ACS OMEGA

http://pubs.acs.org/journal/acsodf

Article

Data-Driven Proxy Model for Forecasting of Cumulative Oil Production during the Steam-Assisted Gravity Drainage Process

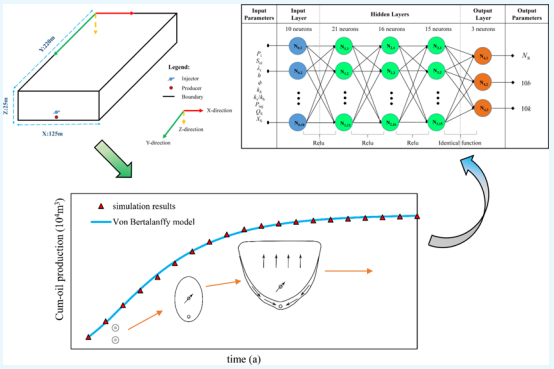
Yang Yu,* Shangqi Liu, Yang Liu, Yu Bao,* Lixia Zhang, and Yintao Dong

Cite This: *ACS Omega* 2021, 6, 11497−11509

Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: The purpose of this study is to develop a data-driven proxy model for forecasting of cumulative oil (Cum-oil) production during the steam-assisted gravity drainage process. During the model building process, an artificial neural network (ANN) is used to offer a complementary and computationally efficient tool for the physics-driven model, and the von Bertalanffy performance indicator is used to bridge the physics-driven model with the ANN. After that, the accuracy of the model is validated by blind-testing cases. Average absolute percentage error of related parameters of the performance indicator in the testing data set is 0.77%, and the error of Cum-oil production after 20 years is 0.52%. The results illustrate that the integration of performance indicator and ANN makes it possible to solve time series problems in an efficient way. Besides, the data-driven proxy model could be applied to fast parametric studies, quick uncertainty analysis with the Monte Carlo method, and average daily oil production prediction. The findings of this study could help for better understanding of combination of physics-driven model and data-driven model and illustrate the potential for application of the data-driven proxy model to help reservoir engineers, making better use of this significant thermal recovery technology for oil sands or heavy oil reservoirs.



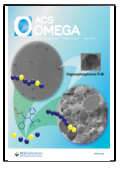
1. INTRODUCTION

The viscous oil is an issue of global importance. As conventional oil resources are depleted, continuous demand for fossil fuels has been promoting the production from unconventional reservoirs with viscous oil during the last few decades.[1,2] In the past, a large quantity of oil sands or heavy oil reservoirs, such as MacKay River oil sands in Canada and Fengcheng extra-heavy oil development area in China, which are difficult to exploit were discovered around the world.[3,4] The steam-assisted gravity drainage (SAGD) process, an effective thermal technique to exploit oil sands/heavy oil reservoirs, has a higher recovery factor than traditional thermal recovery approaches (for instance, cyclic steam stimulation or steam flooding) in general. Also, with the development of the SAGD technology (as depicted in Figure 1), large-scale commercial applications have been realized all over the world.[5,6]

Although the concept of the SAGD process seems quite simple, it is a multiphysical process involving simultaneous heat and mass transfer in reality.[7−9] So that, conventional approaches, such as empirical formula method and analytic productivity formula method, cannot accurately predict the SAGD performance. Recently, reservoir numerical simulation is an effective way to predict the performance for full life cycle of the SAGD process, once adequate inputs are provided.[10] It is one of effective physics-driven modeling methods and is considered as a dependable researching tool in the reservoir engineering field.[11] However, once the reservoir numerical simulation model is complex, the running process will be very time-consuming; meanwhile, the storage requirement of big data raises additional challenges for complicated simulation models.[12] So it is of great value building a more efficient proxy model to meet the requirements of today's fast-paced application scenarios. This is an effort to establish the workflow which could use the prepared data sets to construct the proxy model and offer accurate forecasts at less computational and storage costs.

A data-driven proxy model, an alternative model to a physics-driven model, starts to arouse extensive concern as a result of its capability to learn and memorize throughly the training process with appropriate data sets. Data-driven methods have great potential in the oil and gas industry, and the scope of its application covers upstream and downstream fields which include exploration and development, storage and

Received: February 2, 2021
Accepted: April 9, 2021
Published: April 21, 2021



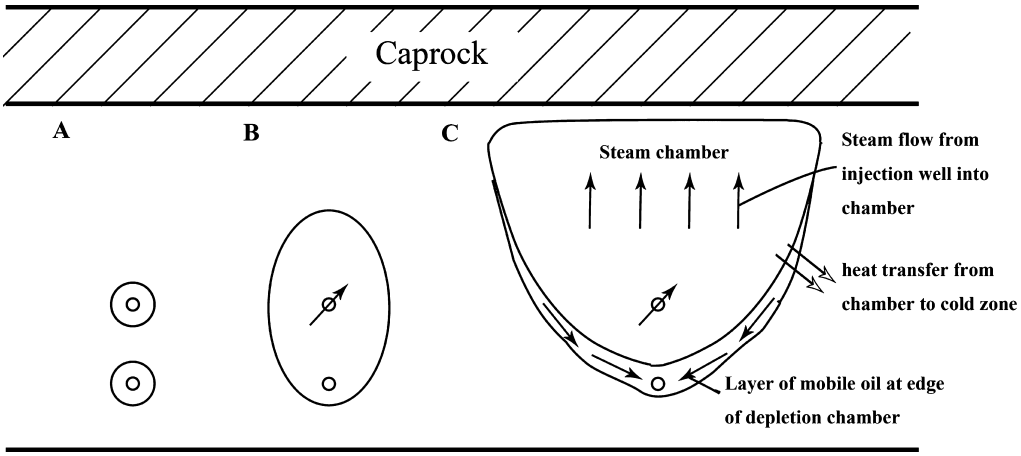


Figure 1. Cross-sectional view of the typical SAGD process: (A) preheating phase, (B) ramp up phase, and (C) lateral expanding phase (modified from Irani et al.[13]).

transportation, and so forth. Especially, many scholars utilized various kinds of data-driven methods to complete performance forecasting tasks in the reservoir engineering field. Gupta et al. (2014) provided a workflow that uses the power of the autoregressive integrated moving average (ARIMA) model to forecast production of shale gas reservoirs.[14] Kulga et al. (2017) proposed an artificial neural network (ANN)-based forecast model to predict daily gas production from tight-gas sand formation and found that the ANN model has a good performance.[15] Amirian et al. (2018) employed artificial and computational intelligence (ACI)-based learning algorithms to realize performance forecasting for polymer flooding in heavy oil reservoirs.[16] Sagheer et al. (2019) built a deep long short-term memory (LSTM) network, in order to solve time series prediction problem of petroleum production.[17] Negash and Yaw (2020) established an ANN model to forecast production of a hydrocarbon reservoir under water injection.[18] Xue et al. (2020) built a data-driven proxy model based on the multiobjective random forest method to forecast dynamic behavior of shale gas production.[19] Zhong et al. (2020) proposed a deep convolutional generative neural network (CDC-GAN)-based data-driven proxy model to predict the field oil production of reservoir developed by the waterflooding technology.[20] Deng and Pan (2021) designed and implemented the echo state network (ESN)-based data-driven proxy model to complete predicting tasks for waterflooding fields.[21]

Aforementioned works reveal that the data-driven proxy model could provide a powerful tool for solving performance forecasting problem and most of them involve time series changes. Several methods, including ANN, support vector machine, random forest regression, and their variants, can be used to construct a data-driven proxy model. Among them, ANN is the most popular approach to solve various forecasting problems with time series.[22−26] As is known to all, the ANN could be roughly divided into two categories, that is, feedforward neural network and feedback neural network. Most of the forecasting problems with time series were solved through feedback neural networks, such as Elman neural network, recurrent neural network, and LSTM neural network.[27−29] The feedback neural network, however, has a more complicated network structure than the feedforward neural network as a result of bi-directional transmission, self-circulation, memory, or other functions. The feedforward neural network, by contrast, is widely used in the performance forecasting field without time series problem for easy intelligibility and accessibility.

In the performance forecasting field of reservoir engineering, many scholars have utilized feedforward neural network, support vector machine, random forest regression, or their variants to complete prediction tasks, and great results have been obtained. However, these tasks usually only involve non-time series problems, such as recovery prediction at a given time step. Forecasting of the cumulative oil (Cum-oil) production profile during the SAGD process is a kind of time series problem (as shown in Figure 2). Throughout the literature review, when it comes to time series problems of performance forecasting, a lot of previously performed studies choose to employ some data-driven methods that can directly solve time series problems or the combination of multiple data-driven methods. Thus, the complexity of the data-driven model and the difficulty of its application are increased. Hence, it is necessary to explore a convenient approach to forecast the Cum-oil production profile during the SAGD process.

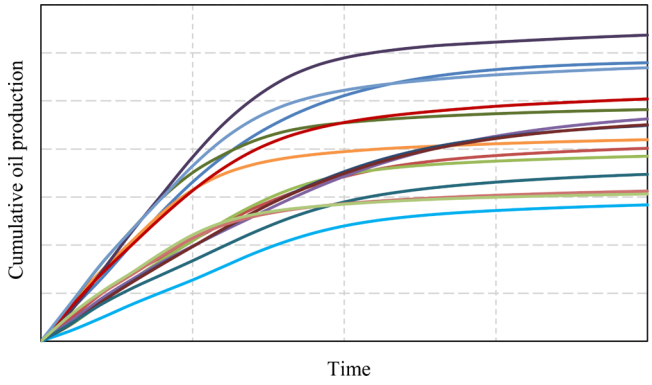


Figure 2. General sketch of Cum-oil production forecasting problem.

This paper focuses on the establishment of the data-driven proxy model which can take full advantage of the feedforward neural network, instead of using more complex neural networks or other methods. Also, the data-driven proxy model can accurately and efficiently predict the Cum-oil production changes with time during the SAGD process under the application of an appropriate knowledge-based performance indicator. The integration of selected reservoir model,

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

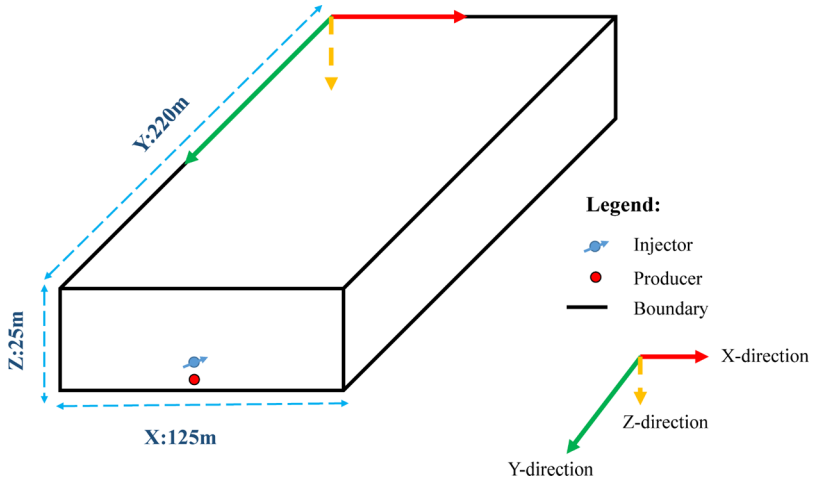


Figure 3. 3D SAGD base model ($\Delta X = \Delta Z = 1$ m; $\Delta Y = 10$ m, 50 m × 4, 10 m).

performance indicator, and feedforward neural network makes it possible to solve time series problems in an efficient way. It is an attempt to combine the physics-driven method and data-driven method. Ultimately, it can help reservoir engineers make better use of the SAGD technology for oil sands and heavy oil reservoirs.

This paper is structured as follows: first, the methodology for reservoir modeling and data generation is explained; then, the methodology for determination of performance indicator and data-driven proxy model construction is presented; next, validation and application of data-driven proxy model are elaborated; and finally, the related discussions are shown and key conclusions are summarized.

2. METHODOLOGY

2.1. Reservoir Modeling and Data Generation.

According to typical properties of MacKay River oil sands, a 3D SAGD base model (Figure 3) was constructed for flow simulation in an oil sand reservoir. As shown in Figure 3, the horizontal injector and producer are parallelly placed at the lower part of the model from the vertical view (Z-direction), and two parallel horizontal wells are located in the middle part of the model in the X-direction. The X-directional length of the 3D SAGD base model is set as 125 m considering the actual distance between two adjacent SAGD well pairs in the field. Also, 200 m horizontal wellbore along the Y-direction is modeled. Therefore, the 3D SAGD base model is 125 m × 220 m × 25 m, and the grid size is 1 m in both X and Z directions, and 10 m, 50 m × 4, and 10 m in the Y direction. The preheating period lasts 150 days and the production period lasts 20 years with the consideration of the realistic SAGD project. All simulation cases are based on the 3D SAGD base model.

Attributes belonging to initial conditions or reservoir characteristics are ungovernable factors which are associated with the reservoir or fluid. Also, attributes which belong to operating parameters are artificial factors relevant to SAGD well pairs. All attributes affect the performance of the SAGD process together. So that, input parameters used in generating simulation cases should cover aforementioned three categories, as present in Figure 4. Through literature review and field experience, a series of typical attributes which could be considered in the numerical simulation model are chosen as input parameters.[30−34] Input parameters attached to initial

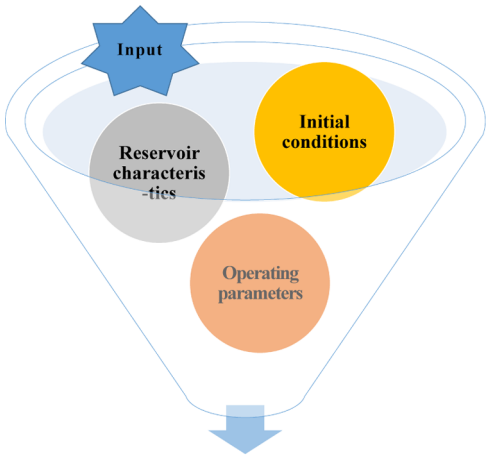


Figure 4. Input and output of the proxy model for Cum-oil production profile forecast during the SAGD process.

conditions include initial reservoir pressure, initial oil saturation, and thermal conductivity of rocks. Input parameters attached to reservoir characteristics include effective thickness, porosity, horizontal permeability, and ratio of vertical permeability to horizontal permeability. Operational pressure, steam rate of production well, and steam quality are considered as operating input parameters. According to actual conditions or experiences of the SAGD process in MacKay River oil sands, the ranges of these parameters are determined. Table 1 shows the ranges of all input parameters, which are divided into three categories, as also shown in Figure 4.

As shown in Table 1, initial reservoir pressure, P_i, values vary from 200 to 600 kPa and initial oil saturation, S_{oi}, values range from 0.65 to 0.85. Thermal conductivity of rocks, λ_r, values vary from 1.56×10^5 to 4.5×10^5 J/(m d °C) and effective thickness, h, values range from 15 to 25 m. Porosity, ϕ, values vary from 0.25 to 0.35 and horizontal permeability, k_h, values range from 2000 to 4000 mD. Ratio of vertical permeability to horizontal permeability, k_v/k_h, values vary from 0.3 to 0.8 and operational pressure, P_{inj}, values range from 1500 to 3000 kPa. Maximum steam rate of production well, Q_S, values vary from 5 to 15 m^3/d and steam quality, X_S, values range from 0.75 to 0.95. The latin hypercube sampling method, one of the

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

Table 1. Value Ranges of Inputs for the Data-Driven Proxy Model

category	parameter	unit	minimum	maximum
initial conditions	P_i	kPa	200	600
	S_{oi}	Fraction	0.65	0.85
	λ_r	J/(m·d·°C)	1.56×10^5	4.5×10^5
reservoir characteristics	h	m	15	25
	ϕ	Fraction	0.25	0.35
	k_h	mD	2000	4000
	k_v/k_h	Fraction	0.3	0.8
operating parameters	P_{inj}	kPa	1500	3000
	Q_S	m^3/d	5	15
	X_S	Fraction	0.75	0.95

experimental design methods, is used to produce 524 sets of data which yield uniform distribution within related value ranges listed in Table 1. Then, the corresponding simulation results are obtained through the commercial numerical simulator (CMG STARS, 2020).

2.2. Construction of the Knowledge-Based Performance Indicator. For building the data-driven proxy model which can accurately and efficiently predict the Cum-oil production profile of the SAGD process, an appropriate knowledge-based performance indicator must be found and clearly defined. Various growth mathematic models derived from the biological growth field have been widely used in the research of population growth problems, cell growth problems, and other domains of life, social, and economic sciences.[35,36] Growth is a common feature in various scenarios including reservoir production. Among the aforementioned application scenarios, the growth mathematic model offers an effective tool to account the growth under given confronting expansion and restraint forces.[37] The capacity of growth mathematic models makes it possible to describe the Cum-oil production profile from interaction of complicated recovery mechanisms involving simultaneous heat and mass transfer across a connectivity network. Importantly, strong analogies between the SAGD process and tumor growth process where growth mathematic models have been successfully applied could be established.[37] It is a significant motivation to take advantage of growth mathematic models for solving Cum-oil production forecasting problem during the SAGD process.

Some successful application cases of production forecasting with growth mathematic models are reported.[37−39] To be able to achieve better performance, different mathematic models, such as logistic model, Gompertz model, and von Bertalanffy model, have been introduced to fit the simulation results. After our attempts and comparisons, the von Bertalanffy model has better performance than other models when fitting the Cum-oil production profile during the SAGD process in our study. Therefore, the von Bertalanffy model is chosen as the performance indicator to fit Cum-oil production profiles. The general mathematical form of the von Bertalanffy model is described as follows[40−42]

$$z(t) = a(1 - be^{-kt})^3 \tag{1}$$

Then, the dz/dt can be described as follows

$$\frac{dz}{dt} = 3abke^{-kt}(1 - be^{-kt})^2 \tag{2}$$

Also, the d^2z/dt^2 is

$$\frac{d^2z}{dt^2} = 3abk^2e^{-kt}(1 - be^{-kt})(3be^{-kt} - 1) \tag{3}$$

The main mathematical characteristics of the von Bertalanffy model are as follows:

(1) From eq 1, it can be seen that $\lim_{t\to\infty} z(t) = a$, so that the boundedness of the von Bertalanffy model is proved. Term a always refers to maximum size or carrying capacity.
(2) According to eq 2, it can be seen that $dz/dt \geq 0$, so this model has the characteristic of monotonically increasing (all constants of the model are greater than 0).
(3) In addition, according to eq 3, it can be known that the curve of this model is S-shaped.

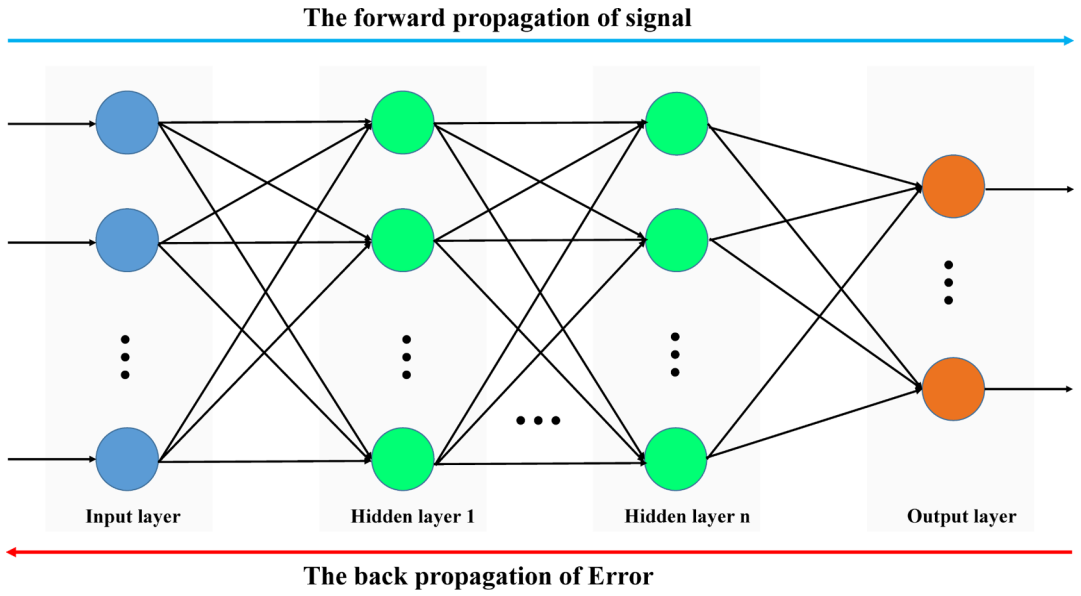


Figure 5. Schematic diagram of the feedforward neural network.

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

In the SAGD process, the characteristic of the Cum-oil production profile is similar with the curve of the von Bertalanffy model. First, not all the resources can be extracted under the specific technical conditions, so $\lim_{t\to\infty} N_p = N_R$. Second, it is obvious that the Cum-oil production curve is monotonically increasing. Third, the Cum-oil production curve is also S-shaped. All the features are consistent with characteristics of the von Bertalanffy model.

Based on the abovementioned analogies and eq 1, the Cum-oil production indicator can be described as follows[39]

$$N_p = N_R(1 - be^{-kt})^3 \tag{4}$$

Based on eq 4, the N_R and coefficients b and k could be used to fit the Cum-oil production profile of the SAGD process. By introducing such an indicator, it is possible to solve the time series problems with the feedforward neural network. Thus, it is feasible to acquire the related Cum-oil production at any desired time step.

Based on the Cum-oil production profiles which are extracted from the simulation results mentioned in Section 2.1, the N_R and coefficients b and k of each data sets are obtained according to eq 4. In the fitting process, the Levenberg−Marquart method is used to acquire better fitting performance. Results show that the Cum-oil performance indicator used in this paper provides great fitting results for the SAGD process. For all cases, the R-square which represent the fitting accuracy is almost equal to 1.00. Also, among the cases, the N_R is found to be in the range 4.84 to 15.56 × 10^4 m^3, coefficient b is found to be in the range 0.54 to 0.82, and coefficient k is found to be in the range 0.11 to 0.49.

2.3. Data-Driven Proxy Model. *2.3.1. Related Basic Theories of the Feedforward Neural Network.* In Section 2.3, the feedforward neural network is employed to construct the data-driven proxy model. The feedforward neural network is a highly nonlinear mapping processing system with self-organization, self-learning, and self-adaptation capabilities, inspired by the biological nervous system. Generally, it is made up of an input layer, one or more hidden layers, and an output layer, and each layer has a different number of neurons (Figure 5).[43,44] The neuron number of input layer and output layer is related to the number of input parameters and output parameters, respectively, while the hidden layers always have several highly interconnected neurons. Generally, the utilization of the neural network includes two parts: training and forecasting. As depicted in Figure 5, the training process of the feedforward neural network could be divided into two sections. The first section is that the signal propagates forward from input layer to output layer. Also, the second section is that the error propagates backward from output layer to start point, in order to correct the weight matrix. Finally, the well-trained neural network could be obtained through several iterations.

During the forward propagation process of the signal, the algorithm of single-layered perception is shown in Figure 6. After inputting the data set, connection weight is used to adjust the weight ratio of each input. The next procedure is the summation process. When the summation result including bias is obtained, the output can be acquired through activation function. Thus, the mathematical formula is given by[45,46]

$$y_j = f\left(\sum_{i=1}^{n} \omega_{ij}x_i + \theta_j\right) \tag{5}$$

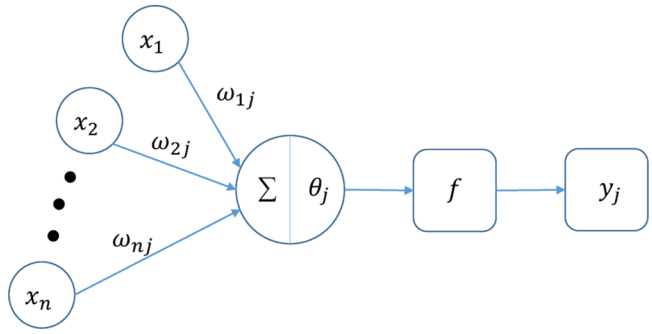


Figure 6. Algorithm of single-layered perception.

During the back propagation process of error, the updating formula of weight and bias is

$$\Delta\omega_{ij} = l\mathrm{Err}_j O_i \tag{6}$$

$$\Delta\theta_j = l\mathrm{Err}_j \tag{7}$$

For convenience, we use the NN (neuron network) to denote the feedforward neural network used here in the following part of this paper.

2.3.2. NN-Based Data-Driven Proxy Model. To build the NN-based data-driven proxy model, the first step is to construct the overall data set which include all the input and output parameters. Ten input parameters are shown in Table 1 and three output parameters are N_R, b, and k which are derived from fitting results of simulation outputs. Due to the inconsistent magnitude of ten input parameters, it is normalized using the following formula

$$\bar{x} = 2\frac{x - x_{\min}}{x_{\max} - x_{\min}} - 1, \qquad \bar{x} \in [-1, 1] \tag{8}$$

The overall data set which consists of 524 sets of input and output data is randomly categorized into three using the 80/10/10 percent split: training data set, validation data set, and testing data set. The training data set is used for training the NN, while the validation data set is used for hyperparameter adjustment to find the suitable structure of the NN. Last but not the least, the testing data set plays a significant role in the blind-testing process (not involved in the training process at all). It is used for evaluating the final forecasting performance of the data-driven proxy model.

In this paper, the NN is implemented in Python 3.7 through programming with the utilization of PyTorch. To find the suitable structure of the NN, the main hyperparameters that need to be adjusted are the learning rate, activation function, loss function, updating optimization algorithm, number of neurons in hidden layers, and number of layers.

After the attempt of tuning the learning rate from 0.0001 to 0.8, 0.004 is selected as the initial value of the learning rate, while a learning rate optimizer named "ReduceLROnPlateau" is used.[47] Such an optimizer allows dynamic reducing of the learning rate once the loss function stops decreasing. Specifically, the factor value is set as 0.1 and the patience value is set as 10. ReLU activation function and L1 loss function have been picked as a final activation function and loss function, respectively. After that, different updating optimization algorithms, such as stochastic gradient descent (SGD) method and Adam method, have been applied in the training process of the NN.[48] Results reveal that the SGD method with appropriate settings outperforms other methods,

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

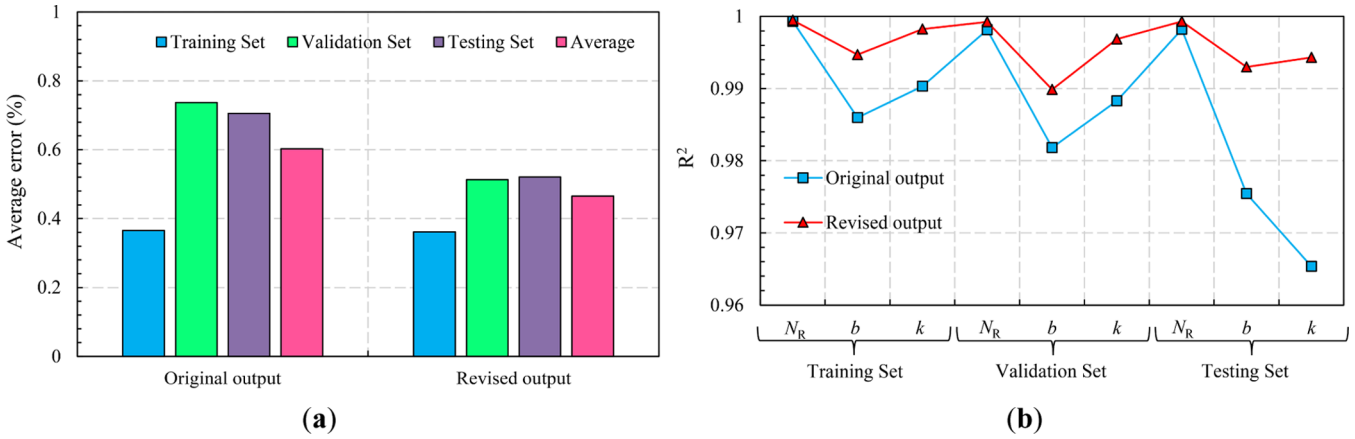


Figure 7. Comparison of different choices of output parameters: (a) average error and (b) coefficient of determination (R^2) for three outputs.

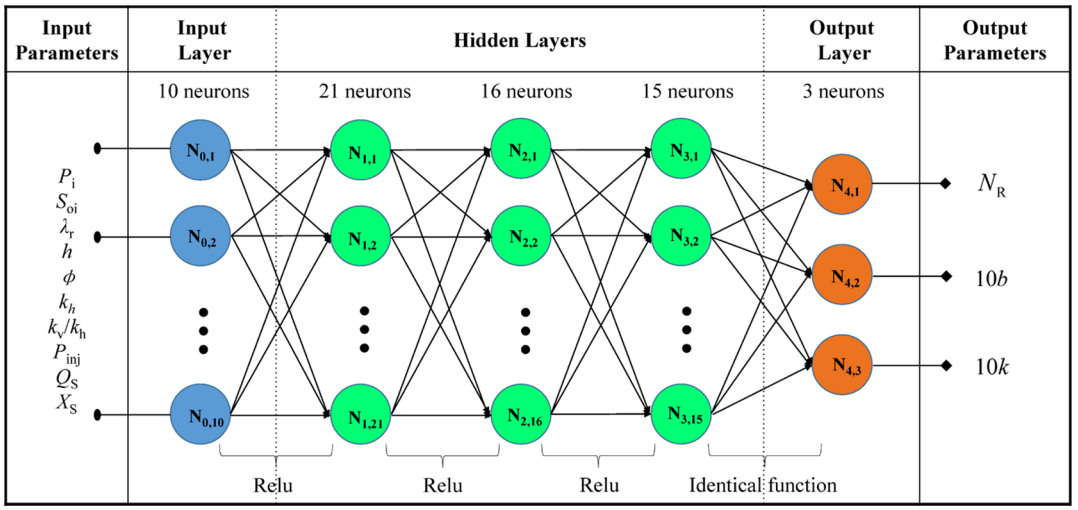


Figure 8. Final topology of the NN.

so the SGD method with a 0.75 momentum value is the final choice. In addition, the L2 regularization method is used, in order to avoid overfitting phenomena.

Based on the aforementioned settings, several varieties of configurations involving one layer or multiple layers (5 to 100 neurons in each layer) are explored. Eventually, the best-performing NN structure is chosen. It can be seen from results that the forecast precision of b and k are relatively low, compared with the N_R. To improve the performance, the revised output parameters have been adopted in the latest NN. The latest NN uses N_R, $10b$, and $10k$ instead of N_R, b, and k considering the magnitude difference between original output parameters. The performance improvement can be seen from Figure 7. Figure 7a shows that the average error of all sets are reduced and from Figure 7b, we can observe that more accurate outputs are obtained.

Figure 8 shows the final topology of the NN and there are three hidden layers with 21, 16, and 15 neurons, respectively. Thus, the construction of the NN-based data-driven proxy model is completed.

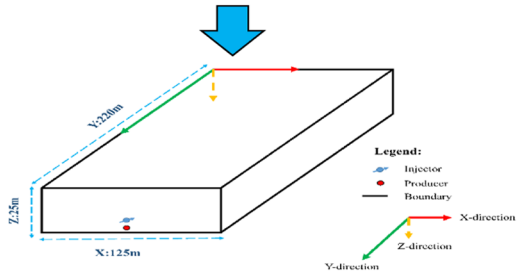
2.4. Overview of Workflow for Building the Data-Driven Proxy Model. According to the abovementioned statement, the workflow of building the data-driven proxy model for Cum-oil production profile forecast of the SAGD process could be summarized as follows (Figure 9):

(1) The first step is to choose the appropriate variables as input parameters and construct the database of the input parameters within given ranges.
(2) After the data generation procedure, a variety of reservoir models with input parameters mentioned in Step (1) are established. Then, each simulation run is completed using the commercial simulator. Therefore, the Cum-oil profiles are extracted as performance characters.
(3) The third step is to construct the knowledge-based performance indicator according to simulation results of various models built in Step 2. In this step, the von Bertalanffy model is chosen to represent the Cum-oil production profiles. The N_R and coefficients b and k of each data sets can be obtained through the Levenberg-Marquart method.
(4) According to the abovementioned research results, the initial framework of the feedforward neural network is designed.
(5). The aim of the fifth step is to find the suitable topology of the NN. In this step, main hyperparameters are constantly adjusted until the error is in the accepted range.
(6) After the training process, the adaptability and accuracy of the trained NN are validated by the blind-testing data set. Thus, it can be used as a data-driven proxy model to

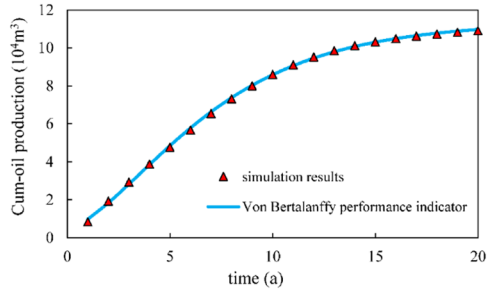
https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

P_i, kPa	S_{oi}, fraction	·····	Q_S, m^3/d	X_S, fraction
200	0.85		10	0.91
500	0.69		11	0.85
600	0.75	·····	15	0.93
⋮	⋮		⋮	⋮
400	0.79		5	0.75

1. Construct the database of the input parameters within the minimum and maximum values used in the study.

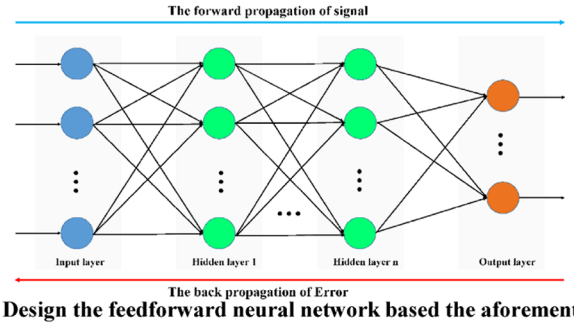


2. Complete the reservoir modeling process based on the data mentioned in the Step 1.

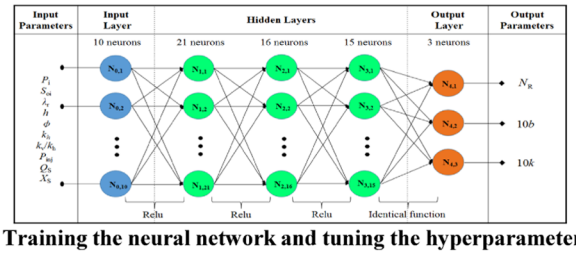


3. Construct the knowledge based performance indicator based on the simulation results of various models built in Step2.

4. Design the feedforward neural network based the aforementioned input and output parameters.



5. Training the neural network and tuning the hyperparameters to find the suitable structure of it.



6. Validate the trained network by blind testing dataset and then use it as a data-driven proxy model for performance forecasting.

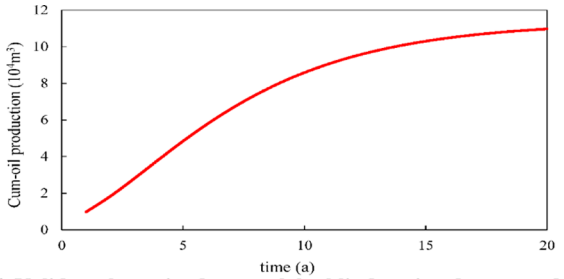


Figure 9. Proposed workflow in this paper.

forecast the Cum-oil production profile during the SAGD process.

2.5. Application of the Data-Driven Proxy Model. After the accuracy of the model has been verified, it can be used to do some application works. The application of the data-driven proxy model will be illustrated in Section 3. First, efficient parametric studies are shown. Second, uncertainty analysis of the SAGD process is conducted with some assumptions shown in Table 2. It is assumed that all parameters yield normal distribution. Third, the ability of the data-driven proxy model to predict the average daily oil production is shown.

3. RESULTS

3.1. Performance of the Data-Driven Proxy Model. In this part, the performance of the data-driven proxy model is evaluated. The data-driven proxy model is mainly used to predict the N_R, b, and k. Once the N_R, b, and k of a given case are determined, the Cum-oil production profile can be obtained using the performance indicator. It is worth noting that the outputs of revised version of NN are N_R, $10b$, and $10k$. Therefore, the simple conversion is needed before the final calculation. The relative error (RE) is defined as follows

Table 2. Different Variable Parameters and Their Value Ranges

parameter	unit	expectation	standard deviation
P_i	kPa	400	20
S_{oi}	Fraction	0.75	0.02
λ_r	J/(m·d·°C)	3×10^5	3×10^4
h	m	20	1
ϕ	Fraction	0.3	0.01
k_h	mD	3000	200
k_v/k_h	fraction	0.55	0.06
P_{inj}	kPa	2250	150
Q_S	m^3/d	10	1
X_S	fraction	0.85	0.02

$$\mathrm{RE} = \left| \frac{y_p - y_r}{y_r} \right| \times 100\% \tag{9}$$

Also, the average absolute percentage error (AAPE) is defined as follows

$$\mathrm{AAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{y_p - y_r}{y_r} \right| \times 100\% \tag{10}$$

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

The AAPE of N_R, b, and k between different models is shown in Figure 10 and so did the 20-year Cum-oil production. Figure 10 shows that the error of each set (training set, validation set, and testing set) is relatively low. The AAPE of all the parameters in each set is less than 2%. Therefore, the accuracy of the model is verified.

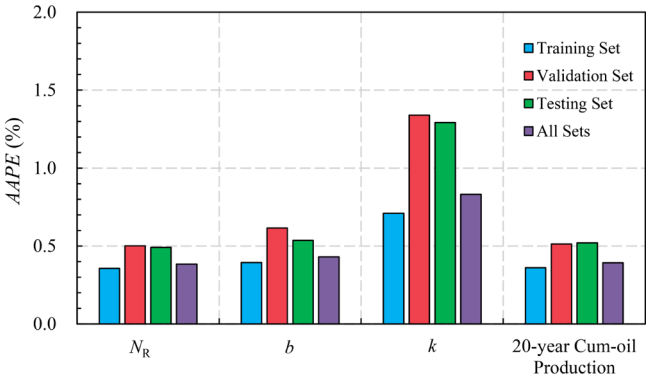


Figure 10. AAPE of N_R, b, k, and 20-year Cum-oil production in each data set.

Then, the error frequencies about RE of different parameters in each set are drawn in Figure 11. It can be seen from Figure 11 that the error in most cases is less than 1% and the error in almost entire cases is less than 3%. It can also be observed from Figures 10 and 11 that the forecasting precision of k is lower than N_R and b. The maximum RE of k is higher than 5%. However, it should be noted that the effect of k on the Cum-oil production profile is comparatively small. Thus, the error of k is acceptable.

Two cases with comparatively better forecasting performance are shown in Figure 12. We can see from Figure 12 that the Cum-oil production curves obtained from two different models fit well. Two cases with comparatively worse forecasting performance are shown in Figure 13. Although the certain deviation between two curves is observed from Figure 13, the general trend of the curve predicted using the data-driven proxy model is consistent with the other. In addition, we can also conclude that the forecasting precision of N_R plays a significant role in the model. In Figure 13, Cum-oil production profiles appear to be marginally overestimated mainly because the value of N_R could not be forecasted as desired so that it also affects the error of 20-year Cum-oil production negatively.

Each reservoir simulation run completed in this paper spends about 50 min due to the complexity of the recovery mechanism on an Intel Core i7-3770 3.40 GHz CPU, whereas the data-driven proxy model just takes around a few minutes for the overall data sets (524 cases).

3.2. Efficient Parametric Studies. The data-driven proxy model is a quite powerful tool to complete the sensitivity analysis work of different input parameters, as a result of its ability of saving time. In this section, S_{oi} and k_h are taken as an example to do the sensitivity analysis work. For instance, we change the value of S_{oi} and leave the remaining parameters unchanged so as to study the effect of S_{oi} on the Cum-oil production profile of the SAGD process.

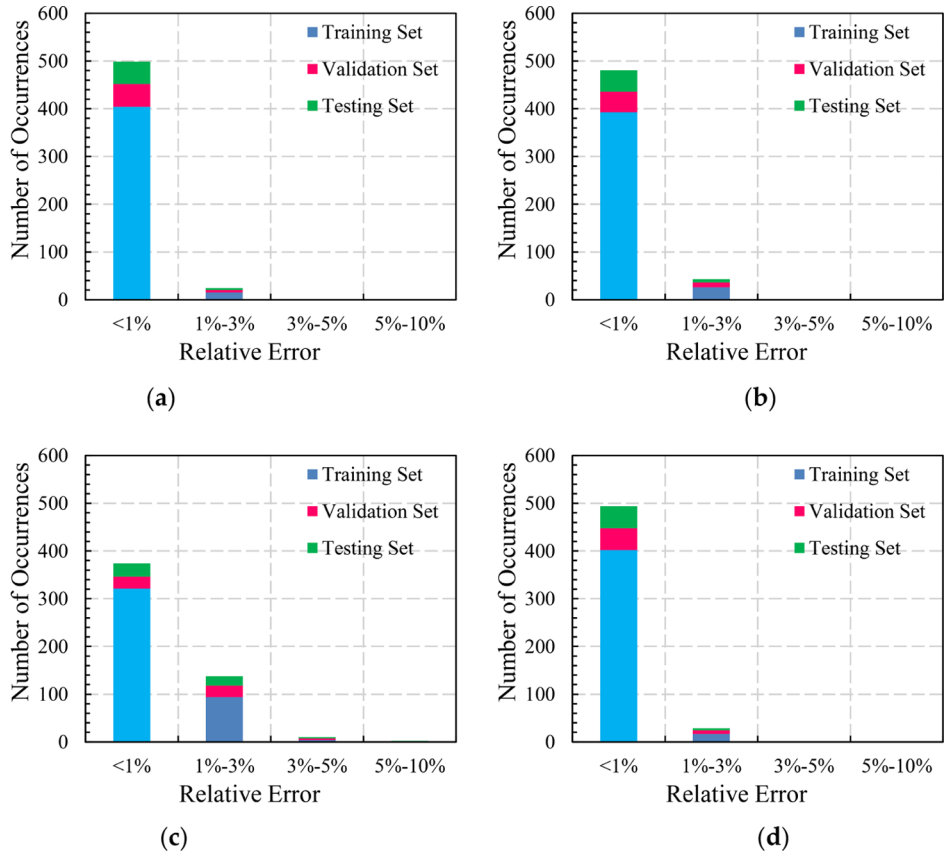


Figure 11. RE frequencies of output parameters and 20-year Cum-oil production for different sets: (a) N_R, (b) b, (c) k, and (d) 20-year Cum-oil production.

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

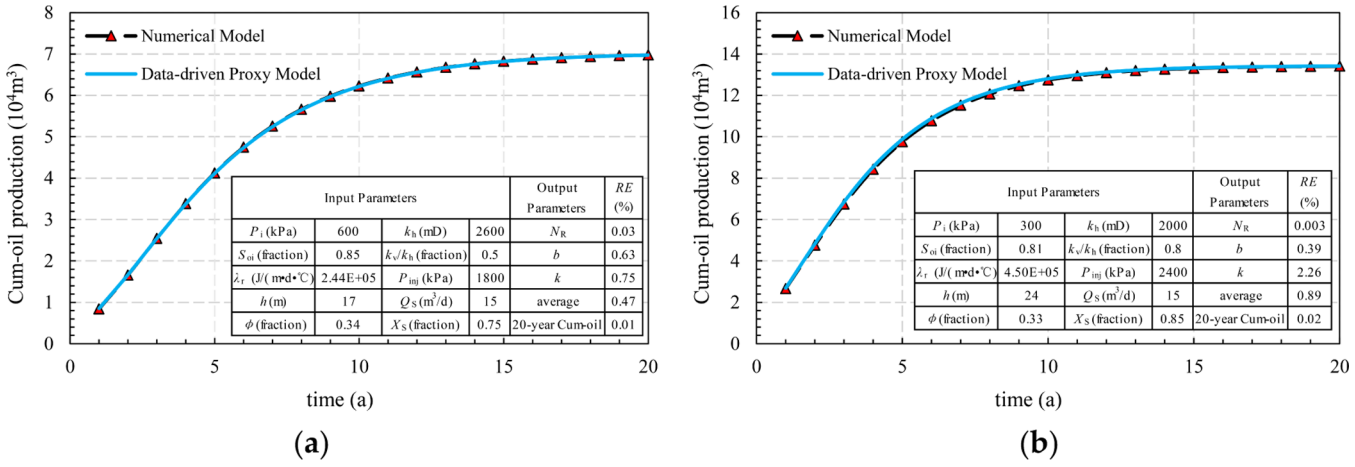


Figure 12. Two cases with comparatively better forecasting performance: (a) average error of three outputs = 0.47%; (b) average error of three outputs = 0.89%.

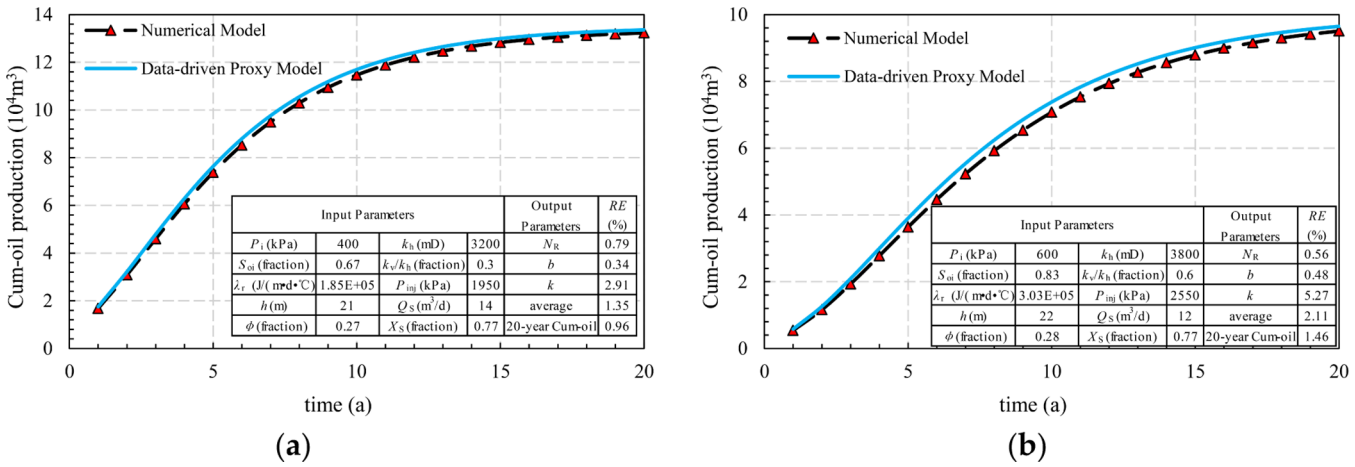


Figure 13. Two cases with comparatively worse forecasting performance: (a) average error of three outputs = 1.35%; (b) average error of three outputs = 2.11%.

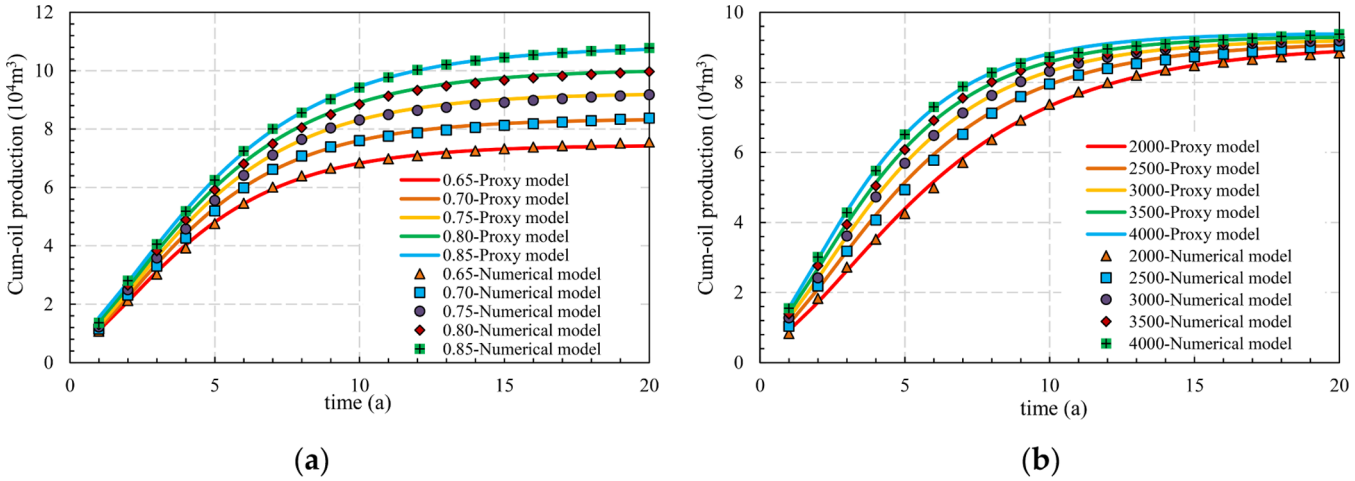


Figure 14. Comparison of sensitivity analysis results between the data-driven proxy model and numerical model: (a) S_{oi}; (b) k_h.

The sensitivity analysis results obtained from the data-driven proxy model and numerical simulation model are shown in Figure 14. The comparison between the two shows a good consistency. Figure 14a shows the sensitivity of Cum-oil production to S_{oi}, and it can be seen that the greater the S_{oi} is, the more the oil can be drained from porous media, and the higher the Cum-oil production is. Figure 14b shows the sensitivity of Cum-oil production to k_h, and it can be seen that the greater the k_h is, the higher the expansion velocity of the steam chamber is, and the higher the Cum-oil production is, but the 20-year Cum-oil production of each case is relatively close.

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

3.3. Uncertainty Analysis of the SAGD Process. The Monte Carlo simulation method is adopted to conduct the uncertainty analysis of the SAGD process through the data-driven proxy model. The expectation curves of the Cum-oil recovery factor for two different production time periods are shown in Figure 15, which are obtained by the Monte Carlo

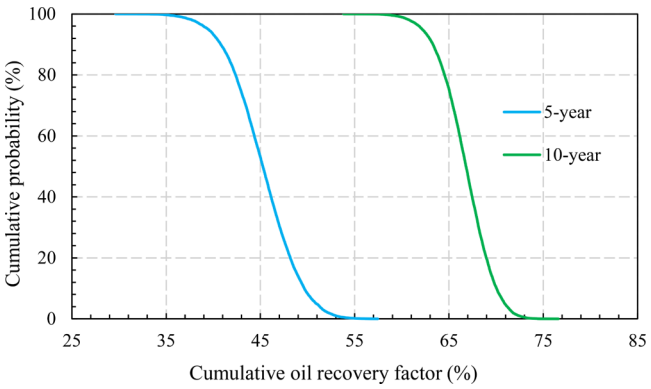


Figure 15. Expectation curves of the Cum-oil recovery factor for two different production time periods.

simulation of 10,000 samples. It takes around just a few seconds. Such an application allows us to quantify the uncertainties of different input parameters to observe their effect on the performance of the SAGD process.

Figure 16 shows the P10, P50, and P90 estimations of the Cum-oil recovery factor for 5 and 10 years. This workflow can

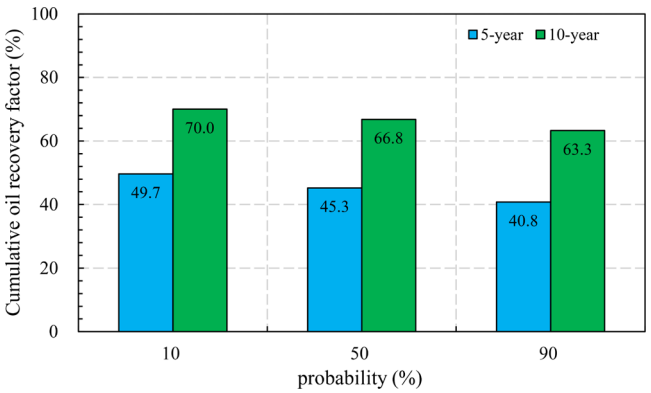


Figure 16. P10, P50, and P90 estimations of the Cum-oil recovery factor for two different production time periods.

be used to compare the effect of different operating parameters on the performance of SAGD with known uncertainties. Furthermore, the Cum-oil recovery factor could be integrated into the economic evaluation process to help reservoir engineers making decisions.

3.4. Prediction of Average Daily Oil Production. This model also could be used to forecast the average daily oil production. For a given case, the N_R, b, and k could be predicted using the data-driven proxy model. Then, the Cum-oil profile could be plotted using the performance indicator based on the aforementioned parameters. Thus, the average daily oil production can be calculated by such a simple formula 11. This function is illustrated in Figure 17 with a given case.

$$\overline{q_o} = (N_p^{\ s} - N_p^{\ s-1})/t_{prod} \quad (11)$$

4. DISCUSSION

The performance indicator is a convenient and effective tool to characterize the production profile. Meanwhile, the fitting results obtained using a performance indicator could be integrated into the NN, which is a powerful approach in the forecasting field. The integration of two parts make it possible to solve time series problems in an efficient way. The data-driven proxy model takes full advantage of the capability of the feedforward neural network, instead of using more complicated neural networks, and the desired effect is achieved. Therefore, it can help reservoir engineers make better use of the SAGD technology for oil sands or heavy oil reservoirs.

It is noteworthy that the error derived from the reservoir numerical model or performance indicator will be carried into the data-driven proxy model because the reservoir simulation models obey some assumptions, compared with the actual situation. Also, the degree of agreement between the simulation results and the performance indicator cannot reach 100%. However, this fact does not prevent the data-driven proxy model from becoming a powerful tool to do forecasting works.

In our study, ten attributes are selected as variable in the numerical simulation model, and cases containing different combinations of those attributes are generated. Considering that such a feature dimension is not high, all of the ten attributes are selected as input parameters for the data-driven proxy model, in order to capture different configurations as much as possible. However, when it comes to a more complex situation including a large number of input features, it is rather remarkable that sensitivity analysis is a useful approach which can assist engineers to complete input parameter determination tasks. Such a way could help engineers to reduce the computational cost while maintaining the performance of the data-driven model at the acceptable level.

In addition, when it is aimed to study the case which had the value range outside our study, the data set used in the training process ought to be expanded and the data-driven proxy model ought to be retrained, in order to include new conditions. The workflow of building the data-driven proxy model presented in this study would offer guidance to the corresponding research.

5. SUMMARY AND CONCLUSIONS

(1) Based on the reservoir numerical simulation approach, the von Bertalanffy performance indicator, and ANN, the data-driven proxy model for Cum-oil production profile forecasting of the SAGD process is established. The data-driven proxy model fully considers initial conditions, reservoir characteristics, and operating parameters.

(2) During the training process of the NN, several attempts of hyperparameter adjustment have been done to find the suitable structure of the network. For this study, the combination of "ReduceLROnPlateau" optimizer, ReLU activation function, L1 loss function, and SGD algorithm is applied. For further improving forecasting performance of the neural network, some strategies or tricks, such as L2 regularization method and output revision, are used. The ultimate structure of the neural network consists of three hidden layers with 21, 16, and 15 neurons, respectively.

(3) The reliability of the data-driven proxy model is verified by testing the data set. Average absolute percentage error

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

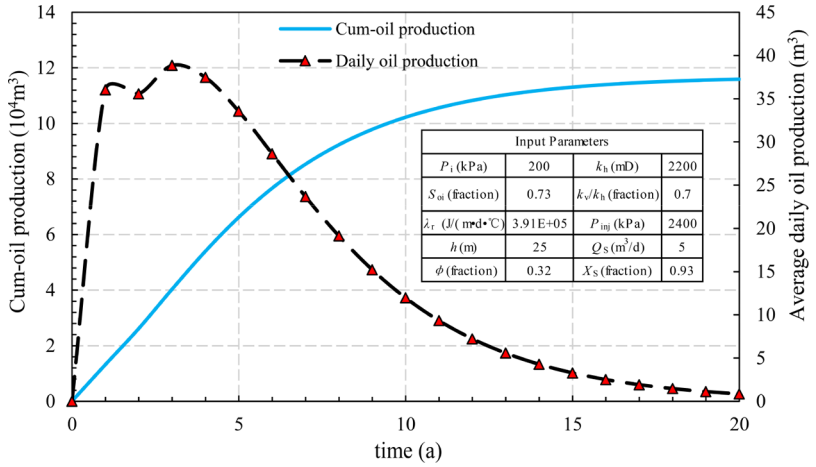


Figure 17. Average daily oil production profile calculated from the Cum-oil production profile.

of related parameters of the performance indicator in the testing data set is 0.77%, and the error of Cum-oil production after 20 years is 0.52%.

(4) The data-driven proxy model could be employed to study large amounts of data efficiently, as shown in the application of parametric studies and uncertainty analysis, and it could also be used to forecast average daily oil production of a given case. Such functions could help engineers to make the decision. Furthermore, the developed workflow also can be extended to more complex situations of the SAGD process.

AUTHOR INFORMATION

Corresponding Authors

Yang Yu − *PetroChina Research Institute of Petroleum Exploration and Development, Beijing 100083, China*; orcid.org/0000-0003-1196-000X; Email: yystzp@hotmail.com

Yu Bao − *PetroChina Research Institute of Petroleum Exploration and Development, Beijing 100083, China*; Email: baoyu03@petrochina.com.cn

Authors

Shangqi Liu − *PetroChina Research Institute of Petroleum Exploration and Development, Beijing 100083, China*

Yang Liu − *PetroChina Research Institute of Petroleum Exploration and Development, Beijing 100083, China*

Lixia Zhang − *PetroChina Research Institute of Petroleum Exploration and Development, Beijing 100083, China*

Yintao Dong − *CNOOC Research Institute, Beijing 100028, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.1c00617

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding

This research was funded by National Science and Technology Major Project, grant number 2016ZX05031.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors would like to acknowledge Computer Modeling Group (CMG) Company for providing the academic licenses for STARS and also want to acknowledge the open source platform named PyTorch. Finally, we would also like to express our gratitude to reviewers and editors.

ABBREVIATIONS

P_i, initial reservoir pressure, kPa
S_{oi}, initial oil saturation, fraction
λ_r, thermal conductivity of rocks, J/(m·d·°C)
h, effective thickness, m
ϕ, porosity, fraction
k_h, horizontal permeability, mD
k_v/k_h, ratio of vertical permeability to horizontal permeability, fraction
P_{inj}, operational pressure, kPa
Q_S, maximum steam rate of production well, m^3/d
X_S, steam quality, fraction
z, von Bertalanffy function, dimensionless
a, constant of the von Bertalanffy model, dimensionless
b, constant of the von Bertalanffy model, dimensionless
k, constant of the von Bertalanffy model, a^{-1}
t, time, a
N_p, Cum-oil production, 10^4 m^3
N_R, ultimate oil production, 10^4 m^3
y_j, output at neuron j of the current layer
x_i, input from the previous layer
f, activation function
n, number of neurons in the previous layer
ω_{ij}, connected weight between neuron i of the previous layer and neuron j of the current layer
θ_j, bias at neuron j of the current layer
$\Delta\omega_{ij}$, variation of weight
$\Delta\theta_j$, variation of bias
l, learning rate
Err$_j$, error of neuron j
O_i, output of neuron i
$\overline{x}$, value of input after normalization
x_{min}, minimum value of the related input parameter
x_{max}, maximum value of the related input parameter
RE, relative error
y_p, predicted value obtained using the data-driven proxy model

y_v, value obtained from simulation results
AAPE, average absolute percentage error
n, number of samples
$\overline{q_o}$, average daily oil production, m^3
N_p^s, Cum-oil production in the (s)th year, 10^4 m^3
N_p^{s-1}, Cum-oil production in the $(s-1)$th year, 10^4 m^3
t_{prod}, production days per year, d

REFERENCES

(1) Guo, K.; Li, H.; Yu, Z. In-situ heavy and extra-heavy oil recovery: A review. *Fuel* **2016**, *185*, 886−902.
(2) Saboorian-Jooybari, H.; Dejam, M.; Chen, Z. Heavy oil polymer flooding from laboratory core floods to pilot tests and field applications: Half-century studies. *J. Pet. Sci. Eng.* **2016**, *142*, 85−100.
(3) Ali, S. M. Life after SAGD − 20 years later. *SPE Western Regional Meeting*; Society of Petroleum Engineers, 2016.
(4) Liu, P.; Zhou, Y.; Liu, P.; Shi, L.; Li, X.; Li, L. Numerical study of herringbone injector-horizontal producer steam assisted gravity drainage (HI-SAGD) for extra-heavy oil recovery. *J. Pet. Sci. Eng.* **2019**, *181*, 106227.
(5) Al-Bahlani, A.-M.; Babadagli, T. SAGD laboratory experimental and numerical simulation studies: A review of current status and future issues. *J. Pet. Sci. Eng.* **2009**, *68*, 135−150.
(6) Jimenez, J. The field performance of SAGD projects in Canada. *International Petroleum Technology Conference*; Society Committees of IPTC, 2008.
(7) Irani, M.; Ghannadi, S. Understanding the heat-transfer mechanism in the Steam-assisted Gravity Drainage (SAGD) process and comparing the conduction and convection flux in bitumen reservoirs. *SPE J.* **2013**, *18*, 134−145.
(8) Jia, X.; Qu, T.; Chen, H.; Chen, Z. Transient convective heat transfer in a steam-assisted gravity drainage (SAGD) process. *Fuel* **2019**, *247*, 315−323.
(9) Sharma, J.; Gates, I. D. Convection at the Edge of a Steam-Assisted-Gravity-Drainage Steam Chamber. *SPE J.* **2011**, *16*, 503−512.
(10) Mohammadi, K.; Ameli, F. Toward mechanistic understanding of Fast SAGD process in naturally fractured heavy oil reservoirs: Application of response surface methodology and genetic algorithm. *Fuel* **2019**, *253*, 840−856.
(11) Luo, E.; Fan, Z.; Hu, Y.; Zhao, L.; Bo, B.; Yu, W.; Liang, H.; Liu, M.; Liu, Y.; He, C.; Wang, J. An efficient optimization framework of cyclic steam stimulation with experimental design in extra heavy oil reservoirs. *Energy* **2020**, *192*, 116601.
(12) Ma, Z.; Leung, J. Y. A knowledge-based heterogeneity characterization framework for 3D steam-assisted gravity drainage reservoirs. *Knowl. Base Syst.* **2020**, *192*, 105327.
(13) Irani, M.; Ghannadi, S. Modeling the conformance improvement using flow control devices in infill wells adjacent to SAGD well pairs: No flashing. *SPE J.* **2020**, *25*, 800−819.
(14) Gupta, S.; Fuehrer, F.; Jeyachandra, B. C. Production forecasting in unconventional resources using data mining and time series analysis. *SPE/CSUR Unconventional Resources Conference—Canada*; Society of Petroleum Engineers, 2014.
(15) Kulga, B.; Artun, E.; Ertekin, T. Development of a data-driven forecasting tool for hydraulically fractured, horizontal wells in tight-gas sands. *Comput. Geosci.* **2017**, *103*, 99−110.
(16) Amirian, E.; Dejam, M.; Chen, Z. Performance forecasting for polymer flooding in heavy oil reservoirs. *J. Pet. Sci. Eng.* **2018**, *216*, 83−100.
(17) Sagheer, A.; Kotb, M. Time series forecasting of petroleum production using deep LSTM recurrent networks. *Neurocomputing* **2019**, *323*, 203−213.
(18) Negash, B. M.; Yaw, A. D. Artificial neural network based production forecasting for a hydrocarbon reservoir under water injection. *Pet. Explor. Dev.* **2020**, *47*, 383−392.
(19) Xue, L.; Liu, Y.; Xiong, Y.; Liu, Y.; Cui, X.; Lei, G. A data-driven shale gas production forecasting method based on the multi-objective random forest regression. *J. Pet. Sci. Eng.* **2021**, *196*, 107801.
(20) Zhong, Z.; Sun, A. Y.; Wang, Y.; Ren, B. Predicting field production rates for waterflooding using a machine learning-based proxy model. *J. Pet. Sci. Eng.* **2020**, *194*, 107574.
(21) Deng, L.; Pan, Y. Data-driven proxy model for waterflood performance prediction and optimization using Echo State Network with Teacher Forcing in mature fields. *J. Pet. Sci. Eng.* **2021**, *197*, 107981.
(22) Li, Y.; Zhu, Z.; Kong, D.; Han, H.; Zhao, Y. EA-LSTM: Evolutionary attention-based LSTM for time series prediction. *Knowl. Base Syst.* **2019**, *181*, 104785.
(23) Bu, S.-J.; Cho, S.-B. Time Series Forecasting with Multi-Headed Attention-Based Deep Learning for Residential Energy Consumption. *Energies* **2020**, *13*, 4722.
(24) Büyükşahin, Ü.Ç.; Ertekin, Ş. Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing* **2019**, *361*, 151−163.
(25) Jakob, S.; Steven, L. Hierarchical temporal memory and recurrent neural networks for time series prediction: An empirical validation and reduction to multilayer perceptrons. *Neurocomputing* **2020**, *396*, 291−301.
(26) Panigrahi, S.; Behera, H. S. A hybrid ETS−ANN model for time series forecasting. *Eng. Appl. Artif. Intell.* **2017**, *66*, 49−59.
(27) Zhang, Y.; Wang, X.; Tang, H. An improved Elman neural network with piecewise weighted gradient for time series prediction. *Neurocomputing* **2019**, *359*, 199−208.
(28) Wei, X.; Zhang, L.; Yang, H.-Q.; Zhang, L.; Yao, Y.-P. Machine learning for pore-water pressure time-series prediction: Application of recurrent neural networks. *Front. Geosci.* **2021**, *12*, 453−467.
(29) 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.
(30) Dong, X.; Liu, H.; Hou, J.; Cheng, Z.; Zhang, T. An empirical correlation to predict the SAGD recovery performance. *SPE/IATMI Asia Pacific Oil & Gas Conference and Exhibition*; Society of Petroleum Engineers, 2015.
(31) Liu, H.; Wang, Y.; Xiong, H.; Wu, K. Semianalytical analysis of chamber growth and energy efficiency of Solvent-Assisted Steam-Gravity Drainage considering the effect of reservoir heterogeneity along the horizontal well. *Energy Fuels* **2020**, *34*, 5777−5787.
(32) Barillas, J. L. M.; Dutra, T. V.; Mata, W. Reservoir and operational parameters influence in SAGD process. *J. Pet. Sci. Eng.* **2006**, *54*, 34−42.
(33) Hashemi-Kiasari, H.; Hemmati-Sarapardeh, A.; Mighani, S.; Mohammadi, A. H.; Sedaee-Sola, B. Effect of operational parameters on SAGD performance in a dip heterogeneous fractured reservoir. *Fuel* **2014**, *122*, 82−93.
(34) Heidari, M.; Pooladi-Darvish, M.; Azaiez, J.; Maini, B. Effect of drainage height and permeability on SAGD performance. *J. Pet. Sci. Eng.* **2009**, *68*, 99−106.
(35) Araujo, R. P.; Mcelwain, D. L. S. A history of the study of solid tumour growth: the contribution of mathematical modelling. *Bull. Math. Biol.* **2004**, *66*, 1039−1091.
(36) Jha, A.; Saha, D. Forecasting and analysing the characteristics of 3G and 4G mobile broadband diffusion in India: A comparative evaluation of Bass, Norton-Bass, Gompertz, and logistic growth models. *Technol. Forecast. Soc.* **2020**, *152*, 119885.
(37) Klie, H. Physics-based and data-driven surrogates for production forecasting. *SPE Reservoir Simulation Symposium*; Society of Petroleum Engineers, 2015.
(38) Clark, A. J.; Lake, L. W.; Patzek, T. W. Production forecasting with logistic growth models. *SPE Annual Technical Conference and Exhibition*; Society of Petroleum Engineers, 2011.
(39) Pang, M.; Tang, H. Von Bertalanffy mathematical model for predicting oil field cumulative production. *China Sciencepap.* **2017**, *12*, 2487−2491.

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509

(40) Mata-Estrada, A.; González-Cerón, F.; Pro-Martínez, A.; Torres-Hernández, G.; Bautista-Ortega, J.; Becerril-Pérez, C. M.; Vargas-Galicia, A. J.; Sosa-Montes, E. Comparison of four nonlinear growth models in Creole chickens of Mexico. *Poult. Sci.* **2020**, *99*, 1995−2000.
(41) Román-Román, P.; Romero, D.; Torres, F. A diffusion process to model generalized von Bertalanffy growth patterns: Fitting to real data. *J. Theor. Biol.* **2010**, *263*, 59−69.
(42) Helidoniotis, F.; Haddon, M.; Tuck, G.; Tarbath, D. The relative suitability of the von Bertalanffy, Gompertz and inverse logistic models for describing growth in blacklip abalone populations (Haliotis rubra) in Tasmania, Australia. *Fish. Res.* **2011**, *112*, 13−21.
(43) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436−444.
(44) Rajabzadeh, A. R.; Ruzich, N.; Zendehboudi, S.; Rahbari, M. Biomass leachate treatment and nutrient recovery using reverse osmosis: experimental study and hybrid artificial neural network modeling. *Energy Fuels* **2012**, *26*, 7155−7163.
(45) Wang, S.; Chen, S. Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling. *J. Pet. Sci. Eng.* **2019**, *174*, 682−695.
(46) Xie, Y.; Zhu, C.; Zhou, W.; Li, Z.; Liu, X.; Tu, M. Evaluation of machine learning methods for formation lithology identification: A comparison of tuning processes and model performances. *J. Pet. Sci. Eng.* **2018**, *160*, 182−193.
(47) Zhang, Y.; Xu, S.; Zhong, S.; Bai, X.-S.; Wang, H.; Yao, M. Large eddy simulation of spray combustion using flamelet generated manifolds combined with artificial neural networks. *Energy AI* **2020**, *2*, 100021.
(48) Ruder, S. An overview of gradient descent optimization algorithms. 2016, arXiv:abs/1609.04747. Available online: https://arxiv.org/abs/1609.04747.

https://doi.org/10.1021/acsomega.1c00617
ACS Omega 2021, 6, 11497−11509