Hindawi
Geofluids
Volume 2022, Article ID 8441075, 15 pages
https://doi.org/10.1155/2022/8441075

Research Article

Development of Decline Curve Analysis Parameters for Tight Oil Wells Using a Machine Learning Algorithm

Weirong Li,[1] **Zhenzhen Dong**,[1] **John W. Lee**,[2] **Xianlin Ma**,[1] **and Shihao Qian**[1]

[1]*Xi'an Shiyou University, China*
[2]*Texas A&M University, USA*

Correspondence should be addressed to Zhenzhen Dong; dongzhenzhen1120@hotmail.com

Received 15 January 2022; Accepted 17 March 2022; Published 4 April 2022

Academic Editor: Jose Antonio Torres

To obtain a reliable production forecast, one has to establish a geological model with well logs and seismic data. The geological model usually has to be upscaled using certain upscaling techniques. Then, a dynamic reservoir model is constructed with another dataset, including completion data, production data, fluid properties, and relative permeability curves. At last, the dynamic model needs to be validated by a history matching process. This approach is data-intensive, time-consuming, and often not rigorously accomplished due to the lack of skillset and time. In this study, 10,000 groups of reservoir/completion input data were generated by Latin hypercube sampling method, and then, 10,000 groups of output (oil rate and cumulative production data) were obtained by numerical simulation. Next, a machine learning technique was applied to establish a model between the input data and determining parameters of a decline curve analysis model by fitting the generated cumulative production rate. Overall coefficients of determination (R^2) of the three Arps decline curve factors were 0.966, 0.990, and 0.945. The validation result shows that the production rate and cumulative production predicted by the proposed machine learning–decline curve analysis (ML-DCA) model agreed well with those simulated by reservoir simulation. As a result of the ML-DCA regression model, a complete understanding can be established of the impact of reservoir properties on the DCA model. The proposed ML-DCA model not only provides a quick and robust method for petroleum engineers to estimate production performance for unconventional reservoirs from reservoir and completion properties without full-field geocellular modeling but also can be used to optimize the completion and operation parameters for wells of interest.

1. Introduction

Unconventional oil and gas reservoirs have been able to be developed economically with advancements in horizontal well drilling and multistage hydraulic fracturing technology. Management and operation of unconventional oil and gas reservoirs demand accurate prediction of production rates, which facilitates better development strategy, more economic feasibility, more reliable reserve evaluations, and more informed business decisions.

Many efforts have been made to develop efficient numerical models for simulation of unconventional oil and gas production considering complex hydraulic and natural fracture geometries and multiple gas transport mechanisms in nanopores [1–4]. High-resolution, three-dimensional (3D), geocellular models characterize geological features and fracture network complexities with grid blocks and their related rock and fluid properties, and they yield a comprehensive geographic distribution of pressure and saturation over a period of time. However, such a model is computationally prohibitive as a large number of simulations are required for history matching and production optimization in a close-loop reservoir management and decision-making context. The challenges become even more discouraging with a lack of rock and fluid data and insufficient production history. Researchers have tried to accelerate simulations through techniques such as upscaling [5, 6], multiscaling [7–9], and streamline methods [10, 11]. However, all of these speed-ups require a full physics-based model as a starting point for simplification.

Decline curve analysis (DCA) as an alternative technique to forecast oil and gas production has been applied in the oil

industry for a long time. The most commonly used curve-fit-based decline curve models for unconventional oil and gas reservoirs include the multisegment Arps decline model [12–15], modified Arps DCA model [16], transient hyperbolic model [17], stretched exponential decline method (SEDM) [18], Duong method [19], and power law exponential (PLE) method [20]. It is well known that curve-fit-based DCA models are easy to apply for unconventional reservoirs once the flow regimes related to depth of investigation are appropriately identified, but they fall short of prediction accuracy as they cannot capture the reservoir features or properties that are highly relevant to production performance.

With the introduction of data analytics in the oil and gas industry, such practical challenges have caused emerging data-driven solutions in the area of data mining and machine learning (ML), in which geological data, completion data, and dynamic data are input to the ML algorithm to unravel hidden physical relationships contained in the data but not represented in existing simulation models to output future production performance. Li and Han [21] applied a neural network to study the correlation between selected reservoir/completion properties and the determining parameters of the logistic growth model (LGM). Thus, a trained ML-DCA model can be used to predict the production trends for both an existing well and a new well according to the given reservoir/completion properties. Sun et al. [22] proposed a recurrent neural network- (RNN-) based sequence-to-sequence model to forecast production. Mukherjee et al. [23] tested and performed various ML algorithms, including linear regression (LR), principal component analysis (PCA), neural network regression (NNR), boosted decision tree (BDT), and binned decision tree (BiDT), to forecast gas production. Tamimi et al. [24] presented a comprehensive intelligent decision support system (IDSS) to forecast and determine parameters of Arps decline curve model from 3,400 unconventional wells. Temizel et al. [25] applied the long short-term memory (LSTM) method for predicting long-term production behavior in unconventional shale reservoirs. Xue et al. [26] proposed a multiobjective random forest (MORF) algorithm to forecast the production rate, with reservoir and completion characteristics as input. Gross et al. [27] proposed a physics-informed ML workflow combining fast reduced-order models (ROMs) with reservoir simulation models to predict production as a function of pressure management in a fractured Marcellus shale reservoir. Doan and Vo [28] used ML techniques to enhance the accuracy of production forecasting in the North Malay Basin.

Data-driven models may not capture many geological features of a reservoir, but they run much faster than full-physics models, and the trained model on real data can avoid uncertainty or making hypothesis. They can provide reliable predictions with enough data in the calibration process. However, when the number of parameters becomes large, data-driven models require a certain number of samples for training and testing before they can be effectively used.

The unique nature of this study is using a simulation model as a basis for the production dataset, which provides an alternative for utilizing the machine learning algorithms when actual data are insufficient or unavailable. This research offers less experienced engineers with an effective ML-DCA model that correlates reservoir and fluid properties, as well as completion parameters, with the DCA model to forecast the production rates without full-field geocellular modeling for unconventional reservoirs. Therefore, without even working with complicated geomodels, less experienced engineers can make robust estimations about the production forecast once the ML-DCA model is established and delivered by experienced reservoir engineers. The implementation between DCA and the simplified reservoir model with ML algorithm makes it easy to extend the workflow for any other type of reservoirs, such as thermal or compositional.

Table 1: Cumulative production-time relationship of Arps decline curve.

Exponential $b = 0$	$Q = \frac{1}{D_i}(q_i - q_t)$
Hyperbolic or superhyperbolic	$Q = \left[\frac{q_i}{D_i(1-b)}\right]\left[1 - \left(\frac{q_t}{q_i}\right)^{1-b}\right]$
Harmonic $b = 1$	$Q = \left[\frac{q_i}{D_i}\right] \ln\left(\frac{q_t}{q_i}\right)$

This study is organized as follows. Firstly, the methods in the proposed ML-DCA workflow, including decline curve model and artificial neural network were introduced, followed by synthetic production profiles generated and established from single-well numerical simulation considering 15 geological and completion parameters that included matrix and fracture network properties for data-driven study. Next, the ML-DCA model was trained and tested by the synthetic production profiles; then, a case study was given to evaluate the prediction performance of the ML-DCA model, which shows the power and accuracy the proposed ML-DCA workflow; and a sensitivity analysis was performed to investigate the effect of geological and reservoir parameters on the production with the ML-DCA model; finally, the limitations and future work were discussed and conclusions were drawn.

2. Methodology

2.1. Decline Curve Analysis. The Arps decline curve is the most common DCA. The Arps hyperbolic decline curve model is [29].

$$q = \frac{q_i}{(1 + bD_i t)^{1/b}}, \tag{1}$$

where q is the predicted production, q_i is the initial production, t is time, b is a constant, and D_i is the initial decline rate. When the constant loss ratio b is 0, the decline curve reduces to an exponential decline; if b is 1, the decline curve becomes harmonic; if b is greater than 0 but less than 1, the decline curve is hyperbolic; and if b is greater than 1, often during transient flow, the decline curve is called superhyperbolic. When one deals with low- and ultralow-permeability wells, transient flow lasting months to years is common,

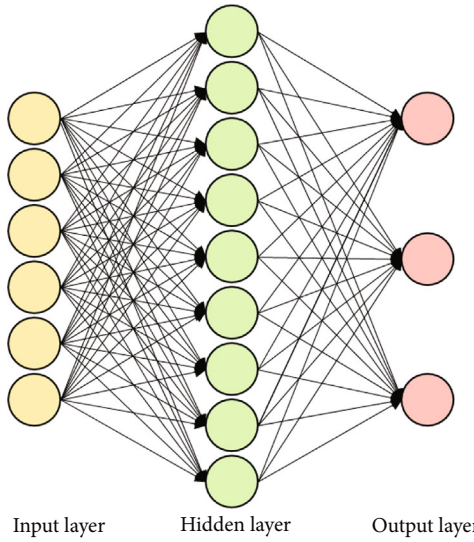


Figure 1: General ANN structure.

$$\begin{pmatrix} \theta_1^{(1)} & \cdots & k_{f1}^{(P)} \\ \vdots & \ddots & \vdots \\ \theta_N^{(1)} & \cdots & k_{fN}^{(P)} \end{pmatrix} \rightarrow \theta(Neural\ Network) \rightarrow \begin{pmatrix} (b, D_i, q_i)_1 \\ \cdots \\ (b, D_i, q_i)_N \end{pmatrix}$$

Figure 2: ML-DCA regression with neural network.

followed by a transition flow regime and later BDF. In practice, b is constant for a long time during transient flow in hydraulically fractured wells and during BDF, so one can apply the Arps model with confidence to those time periods during which b is constant [12].

In this study, cumulative oil production was selected to perform DCA regression as it is much smoother than the production rate. This approach also mitigates the effect of irregular and noisy production data, especially for real field data. Cumulative oil production is the integral of the production rate (Equation (1)), which is defined as

$$Q = \int_{t_0}^{t} q dt = \int_{t_0}^{t} \frac{q_i}{(1 + bD_i t)^{1/b}}, \tag{2}$$

where Q is the cumulative production. Integrating and rearranging Equation (2), the cumulative production-time relationship for different b values can be derived and shown as in Table 1.

2.2. Artificial Neural Network. An artificial neural network (ANN) is a common ML technique inspired by the structure of neurons in the human brain. The ANN was selected as the ML method because it has been successfully applied in many engineering and science problems to extract complex and nonlinear relationships among process variables. In this study, a typical three-layer back-propagation ANN structure was established, consisting of an input layer, a hidden layer, and an output layer, as shown in Figure 1. Several key variables, including hidden layer neurons, training algorithm, and transfer functions, were considered to get the optimal

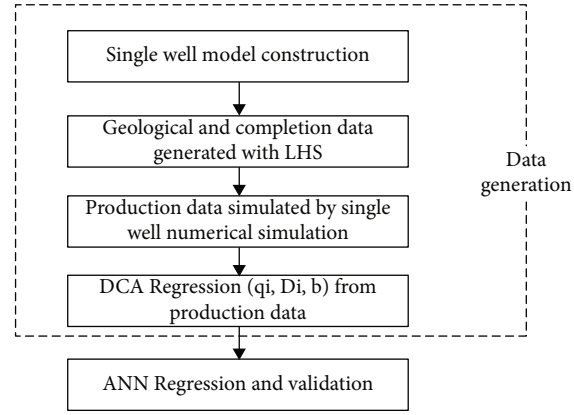


Figure 3: Flowchart of the ML-DCA model for production prediction based on ANN.

ANN structure. The neuron number in the input layer equals the input feature number. The neuron number in the hidden layer was tuned to optimize the objective function. Each neuron received the input information from the output of the neurons in the previous layer and generated and passed output to the next layer. The mathematical model of the i^{th} neuron is expressed as

$$y_i = \phi\left(\sum_{j=1}^{n} w_{ij} x_j + b_i\right), \tag{3}$$

where y_i is the output of the i^{th} neuron on the next layer, x_j is the input from the previous layer, w_{ij} is the weight, n is the input number from the previous layer, and ϕ is the activation function. The weight was tuned for each neuron through optimizing the loss function in the model training process. The R-squared score (R^2) is commonly used as the loss function. It is a statistical measure that represents the proportion of the variance of a dependent variable that is explained by an independent variable or variable in a regression model. R^2 score is given as

$$R^2 = 1 - \frac{\sum_i (\hat{y}_i - y_i)^2}{\sum_i (\bar{y}_i - y_i)^2}, \tag{4}$$

where $\hat{y}_i$ is the predicted value for i, and $\bar{y}_i$ is the mean of y_i.

2.3. Machine Learning-Decline Curve Analysis Algorithm. The ML algorithm provides a statistical technique to analyze a system with an existing dataset without being explicitly programmed. To introduce how the ML algorithm solves a regression problem, one usually defines the input vector, x_i, as

$$x_i = \left(x_i^{(1)}, x_i^{(2)}, \cdots, x_i^{(P)}\right), i = 1, \cdots, N, \tag{5}$$

where N is the sample size or the number of input vectors and P is the number of input parameters, such as permeability and

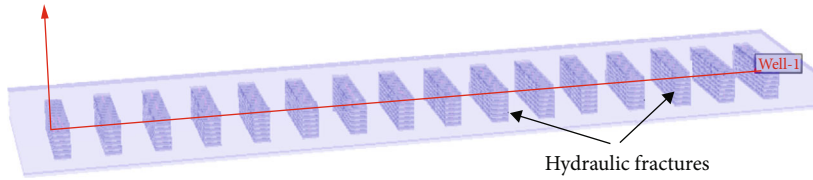


Figure 4: MFHW model to generate the simulated oil and gas production data for DCA.

Table 2: Parameters and associated distribution to generate the input dataset.

Parameters	Minimum value	Maximum value	Distribution function
Grid size, X direction(DI) (ft)	75	125	Uniform[75,125]
Grid size, Y direction(DJ) (ft)	30	80	Uniform[30,80
Grid size, Z direction(DK) (ft)	1	5	Uniform[1,5]
Matrix permeability (mD)	0.0001	1	Power[10,Random[-4,0]]
Porosity	0.05	0.15	Uniform[0.05,0.15]
Horizontal well length (ft)	1800	6000	Triangle[1800,3500,6000]
Fracture half-length (ft)	100	850	Triangle[100,380,850]
Fracture spacing (ft)	75	500	Uniform$[1, 4] \times$ DI
Effective fracture permeability (mD)	1	100	Uniform[1,100]
Layer-up	1	3	Uniform[1,3]
Layer-down	1	3	Uniform[1,3]
Monitored oil rate (bbl/day)	1.5	2.5	Triangle[1.5, 2, 2.5]
Initial reservoir pressure (psi)	2000	6000	Uniform[2000,6000]
Bubble point pressure, psi	400	6000	Initial reservoir pressure/random[1,15]
Operating BHP (psi)	200	3000	Initial reservoir pressure/random[10,30]

Note: the range of each parameter was set based on Chinese oilfield practice, such as Ordos basin [31–33].

porosity. Thus, the input matrix $\mathbf{X}$ can be defined as

$$\mathbf{X} = (x_1, x_2, \cdots x_i, \cdots, x_N)^T = \begin{pmatrix} x_1^{(1)} & \cdots & x_1^{(P)} \\ \vdots & \ddots & \vdots \\ x_N^{(1)} & \cdots & x_N^{(P)} \end{pmatrix}. \quad (6)$$

In this study, the input was a column matrix composed of reservoir, completion, and operation parameters, such as fracture half-length, fracture width, fracture permeability, and porosity.

The corresponding output parameters can be given as

$$\mathbf{Y} = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots \boldsymbol{y}_i, \cdots, \boldsymbol{y}_N)^T \quad (7)$$

where the output vector y_i is the determining parameters of the decline curve model corresponding to the input vector x_i .

$$y_i = \left(y_i^{(1)}, y_i^{(2)}, \cdots, y_i^{(t)}\right), i = 1, \cdots, N, \quad (8)$$

where t is the number of determining parameters for the decline curve model. Clearly, there are three parameters in the Arps decline curve model; thus, the output y_i is a column matrix of b, D_i, and q_i.

The existing measurement dataset was used to train the ML model. The training dataset is given as

$$(\mathbf{X}, \mathbf{Y}) = \{(x_1, y_1), (x_2, y_2), \cdots, (x_k, y_k), \cdots, (x_1, y_1)\}. \quad (9)$$

The purpose of ML is to establish a mapping function, $\varnothing(\bullet)$, from the training dataset

$$\mathbf{Y} = \varnothing(\mathbf{X}). \quad (10)$$

The proposed ML-DCA model can be expressed and established as Figure 2 with the training dataset.

$$\begin{pmatrix} \varnothing_1^{(1)} & \cdots & k_{\mathrm{f1}}^{(P)} \\ \vdots & \ddots & \vdots \\ \varnothing_N^{(1)} & \cdots & k_{fN}^{(P)} \end{pmatrix} \longrightarrow \varnothing(\text{Neural Network}) \longrightarrow \begin{pmatrix} (b, D_i, q_i)_1 \\ \cdots \\ (b, D_i, q_i)_N \end{pmatrix}. \quad (11)$$

2.4. Pearson Correlation. The most commonly used type of correlation that describes the degree of relationship between two variables is the Pearson correlation. Pearson's r measures the linear relationship between two variables, say X and Y. A correlation of 1 indicates that the data points perfectly lie on a line for which Y increases as X increases. A value of -1 also implies that the data points lie on a line;

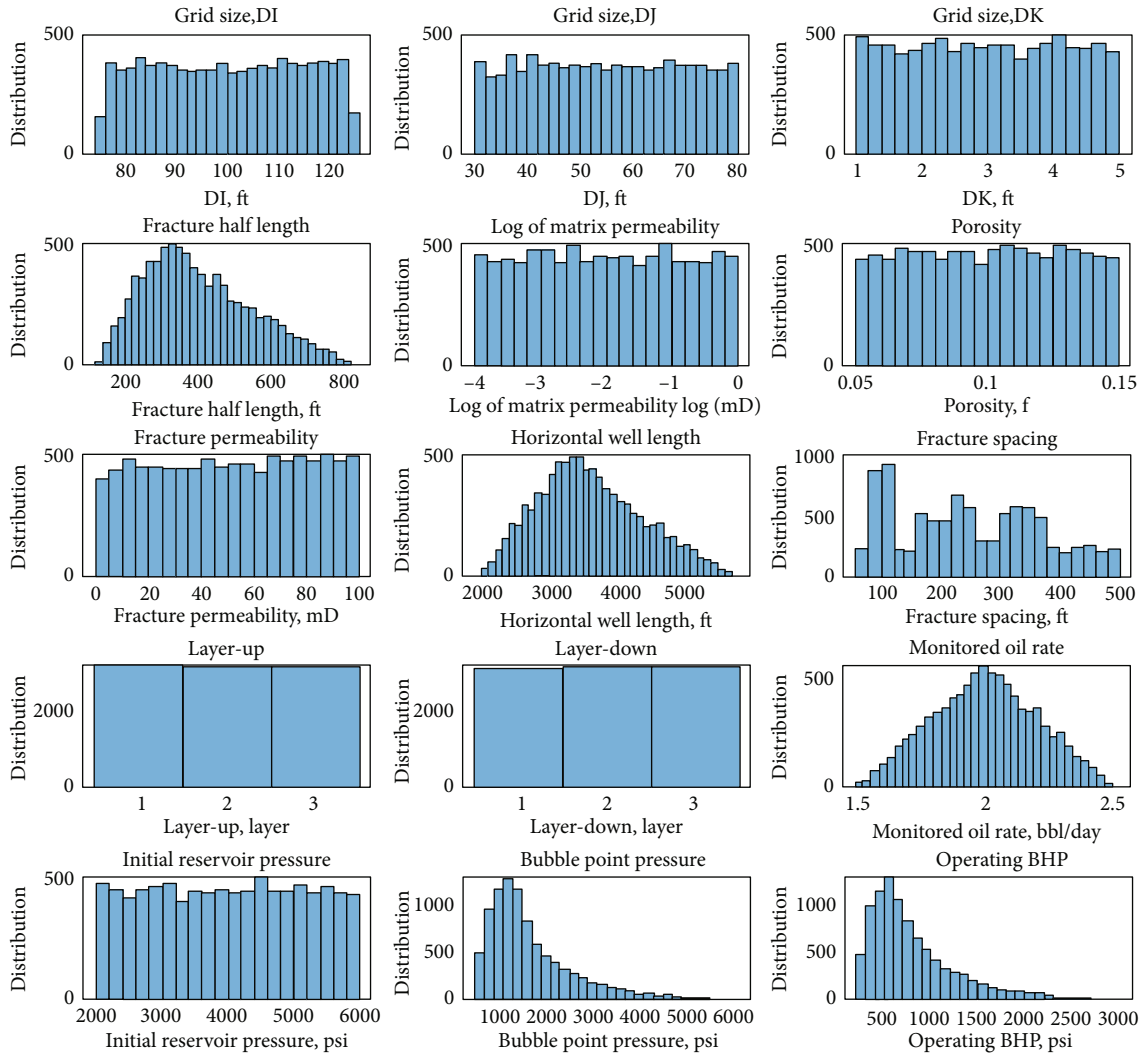


Figure 5: Histograms of geological and completion parameters generated through LHS.

however, Y decreases as X increases. The formular for r is

$$r = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \bar{Y}\right)^2}}. \quad (12)$$

Figure 3 shows the proposed workflow of ML-DCA. First, a 3D numerical model was established for a typical multistage fractured horizontal well (MFHW). Then, the Latin hypercube sampling (LHS) method was performed to generate enough (e.g., 10,000) experimental designs within the predetermined distribution type and ranges of certain input parameters, which were then used to simulate the cumulative oil production rates by running reservoir simulation. Next, the DCA regression algorithm was then performed to fit the simulated cumulative oil production rates and obtain the determining parameters of the DCA model. Finally, the ML algorithm (e.g., ANN, but it can be other algorithms in future work) was trained and tested to investigate the correlations or mapping function between geological and completion parameters and the determining parameters of the DCA model.

Note that the proposed workflow can be extended to investigate various decline curve models, such as SEDM, PLE, and Duong model, by simply replacing the decline model. In addition, all related algorithms were developed with open-source libraries that can be easily integrated with an in-house or commercial simulator.

3. Data Generation

Building a data-driven model requires a large set of geological features and completion data (production, pressure, well logs, etc.) as inputs and the factors of the DCA model as output. The ML method was then used to establish the correlation between reservoir features and production. Ideally, it is better to obtain the dataset used to build the data-driven model from the actual field. However, synthetic data generated from numerical or analytical models can alternatively be used if good-quality, actual data are insufficient or unavailable [26, 30]. In this study, we used a numerical

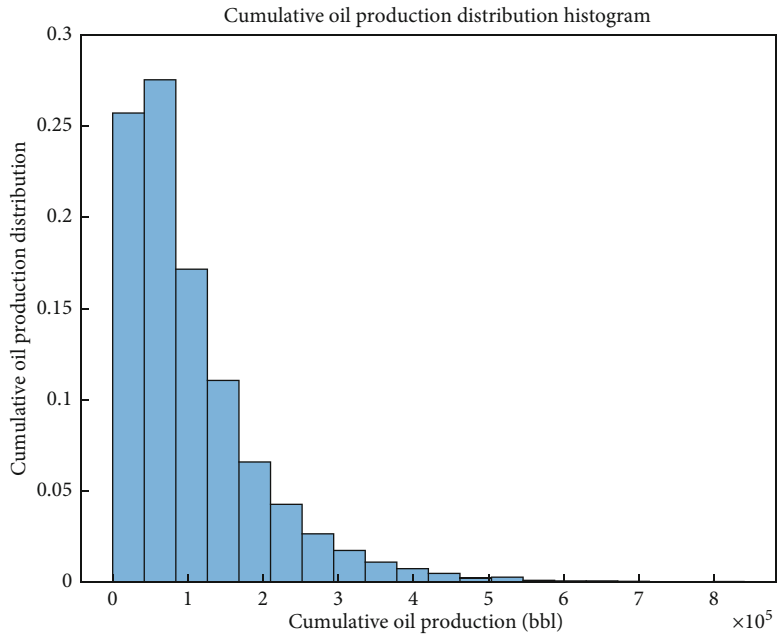


Figure 6: Distribution of 10-year cumulative oil production calculated by numerical simulation.

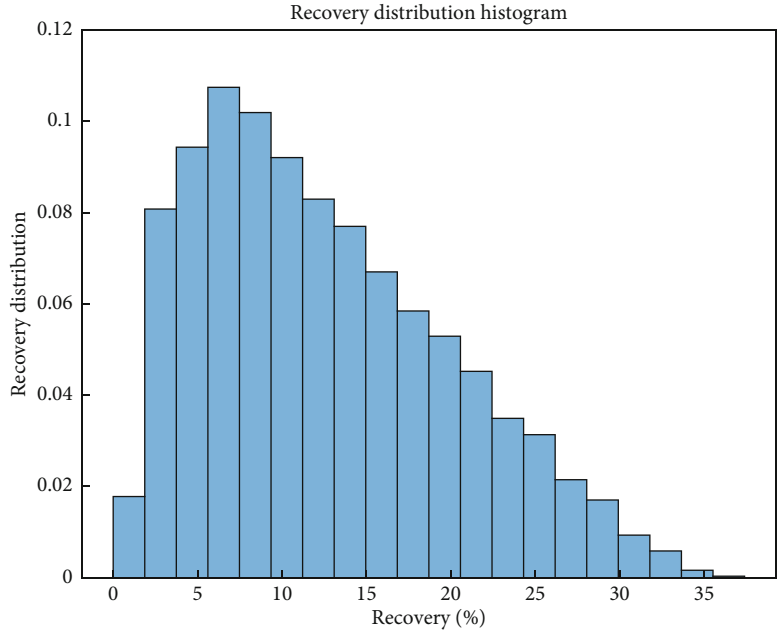


Figure 7: Distribution of 10-year recovery calculated by numerical simulation.

model to generate cumulative production profiles with randomly generated geological and completion parameters. Later, the simulated cumulative production rates were used to establish the relationship between reservoir properties and the DCA model.

3.1. Single-Well Reservoir Model. The MFHW model to simulate production from tight oil reservoirs was a three-phase, 3D rectangular model that was established with a CMG simulator. A tartan grid was used to model the MFHW as it is the best way to catch transient behavior. Otherwise, ones need to use local grid refinement (LGR), which is more time consuming. The grid number ($Nx \times Ny \times Nz$) of the model is set to be $50 \times 21 \times 7$. The grid size for each direction (DI, DJ, and DK) was one of the uncertain parameters that were generated by the sampling method later. Thus, the well control area and drainage area ($Lx = Nx \times DI$, $Ly = Ny \times DJ$, $Lz = Nz \times DK$) was changed with the grid size. The horizontal well was placed in the center of the reservoir model and produced under the constraint

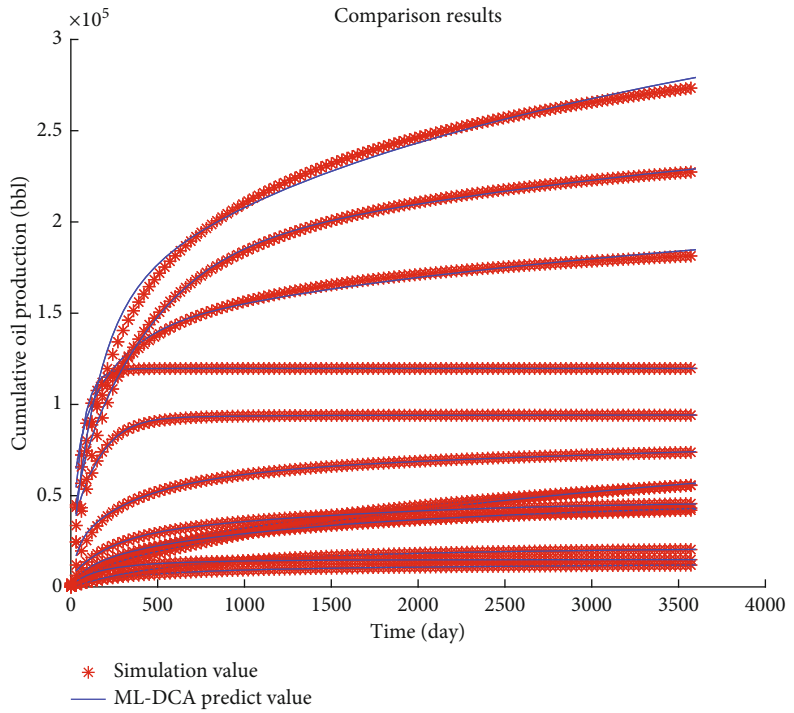


Figure 8: Part of the DCA fitting results.

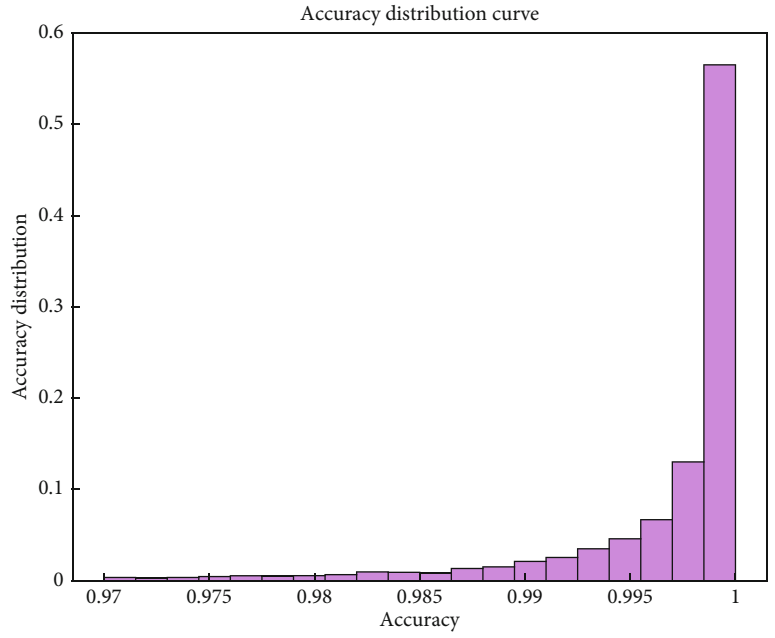
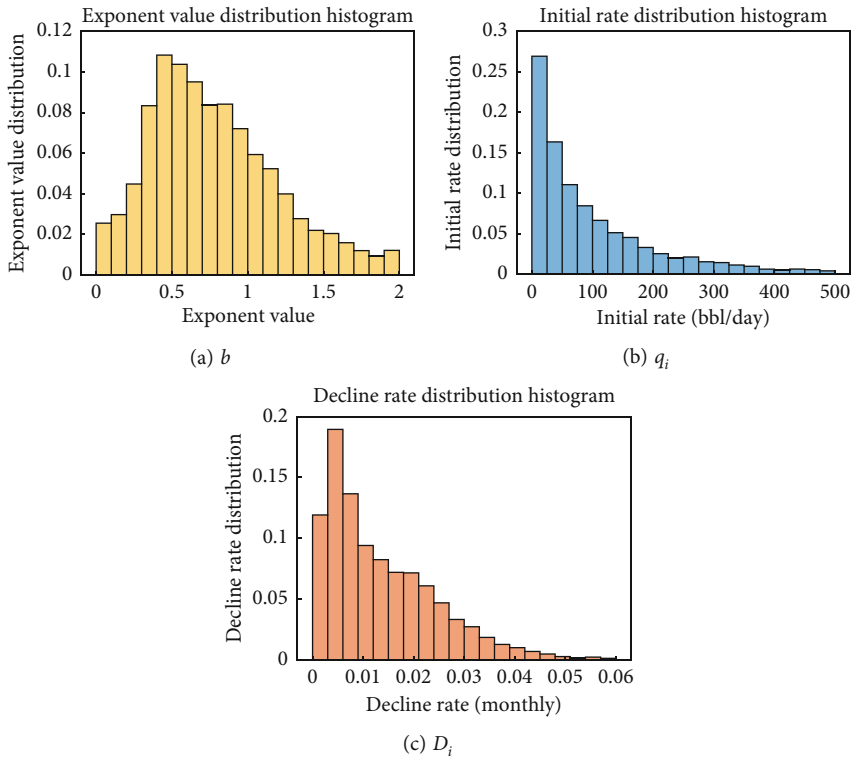


Figure 9: Distribution of R^2 of DCA regression.

of constant flowing BHP. Duration of production was set to be 10 years, which is considered a realistic tight oil production scenario. Figure 4 shows the sideview of the MFHW model, in which the darkened sections of the grid blocks represent the hydraulic fractures.

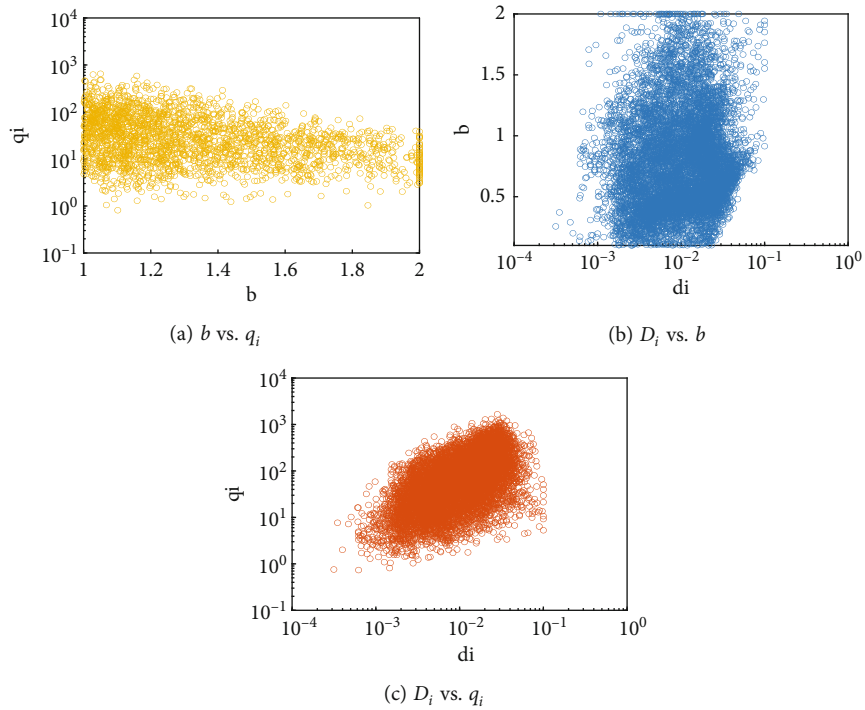
3.2. Latin Hypercube Sampling. Once the base model was established, the experimental design was carried out, varying the design variables to generate experiments that would be used to calibrate the model. In this study, 15 geological and completion parameters were investigated, including grid size in X, Y, and Z directions (DI, DJ, and DK), matrix permeability, porosity, horizontal well length, fracture half-length, fracture spacing, fracture effective permeability, layer-up, layer-down, bubble point pressure, initial pressure, monitored oil rate, and operating bottom-hole-pressure



(a) b (b) q_i

(c) D_i

Figure 10: Probability distribution of the estimated parameters of the DCA model.



(a) b vs. q_i (b) D_i vs. b

(c) D_i vs. q_i

Figure 11: Cross-plot of the fitted DCA parameters for 10,000wells.

(BHP) (Table 2). Among the 15 parameters, the first three features were used to define the drainage area. Matrix permeability and porosity were the tight oil reservoir properties. The following four features, including horizontal well length, fracture half-length, fracture spacing, and fracture effective permeability, define the horizontal well and completion parameters. The layer-up and layer-down specified the number of layers that defines the fracture penetrate above/below the horizontal well. Thus, these two parameters were related to the fracture height. Bubble point pressure was one of the

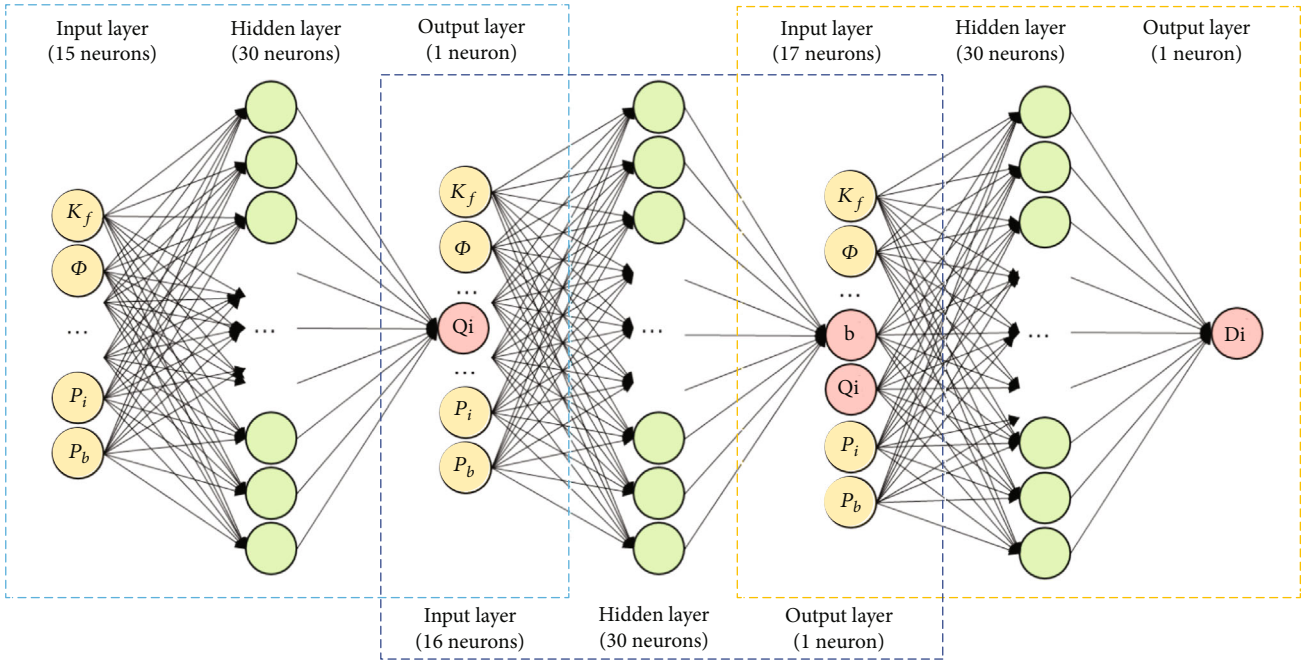


Figure 12: ANN model used in this study to estimate the three Arps parameters from geological and operational parameters.

Table 3: The comparison of overall R^2.

Variables	Sequential network	Simultaneous network
b	0.990	0.923
q_i	0.966	0.904
D_i	0.945	0.912

fluid properties, and initial pressure defined the initial reservoir condition. Bubble point pressure was set to be lower than the initial reservoir pressure for each scenario. The water saturation of the reservoir was constant in this study, and the value was set to 0.4. Relative permeability curve was predefined. The last two features were operation parameters during production. Oil production changed with the operating BHP, which was set to lower than the initial reservoir pressure. The well was shut off when the oil rate was lower than the monitored oil rate defined by well constraints. To obtain effective training and testing of data-driven models, 10,000 samples of each parameter were generated through the LHS method, with the ranges and distribution type listed in Table 2. The probability distributions of each parameter are shown in Figure 5.

3.3. Generation of Cumulative Production. The corresponding 10-year monthly oil production for each combination was then simulated by the numerical model. The distribution of the 10,000 cumulative oil productions and recoveries are displayed in Figures 6 and 7. Basically, the cumulative production of 50% of wells was less than 10×10^4 bbl. The typical oil recovery usually was less than 35%, mainly distributed between 5% and 15%.

3.4. Decline Curve Analysis Regression. To obtain the best predictive model, production data from 10,000 wells were transformed into DCA space. DCA best fit curves were usually computed with the least-squares regression [34]. Thus, least-squares regression was programmed to estimate the three determining parameters of the Arps decline model from the 10,000 synthetic cumulative production profiles generated by the represented reservoir simulation. Figure 8 shows parts of 10,000 fitting results as an example. It shows good agreement between the cumulative oil production calculated by the DCA model and the synthetic cumulative production simulated by the numerical model. Figure 9 is the distribution of coefficient of determination (R^2). 99% of the DCA regression had a R^2 greater than 0.99. In other words, the DCA curve can be a proxy model of the single-well simulation to forecast tight oil production in this study.

Figure 10 gives the histogram of the determined initial rate (q_i), b value, and initial decline rate (D_i) by the ML-DCA model. The distribution can be used for further study if one performs production uncertainty analysis with the DCA model, such as to forecast P10, P50, and P90. b value affects long-term production but does not make much difference in short-time production. Initial production rate and initial decline rate affect the short-term rate. As shown in Figure 11, there was no apparent correlation between b and initial production rate (q_i), meaning that initial production alone cannot represent how well or how poorly a well will produce in the long term. However, b showed a significant correlation with the initial decline rate (D_i) (Figure 11(b)). This observation was similar to Tamimi et al.'s [24] study based on more than 3,400 unconventional wells in DJ basin. Simply put, a higher initial decline rate indicates a higher b value. Thus, it seems a higher decline rate may indicate a bad well, but eventually it produces more in the long term.

Table 4: The optimized network structure for ML-DCA model.

Variables	q_i	B	D_i
Number of neurons in input layer	15 parameters	15 parameters and q_i	15 parameters, q_i and b
Number of neurons in hidden layer	40	50	60
Output in output layer	q_i	B	D_i
Data split		70% for training, 30% for testing	
Function performance		R^2	
Training algorithm		Bayesian regularization	
Iterations to achieve optimal structure	185	250	324
Overall R^2	0.966	0.990	0.945

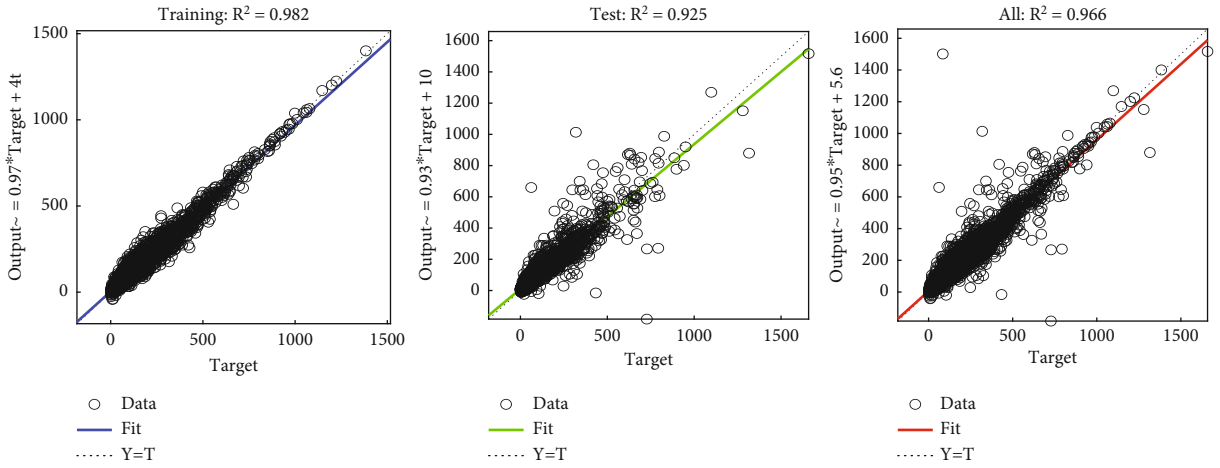


Figure 13: Regression result of parameter q_i with ANN.

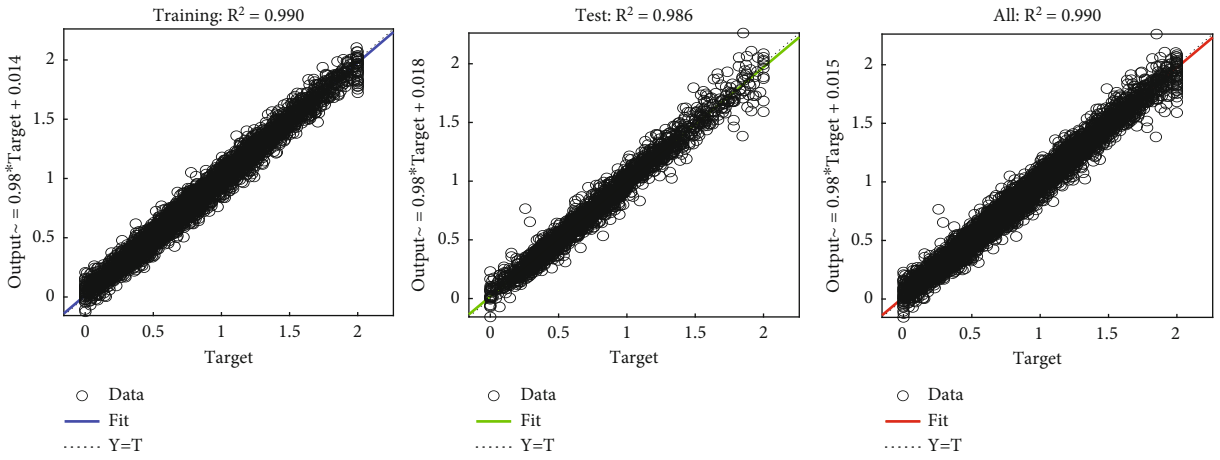


Figure 14: Regression results of parameter b with ANN.

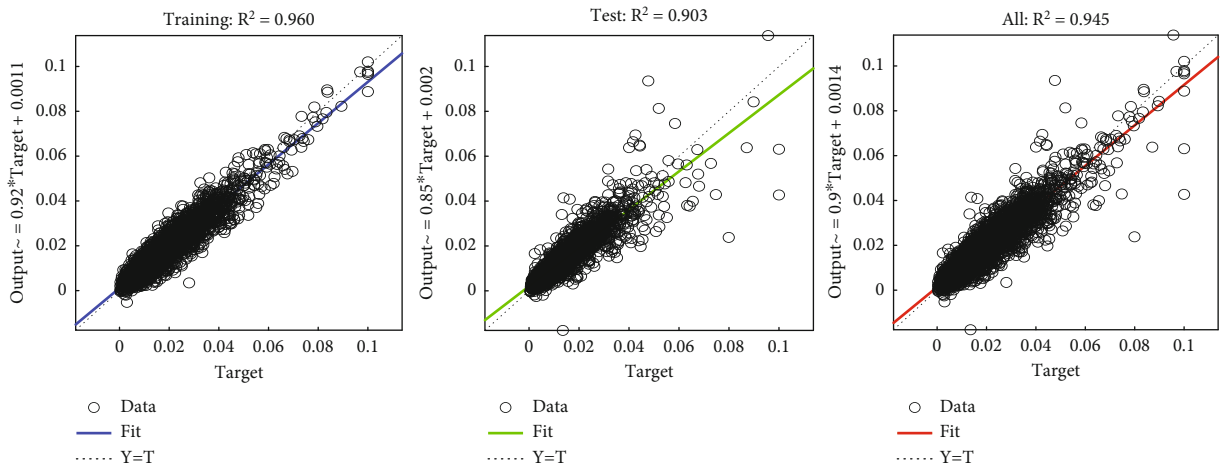


Figure 15: Regression results of parameter D_i with ANN.

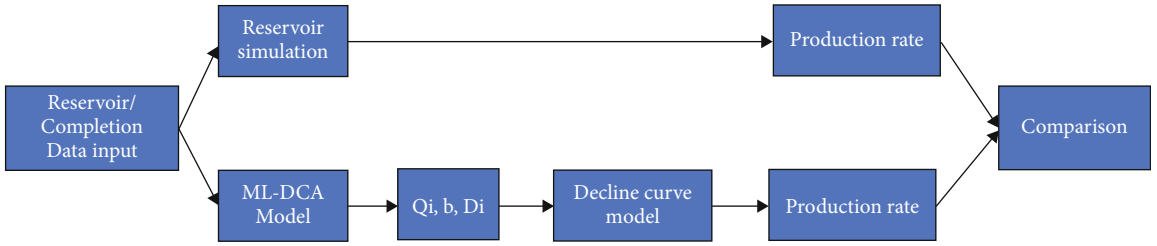


Figure 16: Flowchart to evaluate the production prediction performance.

Table 5: Reservoir features used as inputs for performance validation.

Feature	Value	Units	Feature	Value	Units
Grid size, X direction	84.08	ft	Fracture effective permeability	52	mD
Grid size, Y direction	57.43	ft	Layer-up	2	Layer
Grid size, Z direction	3.21	ft	Layer-down	1	Layer
Matrix permeability	0.0004	mD	Bubble point pressure	845	psi
Porosity	0.08		Initial pressure	3767	psi
Well length	2775	ft	Monitored oil rate	2	bbl/day
Fracture half-length	546	ft	Operating BHP	699	psi
Fracture spacing	252	ft			

4. Results and Analysis

4.1. Neural Network Regression. In this study, three-layer back-propagation neural network was established, as shown in Figure 12. As it can been seen, the three factors of the DCA model were solved sequentially rather than simultaneously, as the sequential network shown as Figure 12 achieved the highest overall R^2 and proven to be the optimal regression workflow to determine the factors of the DCA model (Table 3).

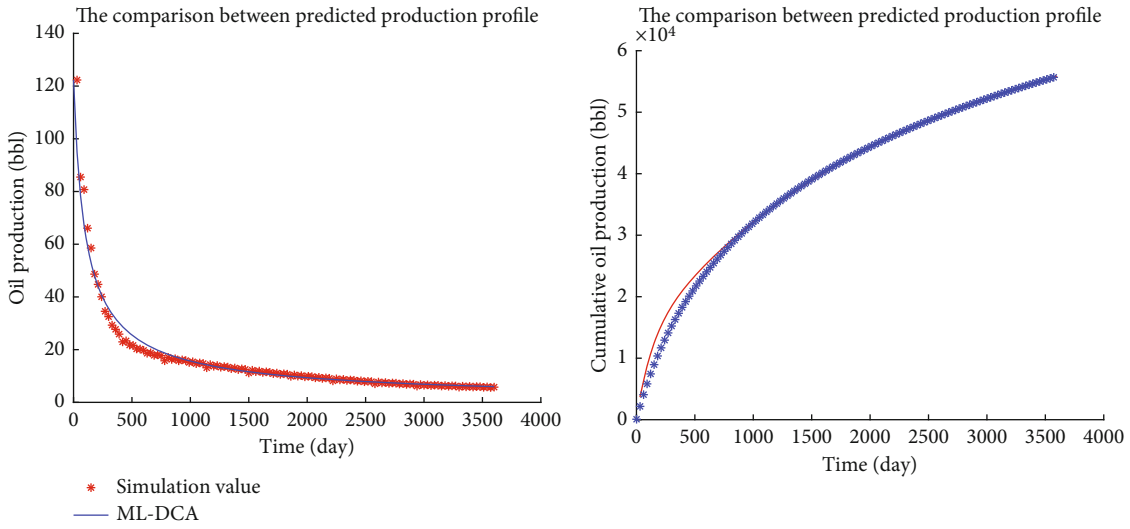
The optimal regression workflow is summarized as follows:

(i) Step 1: initially, there are 15 neurons in the input layer and 1 neuron in the output layer

(ii) Step 2: Bayesian regularization is selected as the training algorithm. Thus, in the 10,000 combinations of input parameters and corresponding cumulative production profiles, 70% of them are used as the training dataset to establish the data-driven model, and 30% of them are used as the testing dataset to determine the DCA factors

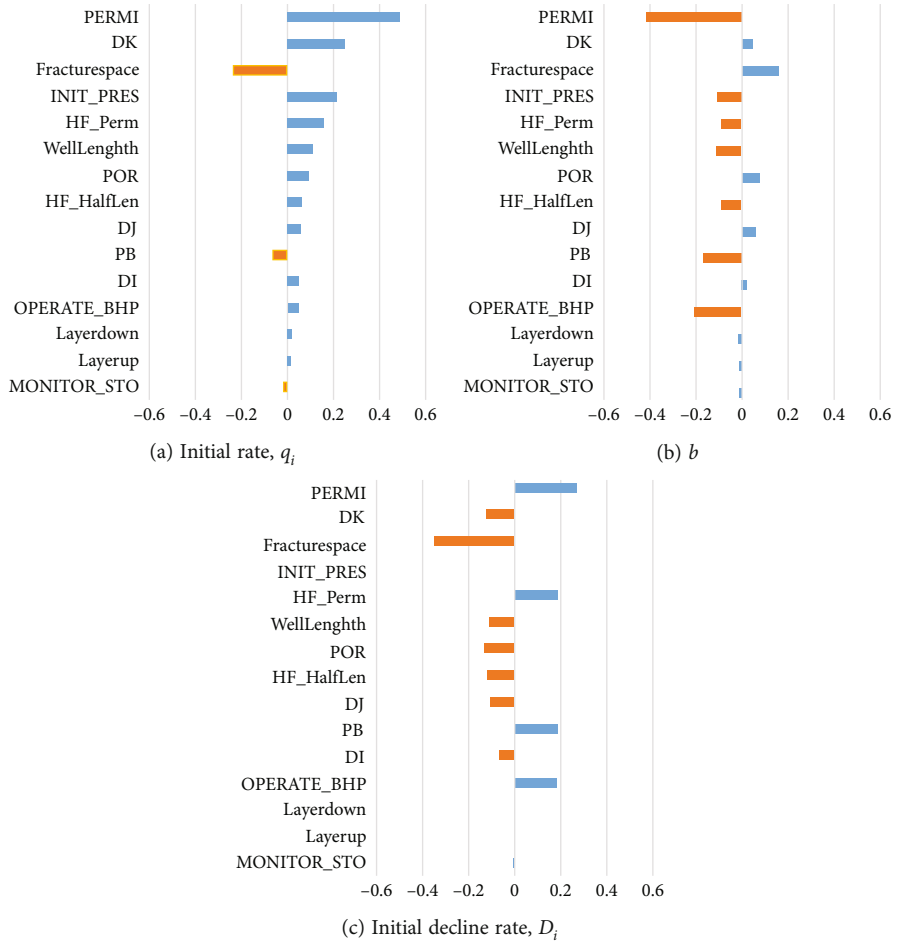
(iii) Step 3: by changing the number of neurons in the hidden layer, the different results for the target DCA factor are obtained

(iv) Step 4: Bayesian regularization, which minimizes a linear combination of squared errors, is performed to modify the connective weights and biases to achieve more accurate results [35]



(a) Oil production (bbl/day) (b) Cumulative oil production (bbl)

Figure 17: Comparison between the production simulated by the numerical model and predicted by ML-DCA algorithm.



(a) Initial rate, q_i (b) b

(c) Initial decline rate, D_i

Figure 18: Sensitivity analysis results of the reservoir and geological parameters on the decline curve model.

(v) Step 5: the ANN's training process continues interactively until the desired level of error or maximum iteration number is reached, and the first final network used to determine the vector of q_i is achieved.

(vi) Step 6: the vector of q_i is added to the input matrix, and Step 2 to Step 5 are repeated to obtain the vector of b

(vii) Step 7: the vector of b is added to the input matrix, and Step 2 to Step 6 are repeated to determine the vector of D_i

The optimum network was achieved and is given in Table 4.

Figures 13–15 show R^2 of the training set, the testing set, and all sets of the three determining parameters predicted by the proposed ML-DCA model, respectively. The axes of the abscissa and ordinate are the actual determining parameters of Arps decline curves and their predicted values, respectively. The smaller the difference between the actual value and the predicted value, the closer the data point is to the 45-degree line. The data in Figures 13–15 (both the training set and testing set) is densely distributed near the 45-degree line indicating the high prediction accuracy of the ML-DCA model. The overall prediction errors of the three factors (q_i, b, and D_i) are 0.966, 0.990, and 0.945, respectively. With the trained and validated ML-DCA model, one can easily and quickly output a decline curve model to forecast the production rate of new wells by only knowing the reservoir and completion parameters.

4.2. Prediction Performance Evaluation. Figure 16 shows the process to evaluate the prediction performance of the ML-DCA model by comparing the production profiles with numerical simulation. Table 5 presents the values of the 15 selected key features used to forecast the production profile of a tight oil well. The comparison result of the well is depicted in Figure 17. Overall, both the production rate and cumulative production between these two methods agreed with each other very well, indicating that the performance of the ML-DCA model is acceptable and reliable for science and engineering applications. It can be concluded that the ML-DCA model has the same accuracy as single-well numerical simulation. Thus, instead of establishing a complex numerical model, engineers who have less experience with numerical simulation can use the ML-DCA model as a proxy model to forecast tight oil production with low cost.

4.3. Sensitivity Analysis. Pearson correlation coefficient was calculated to evaluate the significance of each property in the three determining parameters (q_i, b, and D_i) (Figure 18). Figure 18(a) reflects permeability, grid size in the vertical direction (DK), and fracture spacing to be the top three factors having the most significant effect on the initial rate, followed by initial pressure, fracture permeability, well length, matrix porosity, fracture half-length, etc. Among them, fracture spacing had a significant negative impact on the initial rate. And operating BHP, monitored oil rate, layer-up, and layer-down had a minor effect on the initial rate.

Figure 18(b) indicates that permeability has mainly a negative influence on b. The higher the permeability, the lower the b value. It also can be concluded that operating BHP, bubble point pressure, well length, initial pressure, fracture half-length, and fracture permeability had a negative impact on b value, while fracture spacing, porosity, and grid size had a positive effect on b value.

Figure 18(c) shows the correlation between the initial decline rate (D_i) and reservoir and completion parameters. Fracture spacing, used to define the number of fractures in the reservoir model, had a prominent influence on the initial decline rate D_i. Meanwhile, the grid size (DK, DI, and DJ), well length, porosity, and fracture half-length had a negative influence. Results also indicate matrix permeability, initial pressure, fracture permeability, bubble point pressure, and operating BHP to have a positive impact on the initial decline rate.

5. Discussion and Future Work

It is well known that ML-DCA can be powerful if the data quantity and quality can be improved by including actual field data. However, with most tight oil wells having less than 60 months of production history, these data cannot be directly used to perform history matching. This study proves that this simulation-based proxy tool is reliable with a well-maintained database generated from a single-well reservoir simulation, and its efficiency in computational time allows the practicing engineer to achieve modeling objectives and to reduce uncertainty in a rapid way. As a result, the ML-DCA algorithm not only can be used as a tool to determine DCA factors and predict production, once the initial reservoir conditions, rock properties, and completion and operation parameters are given, but also can be used to optimize the completion and operation parameters for a target reservoir, such as fracture spacing, fracture half-length, and operating BHP. As for future work, further studies can be conducted to investigate the effect of measurement errors and sample size on production prediction performance.

6. Conclusions

In this study, a ML-DCA algorithm, integrating a ML technique with a DCA model, was developed to predict tight oil production performance, serving as a proxy for analytical/numerical reservoir simulation. The following critical conclusions can be summarized:

(1) With the reservoir and completion parameters as the inputs and the DCA factors as outputs, the ML-DCA model can be trained to determine DCA factors accurately, with the overall prediction errors (R^2) of the three Arps decline curve factors being 0.966, 0.990, and 0.945

(2) The production rate and cumulative production predicted by the proposed ML-DCA model agree well with those simulated by reservoir simulation

(3) ML-DCA outperforms traditional DCA methods, especially for a new well or reservoir, as ML-DCA captures the production trend of the training dataset, as well as considering the rock, fluid, completion, and wellbore properties

(4) The proposed ML-DCA outperforms full-field numerical simulation due to its simplicity and low cost

(5) As a result of sensitivity analysis, the reservoir, completion, and operation parameters that affect the three DCA factors can be determined. Among them, fracture spacing and matrix permeability have an essential effect on the DCA model, while the monitored oil rate, layer-up, and layer-down have a minor effect on the DCA model

Nomenclature

b: Arps hyperbolic or superhyperbolic decline exponent, dimensionless
D_i: Initial decline rate in the Arps decline model, D^{-1}
N: Sample size or the number of input vectors in the ML-DCA model
n: Input number from the previous layer in the ANN model
P: Number of input parameters in the ML-DCA model
Q: Cumulative production rate, MSCF or BBL
q: Predicted production rate, MSCF/D or BBL/D
q_i: Initial production in the Arps decline model, MSCF/D or BBL/D
t: Production time, days
x_i: Input from the previous layer of ANN model
y_i: Output of the ith neuron on the next layer in an ANN model
$\widehat{y}_i$: Predicted value
$\bar{y}_i$: Mean value of y_i
w_{ij}: Weight of ANN model.

Greek variables

ϕ: Activation function in the ANN model
μ: Mean value
σ: Standard deviation.

Data Availability

Data is available on request by reaching @dongzhenzhen1120@hotmail.com.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We would like to thank the Project "Shale Oil Development Study of Chang7 Panke Field" and Project "Fracturing Design Optimization of Multistage Fractured Horizontal Wells in the Lower Temple Bay Field, Yanchang Oilfield" for their support and valuable discussion.

References

[1] W. Li, Z. Dong, and G. Lei, "Integrating embedded discrete fracture and dual-porosity dual-permeability methods to simulate fluid flow in shale oil reservoirs," *Energies*, vol. 10, no. 10, p. 1471, 2017.

[2] B. K. Tiwari, S. Al-Sayegh, H. Al-Muraikhi, P. Kumar, P. V. Cueille, and F. Lislaud, "Modelling hydraulic fractures in a full-field dynamic modeling using DPDP simulation techniques-an unconventional approach applied in a tight carbonate oil reservoir, Kuwait, Middle-East," in *Paper SPE-195156-MS presented at the SPE Middle East Oil and Gas Show and Conference*, Manama, Bahrain, 2019.

[3] M. Vo, S. Su, J. Lv, and C. Xiao, "Reservoir modeling and production history matching in a Triassic naturally fractured carbonate reservoir in Sichuan, China," *Improved Oil and Gas Recovery*, vol. 4, no. 2020, pp. 1–14, 2020.

[4] W. Yu, K. Wu, M. Liu, K. Sepehrnoori, and J. Miao, "Production forecasting for shale gas reservoirs with nanopores and complex fracture geometries using an innovative non-intrusive EDFM method," in *Paper SPE-191666-MS presented at the SPE Annual Technical Conference and Exhibition*, Dallas, Texas, USA, September 2018.

[5] E. Alfataierge, E. Chesnokov, and Y. Gorb, "Introduction of upscaling methods derived from the simple averaging method and a comparison with the backus method of upscaling," in *Paper presented at SEG International Exposition and 86th Annual Meeting*, Dallas, Texas, USA, 2016.

[6] M. S. Islam and T. Manzocchi, "The inclusion of two-phase fault rock properties into unscaled models: a novel flow-based geometrical upscaling approach," in *Paper SPE-187228-MS presented at the SPE Annual Technical Conference and Exhibition*, San Antonio, Texas, USA, 2017.

[7] M. Correia and D. Schiozer, "Special connections for representing multiscale heterogeneities in reservoir simulation," in *Paper SPE-200572-MS presented at the SPE Europe featured at 82nd EAGE Conference and Exhibition*, Amsterdam, Netherlands, 2020.

[8] R. Moraes, R. M. Fonseca, M. Helici, A. W. Heemink, and J. D. Jansen, "Improving the computational efficiency of approximate gradients using a multiscale reservoir simulation framework," in *Paper SPE-182620-MS presented at the SPE Reservoir Simulation Conference*, Montgomery, Texas, USA, 2017.

[9] Y. Yang, M. Ghasemi, E. Gildin, Y. Efendiev, and V. Calo, "Fast multiscale reservoir simulations with POD-DEIM model reduction," *SPE Journal*, vol. 21, no. 6, pp. 2141–2154, 2016.

[10] A. Datta-Gupta and M. King, "Streamline stimulation: theory and practice," in *SPE Textbook Series, 11*, Society of Petroleum Engineers, 2007.

[11] S. Tanaka, D. Kam, A. Datta-Gupta, and M. King, "Streamline-based history matching of arrival times and bottomhole pressure data for multicomponent compositional system," in *Paper SPE-174750-MS presented at the SPE Annual Technical Conference and Exhibition*, Houston, Texas, USA, 2015.

[12] J. Lee, "Establishing the basis for multi-segment Arps decline models," in *Paper URTec-5393 Presented at the Unconventional Resources Technology Conference*, Houston, Texas, USA, 2021.

[13] M. F. Tugan and R. Weijermars, "Improved EUR prediction for multi-fractured hydrocarbon wells based on 3-segment DCA: implications for production forecasting of parent and child wells," *Journal of Petroleum Science and Engineering*, vol. 187, no. 2019, article 106692, 2019.

[14] M. F. Tugan and R. Weijermars, "Variation in *b*-sigmoids with flow regime transitions in support of a new 3-segment DCA method: Improved production forecasting for tight oil and gas wells," *Journal of Petroleum Science and Engineering*, vol. 192, no. 2020, article 107243, 2020.

[15] M. F. Tugan and R. Weijermars, "Searching for the root cause of shale well rate variance: highly variable fracture treatment response," *Journal of Petroleum Science and Engineering*, vol. 192, no. 2022, article 109919, 2022.

[16] I. Gupta, C. Rai, C. Sondergeld, and D. Devegowda, "Variable exponential decline: modified Arps to characterize unconventional-shale production performance," *SPE Reservoir Evaluation & Engineering*, vol. 21, no. 4, pp. 1045–1057, 2018.

[17] D. S. Fulford and T. A. Blasingame, "Evaluation of time-rate performance of shale wells using the transient hyperbolic relation," in *SPE-167242-MS presented at the SPE unconventional resources conference Canada*, Alberta, Canada, 2013.

[18] P. P. Valko and W. J. Lee, "A better way to forecast production from unconventional gas wells," in *SPE-134231-MS presented at SPE Annual Technical Conference and Exhibition*, Florence, Italy, 2010.

[19] A. N. Duong, "An unconventional rate decline approach for tight and fracture-dominated gas wells," in *SPE-137748-MS presented at Canadian Unconventional Resources and International Petroleum Conference*, Calgary, 2010.

[20] D. Ilk, J. A. Rushing, A. D. Perogo, and T. A. Blasingame, "Exponential vs. hyperbolic decline in tight gas sands: understanding the origin and implications for reserve estimates using Arps' decline curves," in *Paper SPE-116731-MS presented at the SPE Annual Technical Conference and Exhibition*, Denver, Colorado, USA, 2008.

[21] H. Li and Y. Han, "Decline curve analysis for production forecasting based on machine learning," in *Paper SPE-189205-MS presented at the SPE symposium: production enhancement and cost optimization*, Kuala Lumpur, Malaysia, 2017.

[22] 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 *Paper SPE-190104-MS Presented at the SPE Western Regional Meeting*, Garden Grove, California, USA, 2018.

[23] T. Mukherjee, T. Burgett, T. Ghanchi, C. Donegan, and T. Ward, "Predicting gas production using machine learning methods: a case study," in *Paper SEG-2019-3215692 presented at SEG International Exposition and Annual Meeting*, San Antonio, Texas, USA, 2019.

[24] N. Tamimi, S. Samani, M. Minaei, and F. Harirchi, "An artificial intelligence decision support system for unconventional field development design," in *Paper URTeC-2019-249 presented at the Unconventional Resources Technology Conference*, Denver, CO, USA, 2019.

[25] C. Temizel, C. H. Canbaz, O. Saracoglu et al., "Production forecasting in shale reservoirs through conventional DCA and machine/deep learning methods," in *Paper URTEC-2020-2878-MS presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference*, Virtual, July 2020.

[26] L. Xue, Y. Liu, Y. Xiong, Y. Liu, X. Cui, and G. Lei, "A data-driven shale gas production forecasting method based on the multi- objective random forest regression," *Journal of Petroleum Science and Engineering*, vol. 196, no. 2021, pp. 107801–107813, 2021.

[27] M. R. Gross, J. D. Hyman, S. Srinivasan et al., "A physics-informed machine learning workflow to forecast production in a fractured Marcellus shale reservoir," in *Paper URTEC-2021-5644-MS presented at the SPE/AAPG/SEG Unconventional Resources Technology Conference*, Houston, Texas, USA, 2021.

[28] T. Doan and M. Vo, "Using machine learning techniques for enhancing production forecast in north Malay Basin," *Improved Oil and Gas Recovery*, vol. 5, no. 2021, pp. 1–7, 2021.

[29] J. J. Arps, "Analysis of decline curves," *Transactions of the AIME*, vol. 160, pp. 228–247, 1945.

[30] B. Kulga, E. Artun, and T. Ertekin, "Development of a data-driven forecasting tool for hydraulically fractured, horizontal wells in tight-gas sands," *Computational Geosciences*, vol. 103, pp. 99–110, 2017.

[31] Z. Lei, S. Wu, T. Yu et al., "Simulation and optimization of CO2 huff-n-puff processes in tight oil reservoir: a case study of Chang-7 tight oil reservoirs in Ordos Basin," in *Paper SPE-191873-MS presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition*, Brisbane, Australia, 2018.

[32] R. Zhang, W. Lv, Y. Zhou, B. Qu, Y. He, and X. Ma, "Research on oil-CO2-water relative permeability of the low permeability reservoir based on history matching," *Improved Oil and Gas Recovery*, vol. 2021, no. 5, pp. 1–12, 2020.

[33] Z. Dong, W. Tian, Y. Yang et al., "Assisted history-matching based on multiobjective surrogate reservoir model for tight oil reservoirs," *IOP Conference Series Earth and Environmental Science*, vol. 781, no. 2, article 022086, 2021.

[34] M. Paryani, O. O. Awoleke, M. Ahmadi, C. Hanks, and R. Barry, "Approximate Bayesian computation for probabilistic decline-curve analysis in unconventional reservoirs," *SPE Reservoir Evaluation & Engineering*, vol. 20, no. 2, pp. 478–485, 2017.

[35] M. Sayyafzadeh, M. Haghighi, and J. N. Carter, "Regularization in history matching using multi-objective genetic algorithm and Bayesian framework (SPE 154544)," in *Paper SPE-154544-MS Presented at the SPE Europec/EAGE Annual Conference*, Copenhagen, Denmark, 2012.