

Research Article

Probabilistic Evaluation of Hydraulic Fracture Performance Using Ensemble Machine Learning

Xiaoping Xu,¹ Xianlin Ma ,² and Jie Zhan ²

¹Shengli Oilfield, SINOPEC, Dongying 257001, China

²School of Petroleum Engineering, Xi'an Shiyou University, Xi'an 710065, China

Correspondence should be addressed to Xianlin Ma; xianlinm@126.com and Jie Zhan; zhanjie@xsyu.edu.cn

Received 7 January 2022; Accepted 30 March 2022; Published 16 April 2022

Academic Editor: Yating Wang

Copyright © 2022 Xiaoping Xu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Evaluation of hydraulic fracture (HF) performances is critical to develop unconventional resources such as tight oil and gas. We present a probabilistic evaluation approach that integrates ensemble machine learning with Monte Carlo simulation. In the method, we employ the ensemble learning to develop a predictive model between well productivity and its influential factors including both geological properties and HF treatment parameters. Next, coupling the built prediction model with Monte Carlo simulation, an empirical cumulative probability distribution of the well productivity is generated. The well HF performance is assessed by estimating its cumulative probability value. The proposed method is applied to evaluate the HF performances in a developed block of the eastern Sulige region. The study shows that 19% of the wells were fractured with good quality and 55% of the wells were fractured with average quality, while the rest were stimulated with poor quality. The evaluations provide a guideline for optimization of HF designs of wells that have not been hydraulically stimulated in the region.

1. Introduction

Tight sand gas has become a new exploration target and important source of natural gas supply in China [1–6]. Sulige gas field, located in the central part of the Ordos Basin, is the largest tight gas field with estimated gas reserves of 18.85 Tcf in China [7–9]. The eastern Sulige is one of the main regions of natural gas exploration and development in the gas field [10, 11]. Due to the low quality of rock properties such as permeability and porosity in the gas-bearing formations, hydraulic fracturing stimulations were carried out so that natural gas can be produced economically from the tight formation.

The hydraulic fracturing is a process which injects large amounts of water, proppants (sand), and additives via a wellbore at high pressure into the low permeable formations to break rock and create cracks (hydraulic fractures). The injected proppants keep fractures open [12–14]. These fractures could extend several hundred feet away the wellbore and form high permeable flow paths, which ease the natural gas flow from the tight formation toward the wellbore and

enhance the gas recovery. In the past several years, hundreds of wells were hydraulically stimulated and put on production in the eastern Sulige region. However, the significant differences in the gas production have exhibited in the fractured wells [15]. To further develop the gas field, it is necessary to evaluate the hydraulic fracture performances and to identify the best hydraulic fracturing practices to maximize the well gas production in the field.

Upon completing of a HF job, well production performance such as absolute open flow potential (AOFPP) or cumulative gas production is employed as an indicator for the hydraulic fracture performance of the well. The well production performance is controlled by multiple variables. In this work, we consider geological properties and fracturing treatment parameters. The geological properties include gas saturation, pay thickness, porosity, matrix permeability, and presence of nature fractures. The treatment parameters mainly contain fluid injection volumes, injection rates and pressure, and proppant volumes, which influence fracture geometries. To conduct the performance evaluation, it is essential to build a predictive model to quantify the well

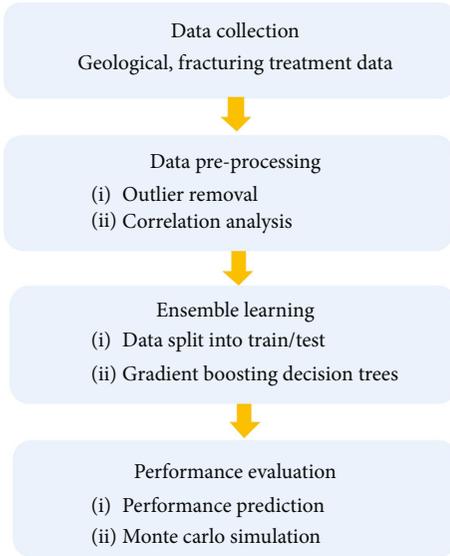


FIGURE 1: Workflow of HF performance evaluation.

production performance with these influential factors. A numerical reservoir simulator is usually employed to build such model by simulating natural gas flow in both tight matrix and the fractures [16–19]. However, the accuracy of the hydrocarbon production predictions highly relies on the characterization of the tight gas reservoirs and representation of induced fracture geometry. Due to limited formation data and subsurface complexities, it is impossible to quantify the properties precisely. Meanwhile, the underlying flow mechanism is so complex in the tight formations that multiple simplified assumptions have been made in the mechanistic modelling [20]. These challenges have made that the estimated hydrocarbon production from the reservoir simulations often differs from actual production data [21]. In addition, the numerical simulation runs are computationally intensive by solving large partial differential equations and make them unfeasible for the probabilistic evaluation of the HF performance, where hundreds of simulations are often required. Numerous data have been collected from various sources and available in the eastern Sulige region. A machine learning method could provide potential solutions to overcome the drawbacks.

In recent years, the machine learning method has become a powerful tool for predicting the production performance. For instance, Nejad et al. [22] developed a neural network model to examine the effect of the completion and fracture treatment designs on gas production in the Eagle Ford Shale. Mohaghegh [23] proposed the concept of shale analytics to describe the applications of artificial intelligence to shale gas development. The shale analytics was applied to investigate how different reservoir and completion parameters affect gas production in several shale plays of the United States. Montgomery and O’Sullivan [24] developed a spatial error model and regression-Kriging to forecast tight oil production using a large well dataset from the Williston Basin in North Dakota. Wang and Chen [25] investigated the performances of four machine-learning algorithms in forecasting first-year oil pro-

duction of Montney tight reservoirs and concluded that random forest outperformed other algorithms. Porras et al. [26] developed a random forest model to predict first-year oil production using geological and completion design parameters. The developed model was used to evaluate hydraulic fracture performance of the horizontal wells in the Viking Formation, Canada. Although the machine learning techniques have shown promising results in the production prediction in the above studies, these studies did not consider the uncertainties related to the complexities of the HF process. The quantification of these uncertainties can provide valuable information to assess the HF performances.

In this study, we develop a probabilistic approach to evaluate the HF performance. Our approach integrates ensemble learning with Monte Carlo simulations to consider the uncertainties caused by the HF treatment complexities. The approach was inspired by Mohaghegh’s combination of neural networks and the Monte Carlo simulations [23]. We applied the developed approach to obtain an empirical cumulative probability distribution of the well productivity. The well HF performance is assessed by estimating its probability value.

2. Methodology

The probabilistic evaluation workflow of the HF performance consists of four main steps, as shown in Figure 1.

- (Step 1) Data collection: collect raw data from a database. The independent variables include geological properties and hydraulic fracturing treatment parameters. The target variable represents post-treatment well production performance.
- (Step 2) Data preprocessing: make raw data ready for building a predictive model. The data preprocessing is composed of two parts: data cleaning and feature selection.
- (Step 3) Ensemble learning: build a model to predict the well production performance using ensemble machine learning.
- (Step 4) Probabilistic evaluation: combine the predictive model with the Monte Carlo simulation to assess the HF performance.

2.1. Data Collection and Preprocessing. We extracted 743 vertical fractured wells from a private database. These wells were chosen without missing values and from the same developed block in the eastern Sulige. As one of China’s natural gas development key areas, the eastern Sulige gas field stretches from the Ordos district in the Inner Mongolia Autonomous Region to the Yulin district of Shaanxi province in China. The reservoir rocks of the gas field mainly consist of the Upper Paleozoic fluvial and deltaic sands [27]. The main gas-producing layers are Permian-age Shihezi formation, Shanxi formation, Benxi formation, and

TABLE 1: Lists of modelling variables.

| Type | Parameters | Unit | Abbreviation |
|-----------------------|------------------------------|-------------------------------------|--------------|
| Geological properties | Formation thickness | m | TH |
| | Well true vertical depth | m | TVD |
| | Formation porosity | % | PO |
| | Matrix permeability | mD | PERM |
| | Gas saturation | % | SG |
| | Pad fluid volume | m ³ | PFV |
| Fracture treatment | Slurry fluid volume | m ³ | SFV |
| | Proppant fluid ratio | % | PFR |
| | Total fluid volume | m ³ | TFV |
| | Average injection rate | m ³ /min | AIR |
| Production | Average injection pressure | MPa | AIP |
| | Absolute open flow potential | 10 ⁴ m ³ /day | AOFP |

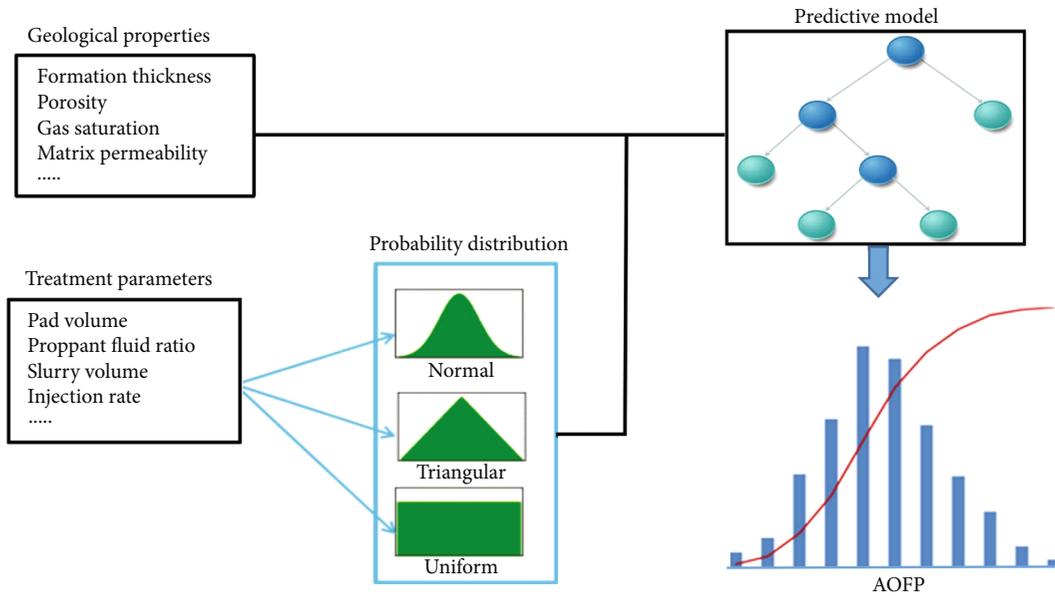


FIGURE 2: Probabilistic evaluation of a fractured well through obtaining cumulative probability distribution of AOFP (modified from [23])

Majiagou formation. The 8th member of Shihezi formation (He8) is one of the most gas productive zones with formation thickness between 45 m and 60 m. The average depth of the tight formations in the study area ranges from 2300 m to 3800 m. The reservoir porosity ranges from 5% to 15%, and the matrix permeability varies from 0.5 to 20 mD [28]. The multistage hydraulic fractured wells have been applied recently in the field. However, the field is still dominated by vertical well development, accounting for more than 80% of the total number of wells [29]. The vertical wells were drilled into the formations He8 and 1st member of Shanxi formation (Shan1) between 2009 and 2016. Based on geological interpretations, most of the vertical wells have several gas-bearing layers with average thickness less than 5 m. To stimulate the separate layer, multilayer staged fracturing techniques such as mechanical packers, casing sleeves, and coiled tubing were developed to enhance fracturing efficiency and increase single-well production. How-

ever, the fractured wells could only keep stable gas production for 1 to 1.5 years; then the production declined rapidly. Therefore, the gas flow rates were allocated properly to maintain a relatively long stable production [30].

For each vertical well, there are twelve variables listed in Table 1.

The target variable is absolute open flow potential (AOFP) to quantify the hydraulic fracture performance. The number of independent variables is eleven. Two kinds of independent variables are identified the most important to the production performance of hydraulically fractured wells: geological properties and HF treatment parameters. The geological properties include formation thickness, true vertical depth (TVD), porosity, matrix permeability, and gas saturation. The HF treatment parameters such as fluid volumes, injection rates, and pressure determine a fracture geometry and conductivity, which play an important role in natural gas flow.

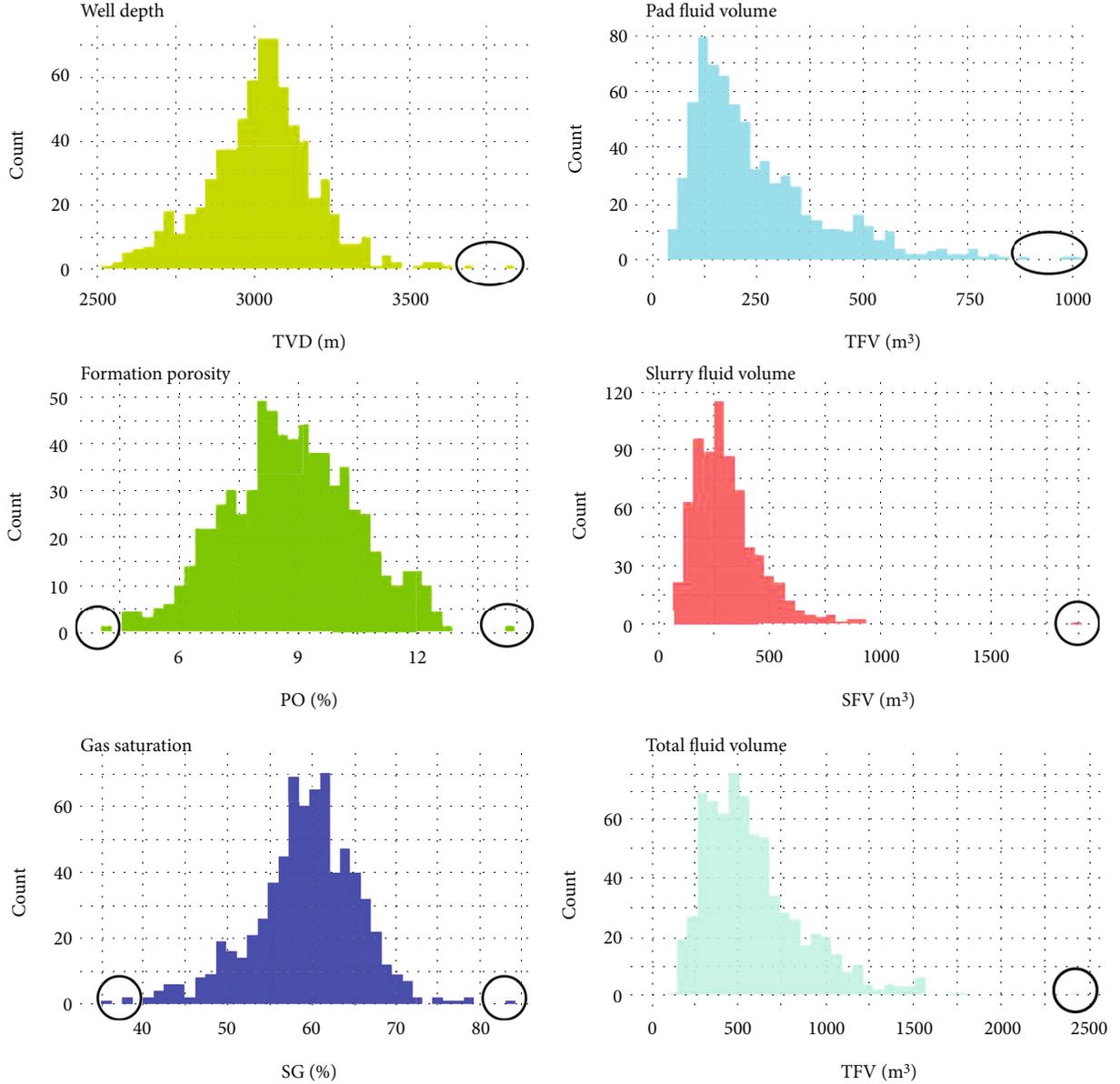


FIGURE 3: Histograms of the main variables.

The collected raw data usually include missing data, incorrect formats, and abnormal data, and the direct usage of the raw data affects later predictive modelling. Data pre-processing is employed to solve the issues, which mainly includes data cleaning, data transformation, and feature selection, which are described in Section 3.

2.2. Ensemble Learning. The ensemble learning method is an advancement in machine learning technique by building multiple learners and combining the outputs of these learners to obtain robust predictions [31]. It has been confirmed that the ensemble learner has better performances than a single learner and has already been applied in many regression and classification problems [32]. Boosting is one of effective ensemble learning methods. Friedman [33, 34] regarded the boosting as the optimization of a loss function and introduced the concept of gradient boosting.

Gradient boosting decision trees (GBDT) is an ensemble algorithm to combine many decision tree models, where each tree is built to minimize the residual error of the previous tree iteratively. The final prediction is integrated from the outputs of all trees. The GBDT algorithm is described briefly as follows.

Assuming that x is a set of features (geological properties and HF treatment parameters) and $F_M(x)$ is a predicted function of the target variable y (absolute open flow potential). Given training data set $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$, a GBDT model $F_M(x)$ is built as the summation of M additive functions $f_m(x)$ based on decision trees $t(x; \mu_m)$ [35, 36]:

$$F_M(x) = \sum_{m=1}^M f_m(x) = \sum_{m=1}^M \theta_m t(x; \mu_m), \quad (1)$$

TABLE 2: Statistical properties of the variables.

| Parameter | Range | Mean | Standard deviation |
|--|--------------|-------|--------------------|
| Absolute open flow (10^4 m ³ /day) | 0.4~21.5 | 5.5 | 3.2 |
| Formation thickness (m) | 1.7~38.0 | 14.6 | 6.6 |
| True vertical depth (m) | 2534~3699 | 3025 | 175.7 |
| Formation porosity (%) | 4.8~12.7 | 8.8 | 1.7 |
| Matrix permeability (mD) | 0.1~5.1 | 0.8 | 0.7 |
| Gas saturation (%) | 30.5~78.5 | 59.2 | 6.3 |
| Pad fluid volume (m ³) | 39.0~830.0 | 249.8 | 158.5 |
| Slurry fluid volume (m ³) | 88.2~933.7 | 310.5 | 153.2 |
| Proppant fluid ratio (%) | 3.1~45.6 | 27.4 | 9.3 |
| Total fluid volume (m ³) | 144.2~1560.0 | 589.7 | 292.6 |
| Injection rate (m ³ /min) | 1.4~11.0 | 3.2 | 1.1 |

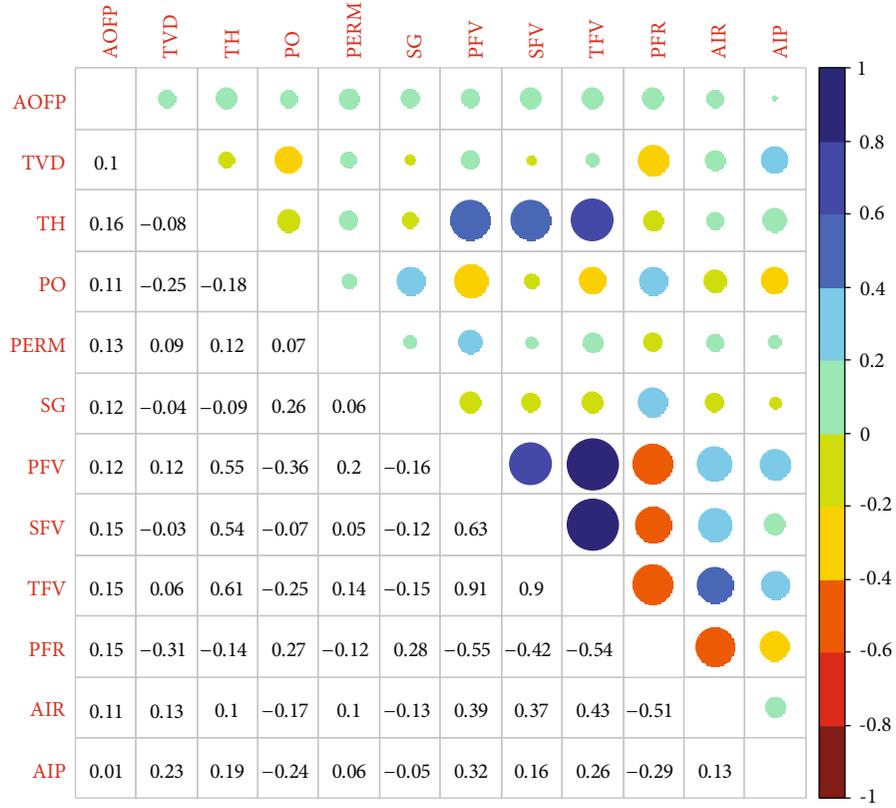


FIGURE 4: The Pearson covariance matrix between features.

where μ_m is the mean of split locations and the terminal nodes of a tree $t(x; \mu_m)$ and θ_m is estimated by minimizing a specified loss function $L(y, F(x)) = (y - F(x))^2$:

$$f_0(x) = \gamma = \operatorname{argmin}_{\gamma} \sum_{i=1}^N L(y_i, \gamma). \quad (2)$$

The number of iteration m starts from 1 to M , and negative gradient g_{im} in the current prediction model is calculated by

$$g_{im} = - \left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)} \right]_{f(x)=f_{m-1}(x)}. \quad (3)$$

The tree $t(x; \mu_m)$ is used to approximate the negative gradient, and a descent step size θ_m is computed by

$$\theta_m = \operatorname{argmin}_{\theta} \sum_{i=1}^N L(y_i, F_{m-1}(x_i) + \theta t(x; \mu_m)). \quad (4)$$

TABLE 3: Optimal hyperparameter values of the GBDT.

| Hyperparameter | Name | Ranges | Optimal values |
|--------------------------|-------------------|------------|----------------|
| No. of trees | N_estimators | 1~200 | 110 |
| Max. depth | max_depth | 2~8 | 3 |
| No. of features to split | max_features | sqrt; auto | sqrt |
| Min. number of samples | min_samples_split | 2~12 | 4 |
| Learning rate | learning_rate | 0.08~0.13 | 0.11 |

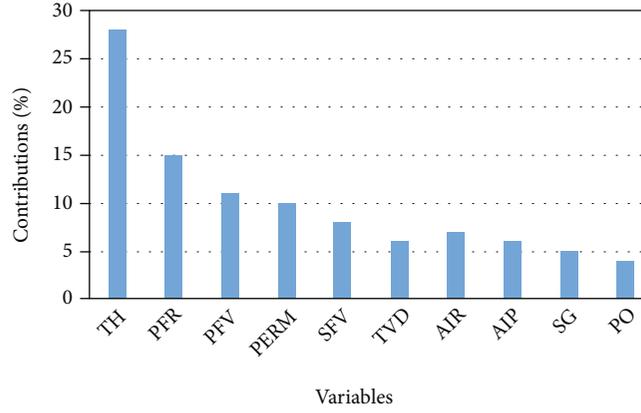


FIGURE 5: Relative variable importance.

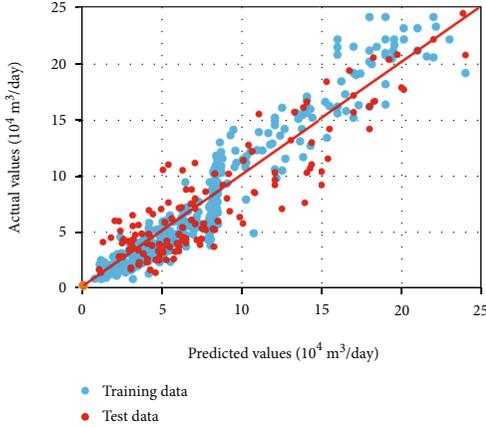


FIGURE 6: Comparison of actual and predicted AOPF for training and test data.

The model is updated based on Equations (2)–(4):

$$F(x) = F_{m-1}(x) + \theta_m t(x; \mu_m). \quad (5)$$

The GBDT has a few advantages, including the ability to cope with skewed variables, computational robustness, and high scalability [37]. More details of the algorithm can be found from Zhu et al. [27].

2.3. Probabilistic Evaluation. Due to the complexity of hydraulic fracturing process, the production prediction has some uncertainty, which strongly affects the HF performance evaluation.

TABLE 4: Properties of well Sudong42-48.

| Parameter | Value |
|--|--------|
| Formation thickness (m) | 17.2 |
| True vertical depth (m) | 3138.9 |
| Formation porosity (%) | 7.2 |
| Matrix permeability (mD) | 0.4 |
| Gas saturation (%) | 54.3 |
| Pad fluid volume (m ³) | 425 |
| Slurry fluid volume (m ³) | 380 |
| Proppant fluid ratio (%) | 9.7 |
| Injection rate (m ³ /min) | 3.5 |
| Injection pressure (MPa) | 67.5 |
| Absolute open flow (10 ⁴ m ³ /day) | 8.5 |

Monte Carlo simulations can be used to quantify the prediction uncertainty by estimating an empirical cumulative probability distribution of the well productivity instead of a determined value. As shown in Figure 2, we fixed the values of the geological properties of a selected well and assumed the fracture treatment parameters as random variables to follow a certain probability distribution such as triangular or Gaussian or uniform. The ranges, means, and variances of these distributions are estimated from the dataset. We run the predictive model thousands of times by randomly sampling from the given distributions. At the end of the runs, an empirical cumulative probability distribution of the selected well AOPF is generated. Through the probability distribution, the HF performance of the well could be evaluated by estimating its probability value. The details are described in the next section.

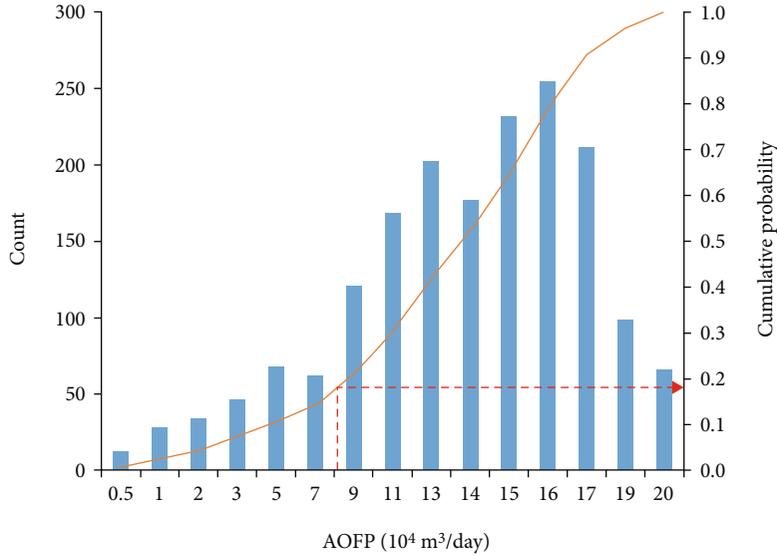


FIGURE 7: Histogram and probability distribution of a well AOPF.

3. Results

3.1. Outlier Removal. Figure 3 displays the presence of outliers through histograms of the main input variables. The outliers in the data are circled in black for the variables. We also applied Rosner’s test to verify the outliers identified in the histograms [38]. A total of 39 wells with the outliers were detected from the dataset and removed. 704 wells were kept to the following predictive modelling.

Table 2 summarizes the range and statistical properties of the independent and dependent variables. The absolute open flow potential per well ranges from 0.4 to 21.5 ($10^4 \text{ m}^3/\text{day}$), averaging $5.5 \times 10^4 \text{ m}^3/\text{day}$. The total injection fluid volume and proppant fluid ratio per well are averaged to be 589.7 m^3 (27.4%), respectively.

3.2. Correlation Analysis. Figure 4 shows the Pearson covariance matrix to quantify the degree of linear correlation among the variables. It is noted that all treatment parameters have a positive correlation with the well productivity (AOPF). The increasing in the fracturing fluid volumes enlarges the reservoir stimulated zones, which creates more contact area between the wellbore and reservoir. As expected, there is the multicollinearity among the pad fluid volume (PFV), slurry fluid volume (SFV), and total fluid volume (TFV) since total fluid volume consists of the pad fluid volume and slurry fluid volume. Therefore, the total fluid volume will be dropped in the following predictive modelling. In addition, the geological properties also have a positive correlation with the AOPF. The formation thickness has the largest value of the Pearson coefficient. More sophisticated feature selection methods including stepwise regression and recursive feature elimination are being investigated.

3.3. Production Forecasting. The 704 wells in the dataset were split into training and test data sets in the ratio of 80:20. The main hyperparameters for a GBDT model

include learning rate, number of trees, minimum number of samples required at a leaf node, maximum depth, and number of features for the best split. The hyperparameter tuning were performed on the training data using grid searching combined with five-fold cross-validation (CV). The optimal hyperparameter values are listed in Table 3.

Figure 5 shows the results of the relative importance of 10 input variables on the AOPF. The formation thickness and proppant fluid ratio are two most important variables, followed by the pad fluid volume and matrix permeability. The variable importance indicates that the formation thickness has a significant impact on the well productivity. Therefore, when fracturing a gas well, we need to evaluate its formation thickness and matrix permeability. In terms of the fracturing operation, we may add more proppant to enhance the fracture conductivity and increase the pad fluid volume to improve the fracturing effectiveness.

Figure 6 compares the actual and predicted AOPF for the training and test data sets using the built GBDT model.

In the figure, the data points are grouped along the 45-degree straight line, and the values of coefficient of determination R^2 for the training and test data set are calculated to be 0.91 and 0.74, respectively. The results show that the predicted AOPF have a good match with the actual AOPF values for both the training and test data sets, which indicates the developed AOPF forecasting model is robust for the evaluation of the hydraulic fracture performance.

3.4. Well Hydraulic Performance Evaluation. We choose well Sudong42-48 to assess its hydraulic fracturing performance as an example. Table 4 shows the input variable and AOPF values of this well.

We fixed the values of the geological properties of this well and assumed the fracture treatment parameters as random variables to follow a certain probability distribution such as triangular or Gaussian or uniform. We run the predictive model thousands of times by sampling the treatment

TABLE 5: Evaluation criteria.

| Hydraulic fracture performance | Cumulative probability |
|--------------------------------|------------------------|
| Good | >0.6 |
| Fair | [0.3, 0.6] |
| Poor | <0.3 |

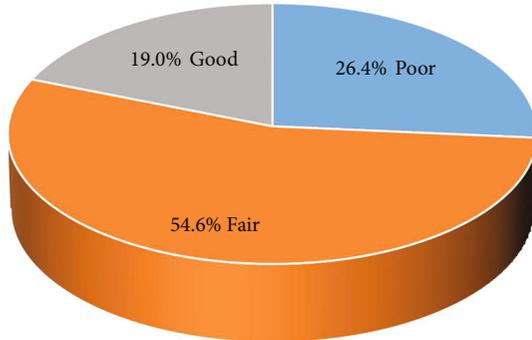


FIGURE 8: Evaluation results of hydraulic fracture performance in the block of the eastern Sulige region.

parameters from the given distributions. An estimated AOFPP was obtained from each run. At the end of the runs, a histogram and empirical cumulative probability distribution of the selected well were generated as shown in Figure 7. The well hydraulic performance is assessed by estimating its cumulative probability based on its AOFPP, as shown in Figure 7. The well AOFPP is $8.5 \times 10^4 \text{ m}^3/\text{day}$, which is marked as a filling circle in the x -axis. The corresponding cumulative probability is 0.18 to determine the hydraulic fracturing performance. The hydraulic fracturing quality of this well was classified as “poor” according to the evaluation criteria given in Table 5 because its cumulative probability value is less than 0.3. Taking the same evaluation process, we have assessed the hydraulic fracturing performances of other vertical wells in the developed block of the eastern Sulige region.

From the variable importance analysis, the proppant fluid ratio (FPR) is the most influential factor among the fracture treatment parameters. If we increased FPR value from 9.7% (original) to 23%, the well productivity was increased by 29%, and the hydraulic fracturing quality of this well could be classified as “Fair.” Therefore, the evaluation is useful to optimize the HF treatment.

We performed the assessments for all 704 fractured wells. The evaluation results of the block are shown in Figure 8. 54.6% of the wells in the block have been fractured with “Fair” quality, 26.4% wells with “Poor” quality, and only 19% wells with “Good” quality. The results indicate that the better HF designs are required to improve the HF performances in the gas field.

4. Conclusions

In the paper, we have developed a probabilistic workflow to assess the hydraulic fracture performance by integrating the ensemble learning with the Monte Carlo simulation. Using

the data from a developed block in the eastern Sulige region, we have applied the workflow to evaluate the hydraulic fracture performance of the wells.

- (1) The absolute open flow potential is regarded as the response variable, while ten geological and fracture treatment parameters are chosen as the input variables
- (2) An ensemble learning model is built to quantify complex relationship between the geological properties, treatment parameters, and the absolute open flow potential. Results indicate that the formation thickness has the most important effect on the well productivity, followed by the proppant fluid ratio
- (3) Among 704 fractured vertical wells, 19% of the wells have been stimulated with “Good” quality, 54.6% of the wells with “Fair” quality, and 26.4% wells with “Poor” quality
- (4) The proposed workflow could be applied to evaluate other well stimulation performances such as multi-stage hydraulic fracturing in horizontal wells or acid fracturing

Data Availability

Data is available upon request.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgments

This research is funded by the Project of National Natural Science Foundation of China (Grant Nos. 51974253 and 51934005), the Youth Project of National Natural Science Foundation of China (Grant No. 52004219), the Scientific Research Program Funded by Education Department of Shaanxi Province (Grant Nos. 18JS085 and 20JS117), and the Natural Science Basic Research Program of Shaanxi (Grant Nos. 2020JQ-781 and 2017JM5109).

References

- [1] H. Li, A. Jia, and D. He, “Technology advancement and prospect of natural gas development in Petrochina,” *Natural Gas Industry*, vol. 30, no. 1, pp. 5–7, 2010.
- [2] C. Zou, R. Zhu, S. Wu Songtao et al., “Types, characteristics, genesis and prospects of conventional and unconventional hydrocarbon accumulations: taking tight oil and gas in China as an instance,” *Acta Petrolei Sinica*, vol. 33, no. 2, pp. 173–187, 2012.
- [3] C. Zou, G. Zhai, G. Zhang et al., “Formation, distribution, potential and prediction of global conventional and unconventional hydrocarbon resources,” *Petroleum Exploration and Development*, vol. 42, no. 1, pp. 14–28, 2015.

- [4] X. Ma, A. Jia, J. Tan, and H. Dongbo, "Tight sand gas development technology and practices in China," *Petroleum Exploration and Development*, vol. 39, no. 5, pp. 611–618, 2012.
- [5] S. Longde, Z. Caineng, J. Ailin et al., "Development characteristics and orientation of tight oil and gas in China," *Petroleum Exploration and Development*, vol. 46, no. 6, pp. 1015–1026, 2019.
- [6] T. Lu, Y. Liu, L. Wu, and X. Wang, "Challenges to and countermeasures for the production stabilization of tight sandstone gas reservoirs of the Sulige gas field, Ordos Basin," *Natural Gas Industry*, vol. 35, no. 6, pp. 43–52, 2015.
- [7] H. Yang, J. Fu, and X. Liu, "Accumulation conditions and exploration and development of tight gas in the Upper Paleozoic of the Ordos Basin," *Petroleum Exploration and Development*, vol. 39, no. 3, pp. 315–324, 2012.
- [8] Z. Tan, T. Lu, Y. Liu, L. Wu, and Y. Yang, "Technical ideas of recovery enhancement in the Sulige Gasfield during the 13th Five-Year Plan," *Natural Gas Industry*, vol. 3, no. 3, pp. 234–244, 2016.
- [9] G. Ji, A. Jia, D. Meng et al., "Technical strategies for effective development and gas recovery enhancement of a large tight gas field: a case study of Sulige gas field, Ordos Basin, NW China," *Petroleum Exploration and Development*, vol. 46, no. 3, pp. 629–641, 2019.
- [10] J. Wang, C. Zhang, and J. Li, "Tight sandstone gas reservoirs in the Sulige gas field: development understandings and stable-production proposals," *Natural Gas Industry*, vol. 41, no. 2, pp. 100–110, 2021.
- [11] J. Guo, Z. Guo, and Y. Cui, "Recovery factor calculation method of giant tight sandstone gas field," *Acta Petrolei Sinica*, vol. 39, no. 12, pp. 1389–1396, 2018.
- [12] H. Belyadi, E. Fathi, and F. Belyadi, *Hydraulic Fracturing Fluid Systems*, Gulf Professional Publishing, 2017.
- [13] T. Jiang, *Fracturing Technology and Case Analysis of Complex and Difficult Oil and Gas Reservoirs*, Sinopec Press, Beijing, 2017.
- [14] J. Li, J. Bai, and L. Zhu, "Volume fracturing and its practices in Sulige tight sandstone gas reservoirs, Ordos Basin," *Natural Gas Industry*, vol. 33, no. 9, pp. 65–69, 2013.
- [15] Z. Liu, Q. Wang, L. Zhang, J. Chuanjie, and G. Haipeng, "Laws for gas well production decline in east area of Sulige gas field, Ordos Basin," *Xinjiang Petroleum Geology*, vol. 36, no. 1, pp. 82–85, 2015.
- [16] C. L. Cipolla, E. P. Lonon, J. C. Erdle, and B. Rubin, "Reservoir modeling in shale-gas reservoirs," *SPE Reservoir Evaluation Engineering*, vol. 13, no. 4, pp. 638–653, 2010.
- [17] W. Yu, B. Gao, and K. Sepehrnoori, "Numerical study of the impact of complex fracture patterns on well performance in shale gas reservoirs," *Journal of Petroleum Science Research*, vol. 3, no. 2, pp. 83–89, 2014.
- [18] X. Li, D. Zhang, and S. Li, "A multi-continuum multiple flow mechanism simulator for unconventional oil and gas recovery," *Journal of Natural Gas Science and Engineering*, vol. 26, pp. 652–669, 2015.
- [19] Y. Xu, J. S. A. Cavalcante Filho, W. Yu, and K. Sepehrnoori, "Discrete-fracture modeling of complex hydraulic-fracture geometries in reservoir simulators," *SPE Reservoir Evaluation Engineering*, vol. 20, no. 2, pp. 403–422, 2017.
- [20] C. R. Clarkson, "Production data analysis of unconventional gas wells: review of theory and best practices," *International Journal of Coal Geology*, vol. 109–110, pp. 101–146, 2013.
- [21] S. D. Mohaghegh, "Reservoir simulation and modeling based on pattern recognition," in *Paper presented at the SPE Digital Energy Conference and Exhibition*, the Woodlands, Texas, USA, 2011.
- [22] A. M. Nejad, S. Sheludko, T. Hodgson, R. McFall, and R. F. Shelley, "A case history: evaluating well completions in the Eagle Ford Shale using a data-driven approach," in *Paper presented at the SPE Hydraulic Fracturing Technology Conference*, the Woodlands, Texas, USA, 2015.
- [23] S. D. Mohaghegh, *Shale Analytics: Data-Driven Analytics in Unconventional Resources*, Springer, Cham, Switzerland, 2017.
- [24] J. B. Montgomery and F. M. O'Sullivan, "Spatial variability of tight oil well productivity and the impact of technology," *Applied Energy*, vol. 195, pp. 344–355, 2017.
- [25] S. Wang and S. Chen, "Insights to fracture stimulation design in unconventional reservoirs based on machine learning modeling," *Journal of Petroleum Science and Engineering*, vol. 174, no. 6, pp. 682–695, 2019.
- [26] L. Porras, C. Hawkes, and I. Arshad, "Evaluation and optimization of completion design using machine learning in an unconventional light oil play," in *Proceedings of the 8th Unconventional Resources Technology Conference*, 2020.
- [27] X. Zhu, C. Liu, D. Zhong, and F. Han, "Diagenesis and their succession of gasbearing and non-gas-bearing reservoirs in the Sulige gas field of Ordos Basin, China," *Acta Geologica Sinica-English Edition*, vol. 83, no. 6, pp. 1202–1213, 2009.
- [28] M. Wang, H. Tang, F. Zhao et al., "Controlling factor analysis and prediction of the quality of tight sandstone reservoirs: a case study of the He8 Member in the eastern Sulige Gas Field, Ordos Basin, China," *Journal of Natural Gas Science and Engineering*, vol. 46, pp. 680–698, 2017.
- [29] J. Fu, L. Fan, X. Liu, and D. Huang, "Gas accumulation conditions and key exploration & development technologies in Sulige gas field," *Acta Petrolei Sinica*, vol. 40, pp. 240–255, 2019.
- [30] A. Jia, D. He, and J. Tang, "The development strategies for Gas fields of low permeability, low abundance and in heterogeneity," in *International Oil & Gas Conference and Exhibition in China*, Beijing, China, 2006.
- [31] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, Springer, second edition, 2017.
- [32] Z. Zhou, *Ensemble Methods: Foundations and Algorithms*, CRC Press, 2012.
- [33] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–12327, 2001.
- [34] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics and Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [35] M. Liu, H. Chen, D. Wei, Y. Wu, and C. Li, "Nonlinear relationship between urban form and street-level PM2.5 and CO based on mobile measurements and gradient boosting decision tree models," *Building and Environment*, vol. 205, article 108265, 2021.
- [36] C. Ding, D. Wang, X. Ma, and H. Li, "Predicting short-term subway ridership and prioritizing its influential factors using gradient boosting decision trees," *Sustainability*, vol. 8, no. 11, 2016.
- [37] J. Tang, B. Fan, L. Xiao et al., "A new ensemble machine-learning framework for searching sweet spots in shale reservoirs," *SPE Journal*, vol. 26, no. 1, pp. 482–497, 2021.
- [38] V. Barnett and T. Lewis, *Outliers in Statistical Data*, John Wiley & Sons, Chichester, UK, third edition, 1995.