URTeC: 5537

Applications of Machine Learning for Estimating the Stimulated Reservoir Volume (SRV)

Ali Rezaei*[1,2], Fred Aminzadeh[1,2], Eric VonLunen[2], 1. University of Houston, 2. FACT Inc,
* Now with Schlumberger.

This paper was prepared for presentation at the Unconventional Resources Technology Conference held in Houston, Texas, USA, 26-28 July 2021.

Abstract

Hydraulic fracturing process is an integrated part of the wellbore completions in unconventional reservoirs. Typically, the process is designed before executing the job, aiming at optimizing the final fracture geometry and increasing stimulated reservoir volume (SVR). The physics-based models used for designing purposes are typically built on several simplified assumptions and do not match the SRV estimates from field observations. This work proposes a data-driven and machine learning-based approach for estimating SRV in unconventional reservoirs.

A dataset from the Marcellus Shale Energy and Environment Laboratory (MSEEL) project is used in this study. The model's input data include stimulation parameters of 58 stages of two wells (MPI-3H and MPI-5H). The model output consists of the size of the corresponding microseismic (M.S.) cloud to each stage. Because of the limited number of stages (58) and to make the predictions close to near-real-time, each stage and its corresponding M.S. events are broken into steps, each having unique operational parameters (e.g., proppant and fluid volume and injection rate, among other parameters). This approach helped us to increase the number of required samples for data-based modeling to 829 samples. A standard procedure, including data cleaning, normalization, exploratory data analysis, and input data split (20/80), is then applied to the data. This is followed by various machine learning algorithms used to predict SRV. The respective performances of different methods are compared against each other.

Microseismic (M.S.) monitoring is commonly used to monitor fracture topology evolution over time during the fracturing process. The recorded M.S. clouds that are observed during the hydraulic fracturing process can give a rough estimate of the stimulated reservoir volume (SRV). In this approach, a volume (or area in 2D) that encloses most of the M.S. events can be estimated at different time windows and used as the model output. All models were validated on the test set, and a good match was obtained. Our approach will be the first step toward real-time data-based modeling of "Dynamic SRV" or DSRV, which can be used to provide a better understanding of fracture propagation in unconventional reservoirs. It also can be used to optimize the well stimulation process before executing the job. Moreover, the developed model can be trained and used for other unconventional reservoirs.

The paper's novelty is utilizing a data-based approach to estimate DSRV or changes in the SRV values using fracturing parameters that can help optimize the stimulation of unconventional reservoirs. Based on the model, we will give guidelines that will allow operators to design more efficient fracturing jobs for maximum recovery and monitor the effectiveness of the hydraulic fracturing process.

Introduction

Unconventional reservoirs play a central role in the increased US oil and gas production in the last decade. Developing these reservoirs depends mainly on multiple parameters such as formation properties, fluid type, and how the horizontal wells are drilled and fractured. Unlike the classical two-wing fracture models in higher permeability reservoirs, in tight formations filled with natural fractures, the induced fractures typically create a complex fracture network (Warpinski, 2008). As a result of the complex fracture network, estimating the extent of the stimulated reservoir using a conventional fracturing simulator is hard and often gives erroneous results. Stimulated reservoir volume (SRV) is a volume created during and after the fracturing process, representing the enhanced permeability zone contributing to hydrocarbon production. Microseismic monitoring is a method that has been extensively used to track the complex growth of fractures and estimate the SRV extent. This study proposes a data-driven approach using fracturing parameters as input to predict SRV in a near-time fashion.

Physics-based modeling and data-based approaches are standard practices for building predictive models that can be used for many applications in the oil and gas industry, including but not limited to the forecastings the production and optimizing the completion design (e.g., Rezaei et al. 2019; Rezaei et al., 2020; Sidiqqui et al., 2019). The physics-based approach includes analytical or numerical solutions to the problem and often is based on several simplifications. The most commonly used analytical and numerical models to predict the final geometry of the created fractures are listed by Adachi et al. (2007) and Lecampion et al. (2018). On the other hand, the data-based approach includes correlations, cross-plots, analog fields, etc., and depends heavily on data availability. The advantage of the physics-based models is that it gives better predictions for "unseen scenarios," but often requires a significant number of parameters, and fails in several cases due to the uncertainties in in-situ rock and fluid properties. While the data-based approach does not require as many parameters yet may fail in extrapolation, it also requires a considerable amount of example data to train the models. These models have gained popularity in recent years due to the massive increase in the amount of produced data (big data). The produced data can be utilized to overcome the shortcomings of the data-based approach. Thus, it has now become feasible to train a data-based model to create reliable tools that can be used for the prediction of oil and gas production. Data-based approaches can be grouped into two main classes: supervised learning and unsupervised learning. Supervised learning is building a function from a set of labeled input data, whereas unsupervised learning is used to find hidden structures of unlabeled data. The most popular data analysis approaches models are least-square regression, gradient descent, support vector regression, random forest, genetic algorithm, gradient boosting models, artificial neural network, self-organizing maps.

Using artificial neural network (ANN) models and other machine learning models in the oil industry has increased in the last decade. The literature contains a wide variety of models, including supervised and unsupervised models. A comprehensive overview of neural networks and other soft computing techniques (namely fuzzy logic and genetic algorithms) in various oil and gas applications was provided by Aminzadeh and DeGroot (2006). More recently, a review of the commonly used data analytics models and their application in the oil and gas industry was presented by Mishra (2017). Usually, these studies are focused on predicting the production (6 months or 12 months). The following are a few examples of using the data-driven approach in the oil and gas industry. Shelley et al. (2012) constructed an artificial neural network model to forecast production. Gupta (2014) used neural network (NN) and time series analysis techniques to predict shale gas wells' performance in unconventional reservoirs. In their method, they used production data of the previous year as the inputs of the model. We expect the next generation of neural networks and other artificial intelligence tools utilization in exploration and production in general, including those in shale resources, will involve integrating human and machine intelligence. This will help address the "SURE" challenge involving integrating data with many different Scales, Uncertainty, Resolution, and Environment (Aminzadeh, 2021).

Nejad et al. (2015) used an ANN model to optimize the well-completion strategy of three operators in Eagle Ford Shale. Zhong et al. (2015) compared the performance of several methods, including Ordinary Least Square Regression (OLSR), Support Vector Regression (SVR), Random Forest (RF), and Gradient Boosting Model (GBM), on the Wolfcamp shale data set. They concluded that the Random Forest performs better in terms of prediction. Bhattacharya et al. (2016) used various supervised and unsupervised techniques on Bakken and Marcellus data sets to identify the geological trends. They found that the support vector machine (SVM) works better for lithofacies classification. Katz et al. (2016) presented a rock permeability forecast that used machine learning and Monte-Carlo committee machines to predict rock permeability.

Mohaghegh (2016) presented a workflow for selection of refrac candidate selection using data analytics. A so-called "Shale Analytics" for using a data analysis approach for shale reservoirs was introduced later by the same author (Mohaghegh et al., 2017). Repchuk et al. (2018) developed a decision forest regression model to forecast production in the Denver Basin. Cai et al. (2018) proposed a nonparametric smoothing model to analyze the well performance in Bakken shale. Luo et al. (2018) built a predictive ANN model based on geologic inputs to forecast the production from Middle Bakken shale. Siddiqui et al. (2019) constructed several ANN models to predict different parameters such as oil API, most probable type of fluid, and first-year oil and gas production prediction from Eagle Ford Shale. Also, see Aminzadeh (2019) for many other utilization of neural networks and other machine learning techniques to improve the performance of different hydraulic fracturing operations.

In this study, we present several machine learning models to address some challenging questions regarding the extent of the stimulated reservoir volume (SRV) in the Marcellus Shale Energy and Environment Laboratory (MSEEL). We aim at using the operational inputs that can be obtained before or during the wellbore stimulation job. The models presented in this study include KNN, AdaBoost, Random Forest, ANN, and a Stack model to predict the value of the SRV as a single scalar during and after the stimulation job. The models are constructed using the data from two wells in MSEEL. The following section provides a summary of the MSEEL project. After that, we describe the pre-processing and the steps we took to make the data ready for analysis. This step is the most time-consuming part of the study. Then, in the Results section, we present the result of the five ML models. Also, the Discussion section contains a discussion on the parameters affecting the models' performance and an approach toward predicting the evolution direction of SRV. Finally, the conclusions are summarized in the last section.

Theory and Methods

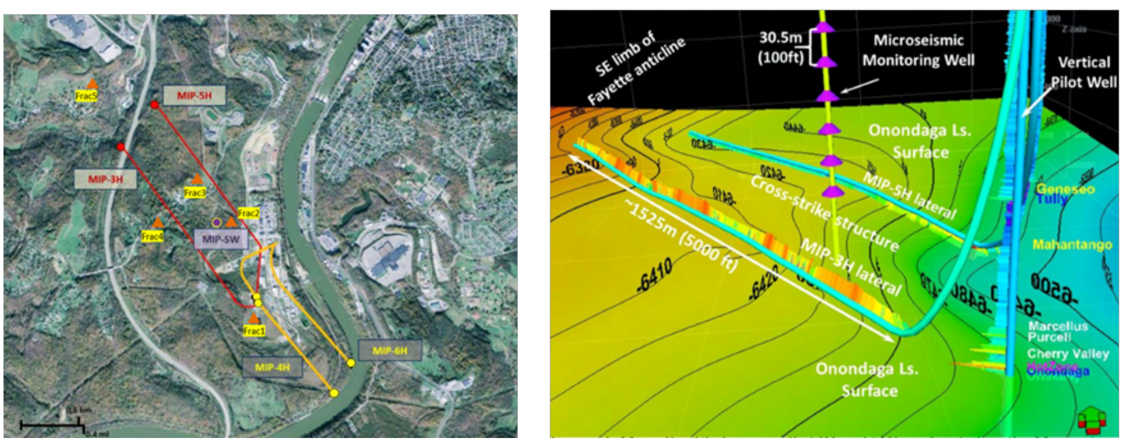
Hydraulic fracturing process typically is executed stage by stage from the toe of the horizontal to the heel, one stage at the time. At each stage, a certain amount of fracturing fluid is injected into the wellbore to break the rock and create new surfaces in the reservoir that can help increase the conductivity ($k \times w$) and hydrocarbon flow toward the wellbore from the stimulated region. As a result of this injection and depending on several variables such as the amount of injected fluid, local stress in the rock, and natural fracture density and orientation of the natural fractures in the pay zone, different "stimulated reservoir volume (SRV)" sizes are generated. The SRV extent is usually tracked using three different methods, namely direct far-field techniques, direct near-wellbore techniques, and indirect fracture diagnostics (Cipolla and Wright, 2000). The direct far-field techniques are suited to give the global visual perspective about fracture growth and are conducted in a separate well. The two commonly used techniques for this type of fracture diagnostics are tiltmeter fracture mapping and microseismic fracture mapping.

On the contrary, the second group of fracture diagnostics techniques is implemented in the same well and gives information about near-wellbore fracture parameters such as height and width and proppant placement. Finally, the most commonly used fracture diagnostics group is indirect fracture diagnostics. The techniques used in this group can provide estimates about fracture conductivity, dimension, and stress. This group includes fracture design model optimization, pressure transient testing, and production data analysis (Cipolla and wright, 2000).

The multi-stage hydraulic fracturing process simulation is a tedious task and requires high fidelity models for many reasons. However, with the abundant data that is being generated in the last decade, one may use a data-driven approach to estimate SRV and its evolution direction. In this study, we focused on the third method discussed above to estimate the extent of SRV in quasi-real-time. We use the available data that can be collected while or before performing hydraulic fracturing as our input. These variables include fracturing fluid type, rate, treating pressure, and proppant properties. On the output side, a volume enclosing the microseismic events (MSE) from hydraulic fracturing processes is estimated and used as the output variable.

Used Dataset

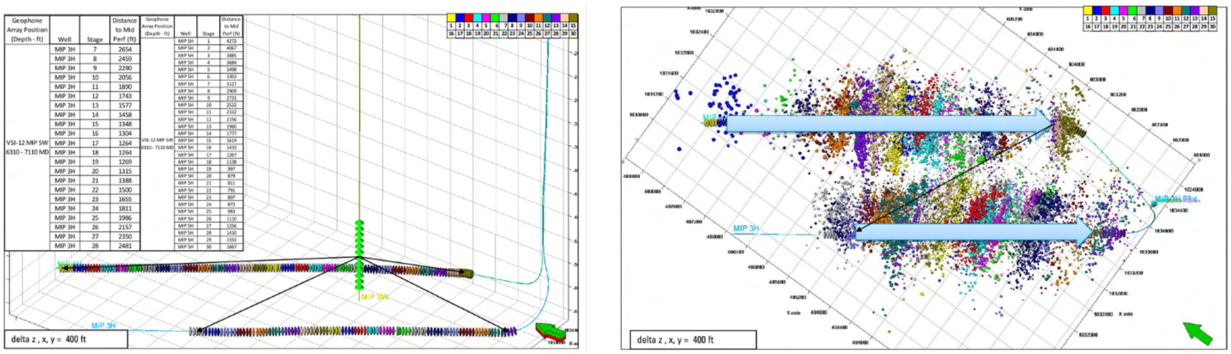
The dataset used in this study is from Marcellus Shale Energy and Environment Laboratory (MSEEL), an unconventional gas reservoir in North East of the U.S. The objective of the MSEEL project is to provide a long-term field site to develop and validate new knowledge and technology to improve recovery efficiency and minimize environmental implications of unconventional resource development. The project involves several universities, companies, and U.S research labs for evaluations in geology, geomechanics, completions, production, and completions areas. Figure 1 shows the subsurface and surface information of MSEEL. The project is located in Morgantown, WV. It consists of four horizontal wells, one vertical microseismic monitoring well, and five surface seismic locations. Out of the four horizontal wells in the site, we had access to fracturing and microseismic monitoring of wells MIP-3H and MIP-5H. Therefore, the data from these two wells were used for training and testing the ML models.



(a) (b)

Figure 1. Surface and subsurface information about the MSEEL project. (a) the surface location of the site. The field is located next to the Monongalia river in Morgantown, WV. It consists of four horizontal wells, one microseismic monitoring well (the dot between wells MIP-3H and MIP-5H), and five surface seismic stations (yellow dots). (b) Location and direction of the two wells used in this study. Figure adapted from Carr et al. (2019).

The two wells are parallel to each other, and the monitoring well was drilled in the spacing between the wells, as shown in Figure 2. Well MIP-3H was stimulated by 28 stages, out of which 22 stages were monitored using microseismic monitoring. Also, well MIP-5H was fractured using 30 stages, and poor MS monitoring was available at stages greater than 22. Figure 2b shows the location of the MSE, colored by stage.



(a) (b)

Figure 2. Schematic of the stages of the two MSEEL wells that were investigated in this study. (a) stages of the two wells (b) microseismic monitoring of the stages, color-coded by the stage number.

Pre-Processing

The pre-processing stage included converting the MSEEL reports from pdf format to MS Excel file and combining them as a single file. The variables include stage, step, step name, slurry volume (bbl), pump rate, pump time, cumulative pump time (new variable), fluid name, ramp up fluid, propp name, propp concentration, propp mass, average treating pressure, maximum treating pressure, and minimum treating pressure. Table 1 shows such a table and its variables for stage 5 of MIP-3H. In the table, the concept of steps is shown. Note that each step has unique properties. This approach helped us to: (1) increase the sample number to ~ 800, (2) apply a model that can predict the SRV volume in near real-time. Also, in order to relate the input variables at each step, a new variable called cumulative pump time is created in the table. This new variable will be used later to correlate the observed MS events to the pump schedule.

Table 1. Input table generated from MSEEL stimulation reports.

Well	Stage	Step	Step Name	Slurry Volume (bbl)	Pump Rate (bbl/min)	Pump Time (min)	Pump Time Cum	SQRT (sec^0.5)	Fluid Name	Ramp Fluid Volume (gal)	Prop Name	Prop Conc (PPA)	Prop Mass (lb)	Average Treating Pressure (psi)	Maximum Treating Pressure (psi)	Minimum Treating Pressure (psi)
MIP-3H	8	1	Rate	20.0	15.0	1.30	1.30	8.83	1	840	0	0.00	0	5470	5833	4303
MIP-3H	8	2	Acid	71.4	15.0	4.80	6.10	19.13	2	2999	0	0.00	0	6028	6106	5839
MIP-3H	8	3	PAD	595.2	80.0	7.40	13.50	28.46	1	25000	0	0.00	0	7996	9143	6028
MIP-3H	8	4	0.25 PPA	529.8	80.0	6.60	20.10	34.73	1	22000	100	0.20	5500	8991	9089	8900
MIP-3H	8	5	0.5 PPA	803.5	80.0	10.00	30.10	42.50	1	33000	100	0.50	16500	8877	8902	8855
MIP-3H	8	6	0.75 PPA	902.7	80.0	11.30	41.40	49.84	1	36667	100	0.70	27500	8913	8948	8893
MIP-3H	8	7	1.00 PPA	1149.9	80.0	14.40	55.80	57.86	1	46200	100	1.00	46200	8951	8988	8915
MIP-3H	8	8	1.5 PPA	1025.7	80.0	12.80	68.60	64.16	1	40333	100	1.50	60499	9058	9150	8944
MIP-3H	8	9	1.75 PPA	1156.1	80.0	14.50	83.10	70.61	1	44982	100	1.80	78718	9153	9197	9124
MIP-3H	8	10	2.0 PPA	748.3	80.0	9.40	92.50	74.50	1	28812	100	2.00	57624	9066	9200	8361
MIP-3H	8	11	2.5 PPA	397.5	80.0	5.00	97.50	76.49	1	14994	100	2.50	37485	9127	9195	9059
MIP-3H	8	12	.50 PPA	179	80	2.20	99.70	77.34	1	7350	40	0.50	3675	9139	9190	9096
MIP-3H	8	13	.75 PPA	271.5	80.0	3.40	103.10	78.65	1	11004	40	0.80	8803	9040	9219	8889
MIP-3H	8	14	1.0 PPA	373.3	80.0	4.70	107.80	80.42	1	14994	40	1.00	14994	9129	9188	9004
MIP-3H	8	15	1.50 PPA	381.4	80.0	4.80	112.60	82.19	1	14994	40	1.50	22491	9171	9203	9135
MIP-3H	8	16	2.00 PPA	508.5	80.0	6.40	119.00	84.50	1	19572	40	2.00	39144	9123	9182	9047
MIP-3H	8	17	2.5 PPA	132.6	80.0	1.70	120.70	85.10	1	4998	40	2.50	12495	8937	9074	8777
MIP-3H	8	18	3.0 PPA	81.2	80.0	1.00	121.70	85.45	1	2999	40	3.00	8996	8904	9037	8770
MIP-3H	8	19	Flush	270.0	80.0	3.40	125.10	86.64	1	11340	0	0.00	0	9059	9167	6093

Table 2 shows the generated output table from the recorded microseismic events. The data contained the x, y, and z location of the events and their magnitude. However, the magnitudes were excluded from this study and will be used later for verifying the calculated SRV. Depending on the cumulative time from the input table, we created similar columns in the output table. Next, a wrapper file was created to loop through the stage, step, and cumulative time and find the corresponding MS events for that specific step.

Table 2. The generated table for microseismic.

Well	Stage	Step	Time	Time difference	Cummulative time	XLoc	YLoc	ZLoc
MIP-5H	2	1	10:30:10	0:00:00	0:00:00	1831598.25	407735.31	-5986.18
MIP-5H	2	1	10:30:34	0:00:24	0:00:24	1831645.44	407753.94	-5968.90
MIP-5H	2	1	10:30:41	0:00:07	0:00:31	1831692.37	407653.12	-6271.44
MIP-5H	2	2	10:32:38	0:01:57	0:02:28	1831319.54	407590.36	-5865.04
MIP-5H	2	3	10:37:08	0:04:30	0:06:58	1831496.50	407606.55	-5840.98
MIP-5H	2	3	10:44:45	0:07:37	0:14:35	1831301.72	407706.51	-6135.57
MIP-5H	2	7	11:01:14	0:16:29	0:31:04	1831879.96	407641.13	-5475.60
MIP-5H	2	9	11:11:24	0:10:10	0:41:14	1831688.69	407965.96	-6320.97
MIP-5H	2	9	11:11:53	0:00:29	0:41:43	1831600.45	408043.02	-6503.97
MIP-5H	2	9	11:14:01	0:02:08	0:43:51	1831564.41	407803.99	-6201.43

The result of this approach is a group of distinguished microseismic events for each step. Figure 3 shows an example of such grouping for MIP-3H stage 7. Note that the MS events are color-coded for each step (note: the steps are smaller time periods inside each stage). Also, note that two parameters are changing from a step to another step. These parameters are the extent of SRV and its intensity (i.e., the density of the MS events in the enclosed geometry). In this study, we only focus on estimating the extent of the SRV. However, it is critical to consider the intensity of the changes in the SRV. Another useful piece of information that can be obtained from the MS points in the plots is the direction of the SRV growth. For example, the direction of SRV growth is NW-SE and NE-SW in the 2D plots shown below. This can be further implemented to track the evolution of the SRV in near real-time.

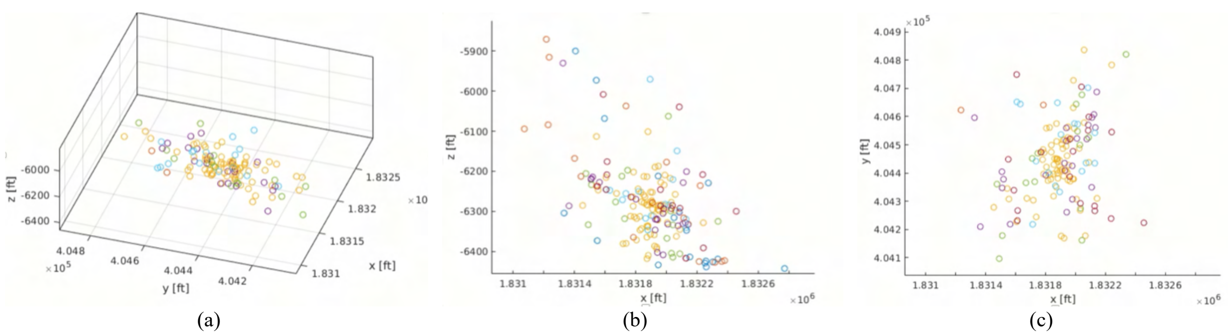
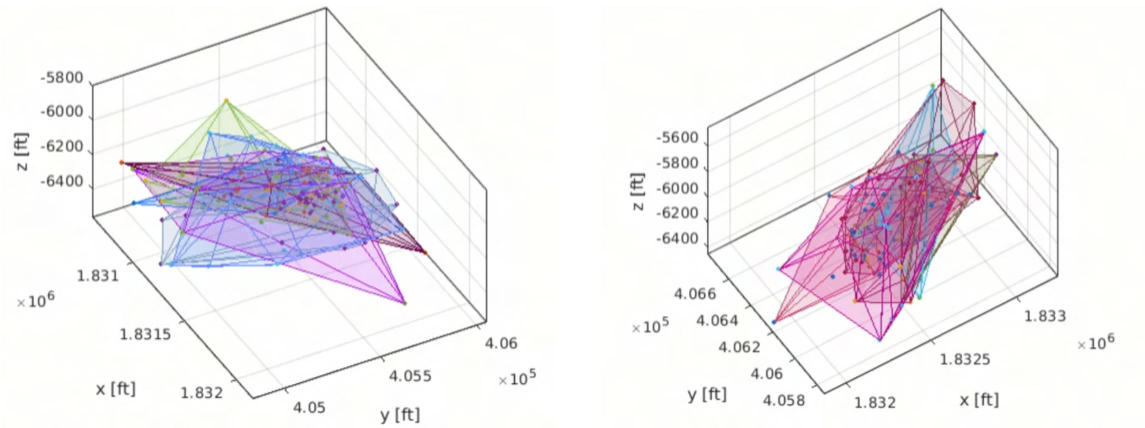


Figure 3. Example of MS event observed in MIP-3H stage 7. The microseismic events are color-coded for each step (a time period with a unique fluid, ump and proppant properties inside a specific stage. Each stage could be divided into several steps) discussed above. In the picture, 3D and two 2D snapshots of the MS events are shown.

The last step toward creating the output (i.e., DSRV) is to estimate a volume that encloses the MS events. For this purpose, a threshold may be devised that eliminated the isolated MS events and does not seem to contribute to the overall SRV estimations. In this study, we enclosed all points in an irregular geometry, and later in the cleaning data section, the samples with volumes that are much bigger than a certain threshold are removed from the dataset. Figure 4 shows two examples of the estimated SRV for MIP-3H stage 7 and MIP-5H stage 12. The volumes that are calculated as SRV are color-coded for the steps.



(a) (b)

Figure 4. Example of the calculated SRV from MS events for two stages of MSEEL. (a) MIP-3H stage 7 (b) MIP-5H stage 12.

Once the SRVs (or DSR) are known for each step, they were added back to the input table (Table 1) as a new column and used as our target variables. The next step toward building the ML models is to perform some exploratory data analysis to find the relationship between the input parameters and their relationship between the output parameters. This can be done in several ways. An example of such plots is shown in Figure 5.We started with all the15 variables that were available in the input table. As in the confusion matrix (Figure 5), several relationships were identified between the input samples. For example, average, max, min pressures have similar effects on the SRV, as highlighted with green boxes in the figure. Also, some of the variables seem to have linear correlation (see, for example, ramp-up fluid volume and slurry volume). Moreover, some of the inputs resulted in multiple SRVs (e.g., pump rate) and could negatively impact the model performance. Finally, some of the variables have outliers that needed to be removed. An example of this case is the estimated SRV, as explained previously. Some of the SRV values were considerably bigger than the others, indicating the non-productive MS events.

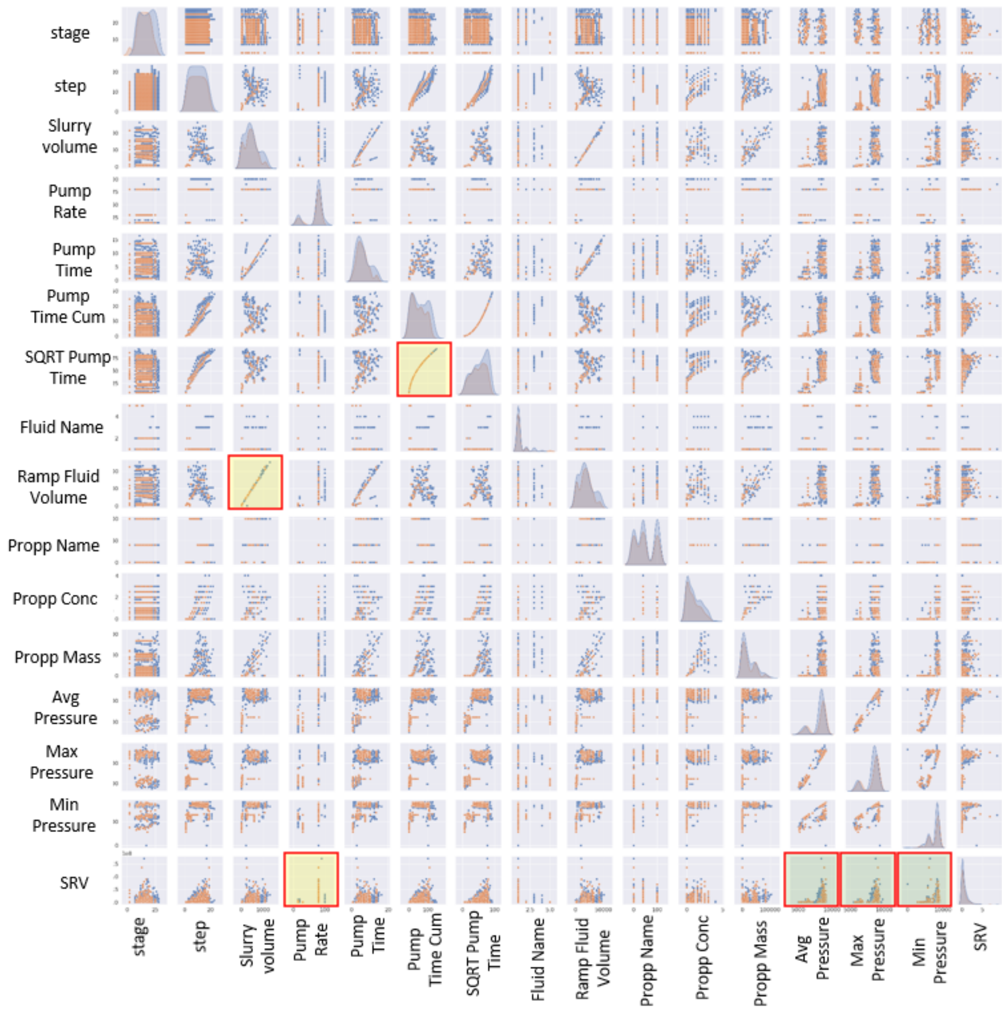


Figure 5. Confusion matrix of all the initial variables. The variables are color-coded for different wells. The variables of MIP-3H are shown by blue colors, and variables of MIP-5H are shown by orange color.

As a result, some of the mentioned columns were removed from the input features, and some were corrected by removing outliers to avoid any model bias. Finally, eight parameters were selected from the analysis done in the previous step. The final input/output table is shown in Table 3. The parameter "well name" is selected metadata. The input columns that were used to create the models were a stage, step, slurry volume, pump time, pump time cumulative, propp concentration, propp mass, and average treating pressure. Note that some of the SRVs in the output column are zero, indicating no creation of SRV. Some of these rows with zero SRVwere also removed from the data, especially if they were located in the middle of the stage, where the pump rate and prop concentrations were maximum (it was assumed that the high injection rate should result in some SRV creation, especially if the steps before or after created SRV). The zero SRV can be related to an error in MS monitoring or not creating an SRV at all (smaller chance).

Table 3. The final set of input/output.

	Well	Stage	Step	Slurry\nVolume\n(bbl)	Pump\nTime\n(min)	Pump Time Cum	Prop Conc (PPA)	Prop\nMass\n(lb)	Average Treating Pressure (psi)	SRV
0	MIP-3H	7	1	20.0	1.3	1.3	0.0	0	5640.0	0.000000e+00
1	MIP-3H	7	2	71.4	4.8	6.1	0.0	0	6212.0	0.000000e+00
2	MIP-3H	7	3	595.2	7.4	13.5	0.0	0	7549.0	0.000000e+00
3	MIP-3H	7	4	529.8	6.6	20.1	0.2	5500	8249.0	7.118735e+06
4	MIP-3H	7	5	803.5	10.0	30.1	0.5	16500	8247.0	3.704643e+07

After all the changes made on the initial data, a total of 582 samples were remained, from which 482 were selected for training and 100 samples were excluded for testing the performance of the models. It should also be noted that the data was shuffled and normalized to [0 1] interval before creating the ML models.
A heat map of the selected input variables is shown in Figure 6. As can be seen, the parameters that affect the output parameters the most are slurry volume, pump time, prop mass, and the least affecting parameters are stage and prop concentration. Also, the relationship between pump time and slurry volume has strong correlation weights in modeling, as shown in the figure.

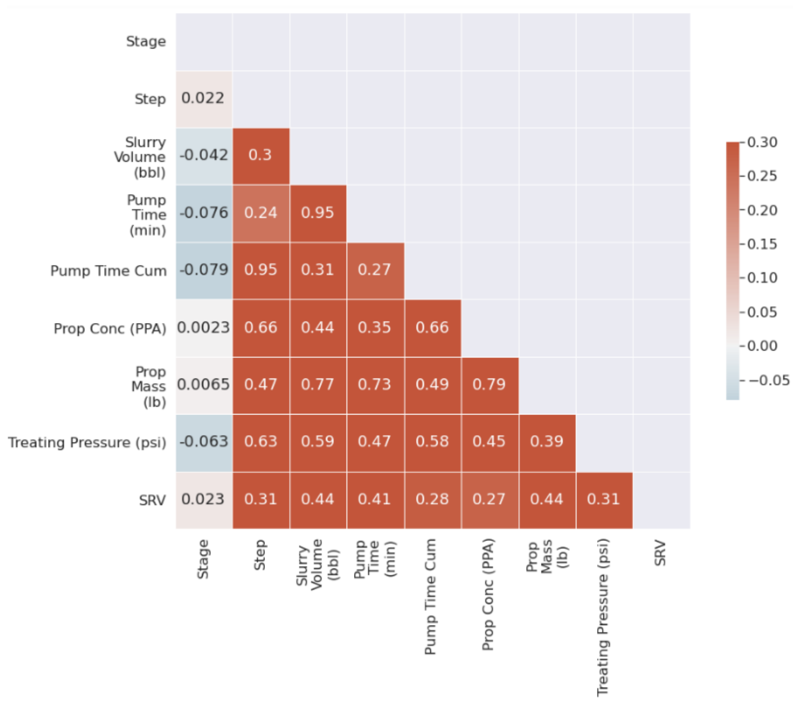


Figure 6. Heat map of the final variables after removing the problematic parameters.

Results

Several machine learning models were trained and tested on the data. The models were AdaBoost, KNN, Random Forest, ANN, and a stack algorithm. The stack algorithm included Random Forest, KNN, AdaBoost models as input learners, and Ridge Regression as the output learner. The created models' performances on the training set, including 482 samples, are shown in Figure 7. The x-axis on the plots shows the predicted values of the SRV, and the y-axis shows the original value from the test set. Also, the line on the plots shows the R-Square of each model. Moreover, the points are color-coded based on the well number. As can be seen, the r-square for AdaBoost, KNN, and stack is 1 for the train set, while Random Forest and ANN have r-squares of 0.96 and 0.85 (overall). All models perform slightly better on the well MIP-3H compared with MIP-5H. The reason for this is that the number of data samples in MIP-5H was smaller, and also there was an error for several points in the middle stages. More discussions will be provided on this point in the discussion section.

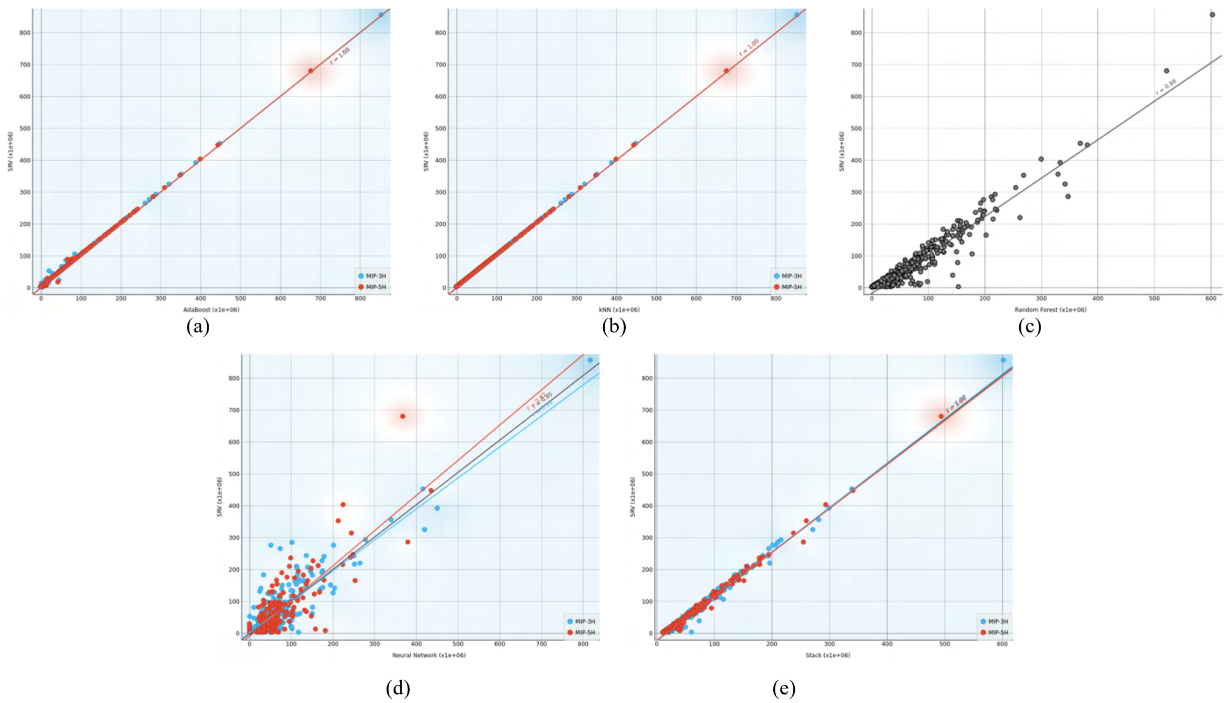


Figure 7. The model performances on the training data. (a) AdaBoost (b) KNN (c) Random Forest (d) ANN (e) Stack. The x-axis in the plots shows the predicted values by the models, and the y-axis represents the actual values.

The test data included 100 data samples from both wells. Figure 8 shows the model performance of the models on the test set. As expected, the performance is lower on the test model. Again, the results are color-coded for the well number and show that the model performance is better in general for MIP-3H. The overall r-square for test sets are 0.69, 0.60, 0.68, 0.72, and 0.64 for AdaBoost, KNN, Random Forest, and Stack models, respectively. As can be seen, among all models, ANN had a better performance on the test set with 0.72 r-square. The performance of the model may be improved by adding more training samples and details about the target formation properties. For example, having an idea about the distribution, density, and orientation of natural fractures may be helpful for predicting SRV. It should be noted that one reason for lower overall r-squares is the lower performance on well MIP-5H, which is caused by the poor prediction of six points. The points are highlighted in the figure with a yellow shaded area. The reason for this weak performance could be an error in data collection or weak engineering or execution of the fracturing job in that stage. In order to further investigate the reason behind models' weak performances on these six points, we performed another investigation on the mentioned points, and the results are discussed next. Note that other metrics may be used to evaluate the performances of the models, but they have been skipped in this study.

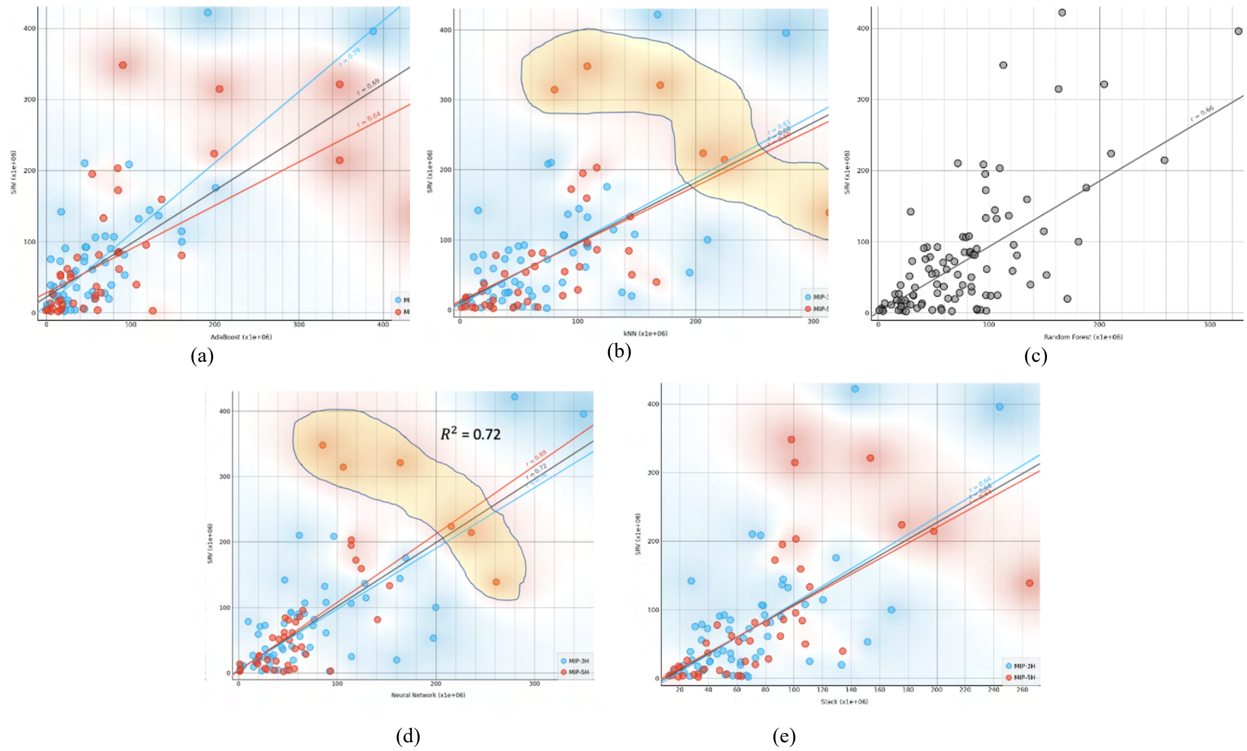


Figure 8. Models' performances, (a) AdaBoost (b) KNN (c) Random Forest (d) ANN (e) Stack. The x-axis in the plots shows the predicted values by the models, and the y-axis represents the actual values.

To further investigate the reason for the weak performance of the models, the results are plotted as a function of three variables, namely stage number, step, and proppant concentration. For this purpose, we arbitrarily selected the AdaBoost model. However, the six points were observed in other models as well. Figure 9 shows the results for the three selected variables. In this presentation, the size of each bubble represents the magnitude of the selected variable. For example, a bigger circle means that the stage number is bigger (i.e., closer to the heel). Figure 9a shows the results for stage number. As can be seen, the miss-predicted points are all from the stage in the middle of the well. It can also be concluded that the steps in five of the points are toward the end of the fracturing job for that specific stage, where the pump rates and proppant concentrations were highest. A look at Figure 9 confirms this observation. In that figure, five of the miss-predicted points were from high proppant concentrations. More investigations may be done to further analyze the reason behind the poor model performance in these six points.

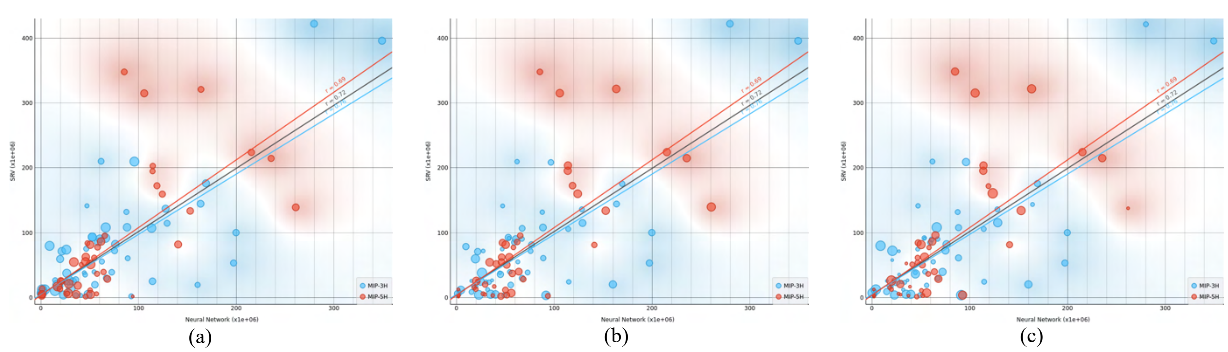
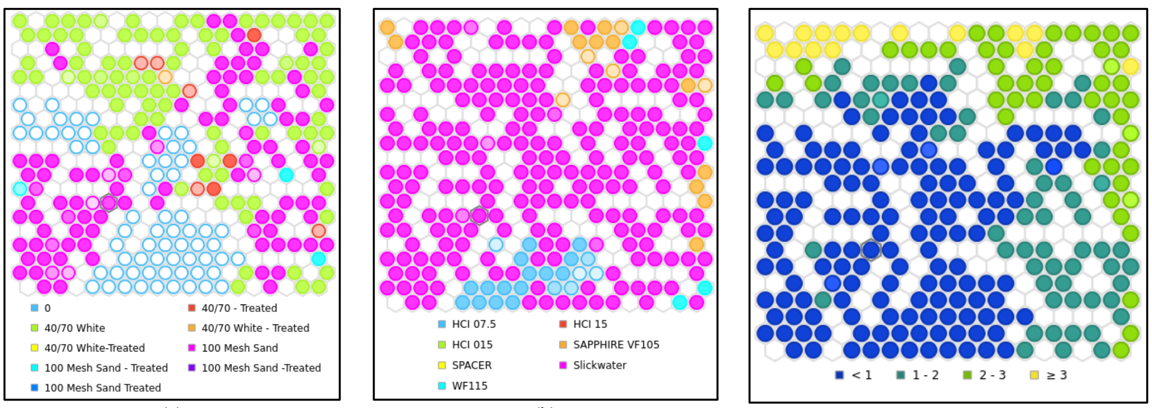


Figure 9. Further analysis of the models' weak performances on well MIP-5H. (a) circle size: stage (b) circle size: step (c) circle size: proppant concentration. The x-axis in the plots shows the predicted values by the models, and the y-axis represents the actual values.

Discussion

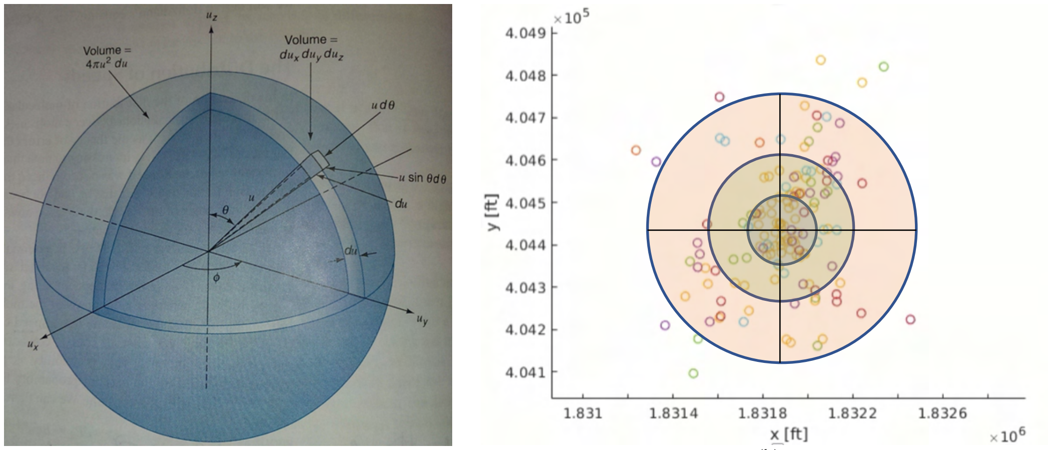
Two types of variables control the created SRV: namely, operational parameters and rock geomechanical and mineralogical variables. Operational parameters include parameters such as the pump schedule (e.g., duration, rate), recorded pressure, and perforation design. Rock-related parameters include rock mineralogy, stress state, natural fractures' properties such as density and orientation relative to the wellbore direction, and layer height in the pay zone, and so forth. In this study, we only limited ourselves to the operational parameters because of limited data availability. Including the rock-related parameters will improve the performance of the models. Also, the number of examples that we worked with was relatively small (580), and it is expected that adding more examples will improve the performance. Another approach that one may take to improve the performance of these models is to apply some unsupervised learning algorithms to the data. As an example, Figure 10 shows the result of the self-organizing map on different properties. The algorithm can be used to cluster the input data based on some variables. In the figure, the algorithm is applied to mesh size, fluid type, and proppant concentration. As can be seen, for example, in Figure 10b, the samples with WF115 seem to have similar behavior. Also, when the proppant concentration is less than one bb/gal (Figure 10c), the points show similar behavior. This clustering helps to further decrease the model errors.



(a) (b) (c)

Figure 10. Clustering the data using a self-organizing map (SOM), unsupervised learning algorithm. (a) proppant size (b) fluid type (c) proppant concentration.

Another parameter that was ignored in this study was MS events magnitude. The magnitude of the events plays a role in the intensity of permeability changes in the stimulated zone. In future studies, this property will be included in our models. Also, one important aspect of SRV evolution in the direction of the evolution. Obviously, the fractures will not always propagate in a symmetrical fashion. The fracture propagation direction can be affected as a result of stress shadowing from previous steps or layer confinement in height. Another idea to further address the evolution direction of SRV is the quadrant idea. In this approach, a sphere (in 3D) is divided into multiple equal size quadrants. Suppose the frac fluid is injected at the point center of a sphere, and the reservoir rock is uniformly diffusive to fluids expanding radially. In that case, the sphere can be a proxy shape for isotropic, homogenous reservoir rock. However, this is not always the case. Figure 11 shows the quadrant idea. In this approach, one needs to start with a small circle centered at the wellbore (Figure 11b). At each step, a new circle is drawn from the same center with a bigger radius. By tracking the number of points in each quadrant, one may estimate the evolution direction of SRV. For example, more points are located in the three quadrants of NE, NE, and SW. Therefore it may be concluded that the SRV is blocked from propagating in the NW quadrant. A look at the design and location of the weels with geological data will be a helpful addition to this analysis to figure out the reason.



(a) (b)

Figure 11. An algorithm for tracking the growth direction of SRV.

Conclusions

A data-driven approach was used to dynamically predict the stimulated reservoir volume (SRV) in this study. The SRV in this approach is called DSRV and can be used for real-time analysis of the SRV growth. In this approach, each hydraulic fracturing stage has been broken down into several steps, each having unique properties such as pump rate, proppant concentration, and average treating pressure. These properties were used in the input feature table. A data set from the MSEEL project, including 58 stages from two well (MIP-3H and 5H) were used. On the output side, we automatically created irregular geometrical shapes around each of the recorded microseismic clouds at each step and represented that as SRV. Then the time of each step was correlated to the SRV volume to create the required table for the model creation. Several machine learning models, including AdaBoost, KNN, ANN, Random Forest, and a stack algorithm, were implemented on the data. The models had an R-Square metric of [0.65 0.78] on the test data, which is an acceptable range for such a limited dataset. The performance of the models can be improved by adding more training examples (we only used ~500 samples). Models showed a weaker performance in the middle stages of MIP-5H that negatively impacted the overall performances. We proposed two approaches to estimate the evolution of SRV: volume prediction and quadrant. The volume prediction in real-time was presented in this study. In future models, we include the quadrant idea to estimate the growth direction of SRV. Model performances can be improved by adding more data samples and more information about the target formation. Also, in order to have a better estimation of SRV, magnitudes of the MS events will be included.

Acknowledgment

This work was completed as part of the **Science-informed Machine learning to Accelerate Real Time decision making for Oil & Gas (SMART-OG) Initiative** (edx.netl.doe.gov/SMART). Support for this initiative was provided by the U.S. Department of Energy’s (DOE) Office of Fossil Energy’s Oil and Natural Gas program through the National Energy Technology Laboratory (NETL). The authors wish to acknowledge Jared Ciferno (NETL, Acting Onshore Oil and Gas Technology Manager), Sailendra Mahapatra (DOE Office of Fossil Energy, Program Manager for Oil and Gas Upstream Research and Machine Learning/Deep Learning/Artificial Intelligence Programs, and Elena Melchert (DOE Office of Fossil Energy, Director, Upstream Oil & Gas Research Division) for programmatic guidance, direction and support. Partial support was provided by the sponsors of **Artificial Intelligence, Machine Learning and**

Data Analytics for Energy Exploration and Production (AIM-DEEP) of the University of Houston: https://aim-deep.petro.uh.edu/

References

Adachi, J., Siebrits, E. M., Peirce, A., & Desroches, J. (2007). Computer simulation of hydraulic fractures. International Journal of Rock Mechanics and Mining Sciences, 44(5), 739-757.

Aminzadeh, F. (2021). Reservoir Characterization: Combining Machine Intelligence with Human Intelligence, E&P Plus, April 2021, Vol. 96 Issue 4, E&P Plus, Hart Energy,

Aminzadeh, F. (2019). Hydraulic Fracturing and Well Stimulation: 330 pages, Wiley & Sons Ltd, ISBN: 978-1-119-55569-

Aminzadeh, F., and de Groot, P. (2006). Neural Networks and Soft Computing Techniques, with applications in the oil industry EAGE Publishing.

Bhattacharya, S., Carr, T.R., Pal, M. (2016). Comparison of supervised and unsupervised approaches for mudstone lithofacies classi_cation: Case studies from the bakken and mahantango-marcellus shale, usa. Journal of Natural Gas Science and Engineering;33:1119-1133.

Carr, T., Ghahfarokhi, P. K., Carney, B. J., Hewitt, J., & Vargnetti, R. (2019). Marcellus Shale Energy and Environmental Laboratory (MSEEL) Results and Plans: Improved Subsurface Reservoir Characterization and Engineered Completions. In Unconventional Resources Technology Conference, Denver, Colorado, 22-24 July 2019 (pp. 215-224). Unconventional Resources Technology Conference (URTeC); Society of Exploration Geophysicists.

Cai, Q., Yu, W., Liang, H.C., Liang, J.T., Wang, S., Wu, K., et al. (2018). Development of a powerful data-analysis tool using nonparametric smoothing models to identify drillsites in tight shale reservoirs with high economic potential. SPE Journal 2018;23(03):719-736.

Cipolla, C.L., and C.A. Wright (2000). "State-of-the-Art in Hydraulic Fracture Diagnostics." Paper presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Brisbane, Australia. doi: https://doi.org/10.2118/64434-MS

Dindoruk, B. (2012). Development of a Correlation for the Estimation of Condensate to Gas Ratio (CGR) and Other Key Gas Properties from Density Data. Society of Petroleum Engineers. doi:10.2118/160170-MS

EIA (2016). Trends in us oil and natural gas upstream costs. US Energy Information Administration.

Gupta, S. a. (2014). Production forecasting in unconventional resources using data mining and time series analysis. SPE/CSUR Unconventional Resources Conference. Canada: Society of Petroleum Engineers.

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S. (2017). Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: Advances in Neural Information Processing Systems. p. 6626-6637.

Katz, S., Aminzadeh, F., Chilingar, G., Long, W. Khylyuk, L. , and Lackpour, M. (2016). Rock Permeability Forecast Using Machine Learning and Monte Carlo Committee Machines, Journal of Sustainable Energy Engineering, Vol. 4, Issue 2, pp 182-200.

Kellogg, R., Chessum, W., Kwong, R. (2018). Machine learning application for wellbore damage removal in the wilmington _eld. In: SPE Western Regional Meeting. Society of Petroleum Engineers.

Kingma, DP, Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980;.

Lecampion, B., Bunger, A., & Zhang, X. (2018). Numerical methods for hydraulic fracture propagation: a review of recent trends. Journal of natural gas science and engineering, 49, 66-83.

Luo, G., Tian, Y., Bychina, M. and Ehlig-Economides, C. (2018). Production Optimization Using Machine Learning in Bakken Shale. In Unconventional Resources Technology Conference, Houston, Texas, 23-25(pp. 2174-2197). Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers.

Mishra, S., Lin, L. (2017). Application of data analytics for production optimization in unconventional reservoirs: A critical review. In: Unconventional Resources Technology Conference, Austin, Texas, 24-26 July 2017. Society of Exploration Geophysicists, American Association of Petroleum Geologists, Society of Petroleum Engineers; p. 1060-1065.

Mohaghegh, S. D. (2016). Fact-Based Re-Frac Candidate Selection and Design in Shale-A Case Study in Application of Data Analytics. Unconventional Resources Technology Conference (pp. 514--527). San Antonio, USA: Society of Petroleum Engineers.

Mohaghegh, S., Gaskari, R., Maysami, M. (2017). Shale analytics: Making production and operational decisions based on facts: A case study in marcellus shale. In: SPE Hydraulic Fracturing Technology Conference and Exhibition. Society of Petroleum Engineers.

Mullen, J. (2010). Petrophysical characterization of the Eagle Ford Shale in south Texas. Canadian Unconventional Resources and International Petroleum Conference. Society of Petroleum Engineers.

Nejad, A.M., Sheludko, S., Shelley, R.F., Hodgson, T., Mcfall, PR (2015). A case history: evaluating well completions in eagle ford shale using a data-driven approach. In: SPE Hydraulic Fracturing Technology Conference. Society of Petroleum Engineers.

Repchuk, K., Reimchen, A., Gregoris, D. (2018). Using data analytics to maximize value in the denver-julesburg basin. In: Unconventional Resources Technology Conference, Houston, Texas, 23-25 July 2018. Society of Exploration Geophysicists; p. 2848-2853.

Rezaei, A., Dindoruk, B., & Soliman, M. Y. (2019). On parameters affecting the propagation of hydraulic fractures from infill wells. Journal of Petroleum Science and Engineering, 182, 106255.

Rezaei, A., Siddiqui, F., Dindoruk, B., & Soliman, M. Y. (2020). Utilizing a Global Sensitivity Analysis and Data Science to Identify Dominant Parameters Affecting the Production of Wells and Development of a Reduced Order Model for the Eagle Ford Shale. In Unconventional Resources Technology Conference, 20–22 July 2020 (pp. 4655-4673).

Schuetter, J., Mishra, S., Zhong, M. and LaFollette, R., (2018). A Data-Analytics Tutorial: Building Predictive Models for Oil Production in an Unconventional Shale Reservoir. SPE Journal, 23(04), pp.1-075.

Shelley, R.F., Guliyev, N. and Nejad, A., (2012). A Novel Method to Optimize Horizontal Bakken Completions in a Factory Mode Development Program. In SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

Siddiqui, F., Rezaei, A., Dindoruk, B., & Soliman, M. Y. (2019). Eagle ford fluid type variation and completion optimization: a case for data analytics. In Unconventional Resources Technology Conference, Denver, Colorado, 22-24 July 2019 (pp. 765-783). Unconventional Resources Technology Conference (URTeC).

Tijmen, T., & Hinton, G. (2012). Lecture 6.5-RMSProp, COURSERA: Neural networks for machine learning. University of Toronto: Technical Report.

Warpinski, N. (2009). Microseismic monitoring: Inside and out. Journal of Petroleum Technology, 61(11), 80-85.

Zhong, M., Schuetter, J., Mishra, S. and Lafollette, R.F., (2015). Do data mining methods matter?: A Wolfcamp Shale case study. In SPE Hydraulic Fracturing Technology Conference. Society of Petroleum Engineers.