Skoltech

Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Viktor Duplyakov

PhD Program: Petroleum Engineering

Title of Thesis: Machine learning on field data for hydraulic fracturing design optimization

Supervisor: Professor Andrey Osiptsov

Co-supervisor: Professor Evgeny Burnaev

Name of the Reviewer: Alexander Bernstein

I confirm the absence of any conflict of interest	**Date: 11-09-2023**

Review

The dissertation work "Machine learning on field data for hydraulic fracturing design optimization" is written on an important and relevant topic, motivated by growing demand for hydrocarbon resources and the need to maximize the efficiency and productivity of hydraulic fracturing operations.

As part of the scientific direction of developing methods for fracture stimulation treatments, based on a deep understanding of reservoir characteristics, fluid dynamics and operational parameters, the dissertation examines one of the most important tasks in this area - optimizing the design of hydraulic fracturing.

The ultimate goal of the dissertation research is to create a data-driven fracturing optimization model for reducing the workload of petroleum engineers and enable them to make informed decisions based on proven results from previous projects, which substantiates the relevance and practical significance of the research. The dissertation as a whole is a systematic and high-quality scientific research. The structure of the dissertation corresponds to the essence of the research and allows to adequately present the results. The topic of dissertation work corresponds to its actual content.

To achieve the goal, the dissertation poses and solves two important and interrelated scientific problems. The first task is to build production forecasting based on historical data with using machine learning methods. The second task is to maximize production by choosing optimal fracture design parameters, one of the results of which should be a recommendation system to guide production stimulation engineers in making informed decisions.

To solve the first problem it was necessary

- reasonably select historical data to construct the required forecast from them, which includes the selection of fields and multi-stage hydraulic fracturing operations on them, the choice of input variables describing them (related to the reservoir, well and fracture design parameters) and output variables characterizing production;

- collect, preprocess and store data in a database, which includes the development of a database architecture and methods for eliminating outliers and filling gaps in data;

- building a production forecast based on the collected data.

To solve the first problem it was necessary

- identify physically explainable dependencies between target and design parameters, developing appropriate methods for this;

- accurately formulate and solve the problem of choosing optimal fracture design parameters to maximize production.

Both tasks were successfully solved and the following important scientific results were obtained:

- an unique digital database containing the field data from over 6000 multi-stage hydraulic fracturing jobs in 23 oil fields was created. The input variables consist of 92 parameters relating to the reservoir, well, and fracture design parameters and output variables consist of 16 parameters, including cumulative oil production. The created database, with its large number of data points, is a rare and representative sample compared to other datasets in the literature, which typically have only tens or hundreds of points;
- missing values are filled in the database through collaborative filtering and outliers are removed by visualizing the data structure using the t-SNE algorithm;
- production forecasting based on collected data was constructed using the combination of Ridge Regression and CatBoost algorithms, with a predictive ability of 64% as measured by the coefficient of determination (R2);
- using the constructed forecast, the physically explainable dependencies between the target (cumulative fluid production) and design parameters (the number of stages, proppant mass, average and final proppant concentrations, fluid rate, and others) were found;
- new methods based on Euclidean and cosine distances were developed for assisting the field engineers in analyzing previous fracturing treatments by searching the similar wells. These methods were also used in a workflow to determine the optimization parameters boundaries for a pilot well during the field testing of the methodology;
- combining various optimization methods (surrogate-based optimization, sequential least squares programming, particle swarm optimization, and differential evolution), the optimal fracture design was obtaining and used in created recommendation system intended to guide production stimulation engineers in making informed decisions.

In general, the considered dissertation work was done on an important and relevant topic. It poses and solves important theoretical and practical problems motivated by real engineering applications. All the obtained scientific results are new and have theoretical and practical significance.

Therefore, the dissertation work satisfies all the requirements for Ph.D. theses in the field of Petroleum Engineering, and its author Viktor Duplyakov deserves to be awarded an academic degree.

Provisional Recommendation
[V] *I recommend that the candidate should defend the thesis by means of a formal thesis defense*
☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*
☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*