Availability

The data used to support the findings of this study are currently under embargo while the research findings are commercialized. Requests for data, 12 months after publication of this article, will be considered by the corresponding author.

Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the research, authorship, and publication of this article.

Acknowledgments

This study was supported by the Scientific Research Start-Up Fund for Introducing High-Level Talents from the Shengli College of China University of Petroleum (kq2019-005).

References

- [1] Q. Yu, "Study on characteristics of seven decline curves," *Petroleum Geology of Xinjiang*, vol. 1, pp. 49–56, 1994.
- [2] Z. Yuchuan, L. Jianyi, Z. Guangdong, Z. Deli, and L. Jin, "Comparative analysis of gas well production decline law analysis methods," *Natural Gas Exploration and Development*, vol. 32, no. 1, pp. 28–31, 2009.
- [3] Z. Yihua, Z. Zhiyin, and Z. Haishuang, "A new method for oil production prediction in ultra-high water cut stage," *Fault Block Oil and Gas Field*, vol. 18, no. 5, pp. 641–644, 2011.
- [4] J. D. Baihly, R. M. Altman, R. Malpani, and F. Luo, "Shale gas production decline trend comparison over time and basins," in *SPE annual technical conference and exhibition. Society of Petroleum Engineers*, Florence, Italy, 2010.

- [5] B. Liang, C. Youlian, and H. Yang, "Application of modern production prediction model," *Natural Gas Exploration and Development*, vol. 27, no. 2, pp. 54–56, 2004.
- [6] Z. Shunli, L. Xiuting, S. Wang, Y. Jun, and Y. Ji, "Study and discussion on a new production prediction method," *Fault Block Oil and Gas Field*, vol. 13, no. 2, pp. 51–53, 2006.
- [7] L. P. Dake, "Fundamentals of reservoir engineering," *Developments in Petroleum Science*, vol. 8, 1978.
- [8] J. J. Arps, "Analysis of decline curves," *Transactions of the AIIME*, vol. 160, pp. 228–247, 1945.
- [9] T. Nanjo and S. Tanaka, "Carbonate lithology identification with machine learning," in *Abu Dhabi International Petroleum Exhibition & Conference*, Abu Dhabi, UAE, November 2019.
- [10] Y. Alaudah, P. Michałowicz, M. Alfarraj, and G. AlRegib, "A machine learning benchmark for facies classification," *Interpretation*, vol. 7, no. 3, pp. SE175–SE187, 2019.
- [11] F. I. Syed, M. Alshamsi, A. K. Dahaghi, and S. Neghabhan, "Artificial lift system optimization using machine learning applications," *Petroleum*, vol. 8, no. 3, 2020.
- [12] Q. Cao, R. Banerjee, S. Gupta, J. Li, W. Zhou, and B. Jeyachandra, "Data driven production forecasting using machine learning," in *Presented at the SPE Argentina Exploration and Production of Unconventional Resources Symposium*, Buenos Aires, June 2016.
- [13] A. Kamari, A. H. Mohammadi, M. Lee, and A. Bahadori, "Decline curve based models for predicting natural gas well performance," *Petroleum*, vol. 3, no. 2, pp. 242–248, 2017.
- [14] Y. Li and Y. Han, "Decline curve analysis for production forecasting based on machine learning," in *Presented at the SPE Symposium: Production Enhancement and Cost Optimisation*, Kuala Lumpur, November 2017.
- [15] A. Vyas, A. Datta-Gupta, and S. Mishra, "Modeling early time rate decline in unconventional reservoirs using machine learning techniques," in *Presented at the SPE Abu Dhabi International Petroleum Exhibition and Conference*, Abu Dhabi, November 2017.
- [16] S. Amr, H. El Ashhab, M. El-Saban et al., "A large-scale study for a multi-basin machine learning model predicting horizontal well production," in *Presented at the SPE Annual Technical Conference and Exhibition*, Dallas, September 2018.
- [17] L. A. Sarduy and U. C. Edelmann, "Method for estimating oil/gas production using statistical learning models," 2018, US Patent 1-5426761.
- [18] J. Sun, X. Ma, and M. Kazi, "Comparison of decline curve analysis DCA with recursive neural networks RNN for production forecast of multiple wells," in *Presented at the SPE Western Regional Meeting*, California, USA, April 2018.
- [19] C. Zhan, S. Sankaran, V. LeMoine, J. Graybill, and D.-O. S. Mey, "Application of machine learning for production forecasting for unconventional resources," in *Presented at the Unconventional Resources Technology Conference*, Denver, July 2019.
- [20] B. Li, T. C. Billiter, and T. Tokar, "Significant error reduction in machine-learning decline curve analysis for unconventional reservoirs," in *SPE/AAPG/SEG Unconventional Resources Technology Conference*, Denver, July 2020.
- [21] C. Ferreira, "Gene expression programming: a new adaptive algorithm for solving problems," *Complex Systems*, vol. 13, no. 2, pp. 87–129, 2001.
- [22] S. Samadianfard, R. Delirhasannia, O. Kisi, and E. Agirre-Basurko, "Comparative analysis of ozone level prediction models using gene expression programming and multiple linear regression," *Geofizika*, vol. 30, no. 30, pp. 43–74, 2013.
- [23] C. Zhou, W. Xiao, T. M. Tirpak, and P. C. Nelson, "Evolving accurate and compact classification rules with gene expression programming," *IEEE Transactions on Evolutionary Computation*, vol. 7, no. 6, pp. 519–531, 2004.
- [24] N. D. Hoang and B. D. Tien, "Spatial prediction of rainfall-induced shallow landslides using gene expression programming integrated with GIS: a case study in Vietnam," *Natural Hazards*, vol. 92, no. 3, pp. 1871–1887, 2018.
- [25] J. Jedrzejowicz, P. Jedrzejowicz, and I. Wierzbowska, "Implementing gene expression programming in the parallel environment for big datasets' classification," *Vietnam Journal of Computer Science*, vol. 6, no. 2, pp. 163–175, 2019.
- [26] C. Ferreira, *Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence*, Springer, Berlin, 2006.
- [27] Y. Changan, "The research on key techniques in intelligent model base based on function finding by GEP," Sichuan University, Chengdu, 2006.