The Production Performance Evaluation of Hydraulically Fractured Wells in the East Sulige Field Using Machine Learning

Xianlin Ma*, Desheng Zhou, and Wenbing Cai
College of Petroleum Engineering, Xi’an Shiyou University, Xi’an, China

Abstract. The paper presents a comprehensive workflow to integrate the machine learning algorithm with the Monte Carlo simulation, and a field example is provided to demonstrate that the proposed workflow could reasonably capture the behaviour of well production data. The workflow helps engineers in learning valuable lessons from their historical operations to optimize the future hydraulic fracturing treatments in the Sulige gas field.

1 Introduction

Unconventional tight reservoirs like the Sulige field has become major energy resources in China [1]. Due to the low permeability nature of the rock formation in the gas field, hydraulic fracturing treatments were carried out to increase and maintain well gas productivity. Hydraulic fracturing is a process which injects high-pressure liquid into a petroleum-bearing rock formation to create new fractures or increase existing fractures via a well. This process facilitates the flow of natural gas toward the wellbore by increasing a contact area between the well and the formation [2]. An enormous amount of data have accumulated from well drilling, stimulation and production in the field during the past several years. The production history demonstrated that the significant variations in the well production performance in the field [3]. To further develop the gas field, it is necessary to evaluate and analyze the data generated to understand which hydraulic fracturing parameters control the well productivity, and identify the best hydraulic fracturing practices to maximize the well gas production. The available data enable the application of machine learning algorithms to evaluate the production performance of the hydraulically fractured well in the gas field. Taking the data from a block in the East Sulige field as an example, we first explain preparation of the geological and hydraulic fracturing data used in the machine learning models. We then describe a workflow to evaluate the well production performance. Finally we demonstrate the workflow usage by an example.

2 Data preparation

The data in this study is from the gas wells in a Sudong block. A total of 244 vertical producing wells were obtained from a well completion and fracture database. For each vertical well, there are 13 input variables. These variables are subdivided into two groups:
geological subsurface parameters and fracturing treatment parameters as shown in Table 1. Absolute open flow (AOF) potential is presented as a response variable to quantify a well productivity.

| Type          | Parameter                          | Unit   | Variable Name    |
|---------------|------------------------------------|--------|------------------|
| Geology       | Formation thickness                | m      | Gross_thick      |
|               | Perforated thickness               | m      | Perf_thick       |
|               | Clay content                       | %      | Clay_vol         |
|               | Porosity                           | %      | Porosity         |
|               | Matrix permeability                | mD     | Permeability     |
|               | Gas saturation                     | %      | Sg               |
| Fracture      | Pad fluid volume                   | m³     | Pad_vol          |
|               | Slurry fluid volume                | m³     | Slurry_vol       |
|               | Proppant concentration             | %      | Proppant_per     |
|               | Flush fluid volume                 | m³     | Flush_vol        |
|               | Total fracturing fluid volume      | m³     | Total_vol        |
|               | Average injection rate             | m³/min | Pump_rate        |
|               | Average injection pressure         | MPa    | Treat_press      |
| Productivity  | Absolute open flow potential       | 10⁴ m³/day | AOF               |

The input variables may be correlated with each other. The Pearson correlation coefficients were calculated to measure the linear correlation between the parameters as shown in Figure 1. This figure shows that the formation thickness and perforated thickness have a strong positive correlation (0.86), and matrix permeability is strongly correlated to porosity. As for the fracture treatment parameters, most of them are strongly correlated each other, with a few exceptions. There are strong correlations among the pad, slurry, flush, and total fluid volumes. As pointed out by Hirschmiller [4], the correlated parameters could compromise the predictive model usage, stability and interpretability, it is often desirable to reduce the number of input parameters although it is impossible to avoid correlated parameters altogether.

Based on the correlation analysis, the following 9 input parameters are selected to build predictive models:
- formation thickness, clay content, porosity, gas saturation
- pad volume, slurry volume, proppant concentration, average injection rate, average injection pressure

### 3 Workflow

We must first decide on which machine learning algorithm works best to predict the well productivity since there are so many machine learning algorithm available such as ordinary least squares regression, support vector machine, neural network, and tree-based methods including decision tree, random forest, gradient boosting model. A comparison of their relative strengths and weakness was listed by Schuetter [5]. The optimal algorithm suitable for the data set is determined by trying the multiple algorithms through trial and error. The data set is randomly split into training and testing sets by ratio 4:1.
Figure 1. Correlation matrix between the input variables.

We used the training set to fit machine learning models, then used the fitted models to predict the well production in the testing set. To assess the model predictive performances on the testing set, the two metrics are used: mean squared error (MSE) and $R^2$ value. Table 2 compares the metrics for 6 machine learning methods, which shows that the gradient boosting model (GBM) has best prediction performance for the testing set. The predictive model built using the GBM will be applied to evaluate the production performance of the fractured wells.

Table 2. Metrics of various machine learning methods on the testing set.

| Methods                  | MSE    | $R^2$ |
|--------------------------|--------|-------|
| Linear Regression        | 7.96   | 0.43  |
| Support Vector Machine   | 7.60   | 0.40  |
| Neural Network           | 8.54   | 0.46  |
| Decision Tree            | 8.05   | 0.43  |
| Random Forest            | 7.25   | 0.45  |
| Generalized Boosted Model| 6.43   | 0.56  |
After determining the best machine learning algorithm for the data set, a probabilistic method was developed for evaluating the well production performance by combining the built predictive model with the Monte Carlo simulation as shown in Figure 2, modified from [6].

**Figure 2.** Workflow to generate the cumulative probability distribution of a well AOF.

In this figure, the values of the geological subsurface parameter inferred from well logs were fixed, while each of the fracture treatment parameters was assumed to follow a certain probability distribution such as triangular or Gaussian or uniform. The ranges, means and variances of these distribution were estimated the data set. The predictive model was run 1000 times by sampling the treatment parameters from the given distributions. The result of each run was an absolute open flow (AOF) potential value under a random combination of 5 fracture treatment parameters (pad volume, slurry volume, proppant concentration, average injection rate, average injection pressure), called a hydraulic fracturing scenario. 1000 times of the runs generated a histogram plot and empirical cumulative probability distribution of the AOF potential for the selected vertical well.

By generating the probability distribution from the Monte Carlo simulation, the well production performance can be evaluated by estimating its cumulative probability based on its actual production value as shown in Figure 3. Figure 3 presented the probability distribution of the selected well generated by the Monte Carlo simulation. The actual AOF value of the well is $8.25 \times 10^4 \text{m}^3/\text{day}$, which was marked “x” symbol in the x-axis. The corresponding cumulative probability $P_x$ is 0.77, which determines the quality of hydraulic fracturing treatment. The hydraulic fracturing quality of this well was classified as “very good” according to the evaluation criteria given in Table 3. The hydraulic fracturing quality of other vertical wells in the block can be assessed like this well.

**Table 3.** Evaluation criteria.

| Well performance | Cumulative probability |
|------------------|------------------------|
| Excellent        | $P_x \geq 0.8$         |
| Very Good        | $0.6 =< P_x < 0.8$     |
| Good             | $0.4 =< P_x < 0.6$     |
| Fair             | $0.2 =< P_x < 0.4$     |
| Poor             | $< 0.2$                |
4 Conclusion

We proposed a probabilistic approach to assess the fractured well production performance by integrating the machine learning algorithm with the Monte Carlo simulation. Using the data collected from the vertical wells of a Sudong block, we presented the workflow, and how to use the workflow by an example. The presented workflow may help to identify the best hydraulic fracturing practices and improve hydraulic fracturing efficiency.

This work is partially funded by the National Natural Science Foundation of China (Grant No. 51974253 and 51934005), Natural Science Basic Research Plan in Shaanxi Province of China (Program No. 2017JM5109) and Scientific Research Program by Shaanxi Provincial Education Department (Program No. 18JS085).

References

1. H. Yang, X. Wei, New Progress Achieved By Natural Gas Exploration in Sulige Area, Natural Gas Industry, 27, 12, 6-11 (in Chinese) (2007).
2. R.W.J. Veatch, Overview of Current Hydraulic Fracturing Design and Treatment Technology – Part 1. J Pet Technol, 35, 4, 677-687 (1983).
3. X. Bai, Y. Wang, S. Yao, et al., Analysis on Productivity Influential Factors in Sugeli Gas Field, J. of Oil and Gas Technology, 30, 5, 344-348 (in Chinese) (2008).
4. J. Hirschmiller, A. Biryukov, B. Groulx, et al., The Importance of Integrating Subsurface Disciplines with Machine Learning when Predicting and Optimizing Well Performance, Society of Petroleum Engineers (2019).
5. J. Schuetter, S. Mishra, M. Zhong, et al., A Data-Analytics Tutorial: Building Predictive Models for Oil Production in an Unconventional Shale Reservoir, Society of Petroleum Engineers (2018).
6. S.D. Mohaghegh, Shale Analytics: Data-Driven Analytics in Unconventional Resources, Springer (2017).