URTeC: 3723843

Completion Optimization in the Bakken Petroleum System Using Data Mining

Alexander Chakhmakhchev*[1], Nicholas A. Azzolina[1], Bethany A. Kurz[1], Xue Yu[1], Chantsalmaa Dalkhaa[1], Justin T. Kovacevich[1], James A. Sorensen[1], 1. Energy & Environmental Research Center

This paper was prepared for presentation at the Unconventional Resources Technology Conference held in Houston, Texas, USA, 20-22 June 2022.

Abstract

Like oil producers in other unconventional plays, operators in the Bakken petroleum system (BPS) must reduce capital and operating costs by optimizing operations. This work summarizes the results from more than 12,000 producing wells in the BPS, which provided input data for optimization calculations. Straightforward interpretations of the relationship between production and completion parameters based on bivariate (two-dimensional) scatterplots were difficult because of the nonlinear nature of the dependencies between variables. Therefore, the primary goal of this study was to identify optimal completion practices using publicly available well completion and production information and applying data-mining techniques that could accommodate nonlinear relationships.

Optimization work was conducted using the data-mining tool, Gradient Boosting. The target or predicted variable was cumulative 6-month oil production, and the predictors included ten completion design parameters. To reduce the influence of geologic or reservoir heterogeneity on the results of the calculations, the optimization work was conducted on three groups of wells located in three subareas of the BPS representing low-, moderate-, and high-productivity regions, with approximately 300 wells in each group. The statistical modeling produced 1) variable (completion parameter) importance graphs and 2) one-variable dependence graphs, which were used to estimate optimal values of completion parameters that maximized 6-month production while minimizing the size of the stimulation job (e.g., volume of fluid or pounds of proppant).

Across all three subareas, the three most important features always included total proppant and total fluid, which supports other work that showed these features to be significantly related to oil production. The results suggested different optimal completion configurations for the three subareas. The high-productivity subarea benefitted from higher total proppant, slightly lower total fluid, and higher maximum treatment pressure, and the moderate- and low-productivity subareas maximized oil production with less total proppant, more total fluid, and lower maximum treatment pressure.

The differences in completion strategies among the three areas were attributed to observed heterogeneity of geologic and reservoir characteristics, including formation depth, temperature, pressure, maturity level, total organic carbon content, and thickness of the reservoirs. This innovative approach of reducing the impact of geologic variability by running calculations for smaller areas improved the statistical models (improved the goodness-of-fit) and strengthened the model interpretations. The completion optimization results can help oil and gas operators to tailor their completion designs in different subareas of the BPS, which could significantly reduce their costs and maximize oil production.

Introduction

The Bakken petroleum system (BPS) of the Williston Basin has been one of the most active and prolific unconventional oil plays in the United States. The Bakken Production Optimization Program (BPOP) is a research program led by the Energy & Environmental Research Center (EERC) and funded by the North Dakota Industrial Commission (NDIC) and industry partners, with the goal of simultaneously improving BPS oil recovery while reducing its environmental footprint. Since 2016, through BPOP, the EERC has been evaluating BPS oil production and well completion data using statistical and machine learning (ML) methods (Pekot and others, 2016; Dalkhaa and others, 2019; Chakhmakhchev and others, 2020). An expanded analysis conducted in 2020 using over 12,000 BPS wells and ML methods to predict well performance using completion design parameters demonstrated an overfitting problem on a basin scale (Chakhmakhchev and others, 2020). Stated differently, while the ML-based models performed well on the training data set and could accurately explain the variation in production from completion parameters, the models did not perform equally well on the test data set for a different set of wells that were not included in the model training and tuning, which was an indication of model overfitting. The overfitting suggested that variables not included in the model (e.g., geologic factors) were important and that further work incorporating geologic factors would reduce prediction errors.

Early Efforts at Completion Optimization in Unconventional Plays: Optimization of completion design based on benchmarking and well performance prediction is challenging in an unconventional resource play like the BPS. Over relatively small spatial scales, well production performance can vary dramatically, making simple evaluations based on bivariate analysis almost meaningless (Pearson and others, 2018). The multiple options available in completion design provide additional complexity to the challenging task of predicting well behavior. The use of multivariate models incorporating large data sets of geologic, well, and completion parameters provides a more robust solution and a data-driven, empirical approach to optimizing completion design.

Completion optimization studies for unconventional reservoirs began to emerge in the United States when drilling accelerated and the producing well count achieved several thousand units. For example, in 2012, the number of unconventional wells reported as Barnett producers was about 17,000 wells. By that time, thousands of wells in multiple unconventional plays including Eagle Ford, Barnett, Niobrara, the BPS, and others were stimulated, and the data became available for experimental data-mining modeling through public sources and commercial subscriptions.

In the early phases of unconventional development, oil and gas operators were dealing with unpredictable well behavior and testing different completion strategies, trying to maximize production at lower cost. Early data analysis approaches applied parametric statistical analysis to determine correlations between well performance and completion/engineering parameters to identify and possibly optimize operations. However, because production and completion data frequently demonstrate weak or nonlinear correlations and are further confounded by missing values (open source, public data), traditional parametric approaches have been less effective and innovative data-mining techniques based on ML principles have been increasingly used to evaluate and optimize well stimulation practices.

Huckabee and others (2010) provided one of the earliest published efforts applying ML techniques for unconventional completion optimization. Using data from the Pinedale Anticline field in western

Wyoming, producing tight gas, the authors evaluated completion design parameters (proppant mass, flowback method) and geologic/petrophysical parameters (reservoir characteristics, geology) to better understand their impact on gas well performance. In the initial phase of their work, Huckabee and others (2010) concluded that more traditional single-variable analysis is challenging because of uncertainty in well performance outcomes, complexity of subsurface variability, and nonsequential and coincident completion variable modifications. To address deficiencies of single-variable analysis, a neural network model was applied to identify "sweet spots" and develop fracture design optimization. The results of calculations showed that over 80% of production performance was controlled by subsurface geologic and petrophysical characteristics and only 20% of the variability in production performance could be explained by completion design parameters.

Swindell (2012) provided another example of applying ML to unconventional production data using completion design data from approximately 300 wells completed in the Eagle Ford Shale. The study suggested that 4 to 5 million pounds of proppant and 5000 ft of perforated lateral length would maximize estimated ultimate recovery (EUR) and that proppant volumes or perforated lateral lengths exceeding these thresholds resulted in diminishing well performance.

LaFollette and others (2012) applied boosted regression trees to analyze a large data set of 15,000 wells representing Barnett Shale gas production. The model used well location, architecture, and hydraulic fracturing details to predict 6-month cumulative production. After eliminating highly correlated variables, six input variables were identified as the most important features to include in the model, which included well location (X- and Y-coordinates) and total vertical depths as a proxy for reservoir quality and the following operational parameters: volume of injected fluid, injection rate, 20/40 mesh proppant size, and perforated length. To visualize the influence of the parameters on production levels, the authors generated a series of graphs demonstrating the impact of individual variables on production. They concluded that, in general, larger fracture jobs resulted in improved well performance; however, the authors noted that the larger fracture treatments might create undesired communication with water-saturated formations, resulting in gas production loss (LaFollette and others, 2012).

The work of Griffin and others (2013) was focused on completion design optimization in the BPS and used nonlinear regression methods. The authors underscored the necessity of integrating geologic attributes into the model and the elimination of redundant completion parameters that demonstrated significant interdependency (correlation). The well performance in their work was characterized by 90- and 180-day cumulative production. An interesting outcome of their study was that cumulative water cut was the best indicator of reservoir quality and correlated with several geologic inputs.

One of the first studies that evaluated optimization in the Eagle Ford used a set of input variables representing completion and engineering parameters, well location, oil API (American Petroleum Institute) gravity, gas-to-oil ratio (GOR), and tubing pressure to predict well performance defined as peak oil daily production (Gao and Gao, 2013). Several important results came out of the study. First, the authors demonstrated that multivariate adaptive regression splines (MARS) could accommodate nonlinearity in the relationships between completion and production parameters and interactions between variables. Second, the importance of completion parameters for predicting well performance was shown to vary depending on the geographic location and that the variation possibly related to the Eagle Ford Shale thermal maturity or other geologic factors. Lastly, the lateral length of the well and proppant tonnage were shown to be less important to production than formation depth and tubing flowing pressure, which was an unexpected result and differed from past studies. A significant result of the study was utilization of one-variable dependence plots (partial dependence plots) for the purpose of completion optimization (Gao and Gao, 2013). Importantly, the work of Gao and Gao (2013) suggested the possibility of improving data-mining results using other statistical models based on decision trees such as gradient boosting. The effective performance of decision trees and gradient boosting was also documented by Friedman (2000, 2001), who showed that gradient boosting accommodates nonlinearity of data and provides similar tools for data mining such as quantifying the relative importance of input variables (feature importance graphs) and providing the ability to generate partial dependence plots for assessing completion optimization.

Well Completion Optimization Calculations in the BPS: Several relatively recent studies have been conducted in the BPS to identify the impact of completion parameters on production performance (Theloy and Sonnenberg, 2013). The work of Flowers and others (2014) used publicly available data NDIC and FracFocus to evaluate the impact of various proppant types on oil well performance. The authors limited the area of their research to a set of 205 wells from the Central Basin to minimize the differences in depth, permeability, reservoir pressure, temperature, and hydrocarbon content. They also divided the study area into subregions by using cluster analysis that incorporated the well location (X- and Y-coordinates) and formation depths. They concluded that ceramic proppants provided greater performance than other proppant types. Subsequent work by Lolon and others (2016) also limited the study area to the Central Basin (Rough Rider Area) and used publicly available well and completion data; however, the authors significantly expanded the data set to include 3061 wells and expanded the data analysis by using a suite of statistical and ML methods. The authors divided the data set into two subsets: 2283 wells producing from the Middle Bakken Formation and 778 wells producing from the Three Forks Formation. Thirteen completion parameters were used to predict 6-month cumulative production. Three data analysis methods were applied to the data to compare results among the different methods: multiple linear regression, gradient boosting, and random forest. Lolon and others (2016) made several important observations about the relationships between completion data and well performance. Evaluation of the different models showed that the three methods did not yield identical results and identified different sets of important parameters depending on the technique. The authors recommended that model selection should not be based solely on R-squared but should instead perform cross-validation. Like previous work by Griffin and others (2013), Lolon and others (2016) showed that water cut was the most influential predictor of oil production and attributed water cut as a proxy for reservoir quality. After controlling for water cut, the study showed that total fracture fluid and proppant pumped were the most important completion variables. In addition, the study posited that the feature importance graphs could be used for completion optimization purposes in the Bakken and Three Forks reservoirs (Lolon and others, 2016).

The work of Male and others (2018) is another example of applying advanced analytics to understanding the impact of completion design and well spacing and used tree-based ML approaches like those employed by Friedman (2000, 2001). Unlike other studies where the predicted variable was 6- or 12-month cumulative production, the work of Male and others (2018) used time to boundary-dominated flow and terminal decline rate to characterize well performance. Several sets of input variables were used in modeling runs, including completion design parameters, reservoir, fluid characteristics, water cut, depth, and completion dates. An important conclusion was that both increased drilling density (i.e., number of wells per drill spacing unit [DSU]) and hydraulic fracturing intensity resulted in a modest increase of initial production but also led to a higher terminal decline rate. The authors also concluded that EUR per lateral foot had been decreasing since 2009. The results and conclusions of Male and others (2018) provide another example of the potential drivers controlling hydrocarbon recovery and the value of both feature importance graphs and partial dependence plots in the analysis. In addition, the study also showed that more intense drilling and completion did not necessarily translate into higher EUR values.

The work of Luo and others (2018) was also focused on production optimization in the BPS. Like other optimization studies, their work used a mix of completion parameters and geologic variables such as structural depth, formation thickness, porosity, and water saturation to capture the geologic heterogeneity across the BPS. As in earlier studies, parametric methods like the Pearson correlation coefficients had close to zero values (i.e., not significant) between predictors and 12-month cumulative production, suggesting limitations of traditional methods for identifying important parameters. A neural network approach was used to predict 12-month cumulative production in 2000 wells drilled in 2013–2014. The top eight parameters with the highest ranking based on the neural network approach were 1) formation thickness from stratigraphic analysis, 2) normalized volume of proppant, 3) depth, 4) porosity, 5) stage count, 6) normalized volume of fluid, 7) normalized stage length, and 8) water saturation. The authors concluded that among the considered geologic factors, formation thickness and structural depth had the most significant impact on production, while volume of proppant and number of stages were the most significant completion parameters. They also suggested that at lower porosity, production increased more rapidly with more proppant injected, but at a higher porosity the impact of proppant mass was less pronounced. Similarly, the study showed that increased proppant mass improved production in locations where the reservoir had greater thickness (>40 ft) but did not increase production in locations where the reservoir was thinner (<29 ft) (Luo and others, 2018). These results highlight the interdependency of geologic factors and well completion parameters (Bhattacharya and others, 2019).

Scope of the Present Study: To account for the geologic heterogeneity across the BPS, the current work calculates optimized well completion parameters in three subareas characterized by similar well performance, geology, and geochemistry within each area. Well performance based on 6-month cumulative production and geographical location was used to create low-, moderate-, and high-productivity subareas within the BPS and select wells for further investigation. Optimization calculations were performed using statistical modeling in each area, and results of calculations were compared and explained in geological and geochemical terms.

Methodology

The master database used in this work contained well production and completion information for more than 12,000 producing wells in the BPS. The data were derived from NDIC and Enverus DI sources (Enverus 2020). The statistical modeling was performed using the commercially available Salford Predictive Model® package by Minitab Salford Systems and the decision tree-based algorithm, Gradient Boosting (SPM-GB). The principles of the SPM-GB engine have been extensively documented by Friedman (2000). The SPM-GB algorithm was previously tested using Eagle Ford and BPS well and production data sets (unpublished results) and demonstrated high tolerance to imperfect, incomplete data and outliers. The SPM-GB engine does not require data cleaning (preprocessing) and can accommodate nonlinearity of data. The output from the SPM-GB calculations are the relative importance of the input variables (feature importance diagram) and partial dependence plots that demonstrate the effect of each input feature on well production. Feature importance diagrams and partial dependence plots were used to estimate optimal completion parameters for three different production areas of the BPS. The results of optimization calculation for the low-, moderate-, and high-productivity subareas provide information that can be used by operators to optimize completion practices and significantly reduce completion costs.

Results

Identifying Low-, Moderate-, and High-Productivity Subareas of the BPS: As extensively discussed in the literature (see Introduction) and shown in the preceding data analysis performed by the EERC research team, geologic variables play an important role in production and must therefore be included in the modeling or have their effect minimized by narrowing the study area to smaller regional subsets that contain less geologic heterogeneity. Recognizing the importance of geologic factors in optimization calculations, the EERC research team has been developing a workflow that integrates in-house subsurface data for the purpose of optimization calculations. These efforts are ongoing and have not been finalized; thus, in this study, an alternative approach was taken to minimize the effect of geologic and reservoir heterogeneity by limiting the optimization calculations to targeted subareas of the BPS. Wells located within three relatively localized subareas of the BPS were selected to reflect low (less than 50 MMbbl cumulative 6-month production)-, moderate (50 to 100 MMbbl cumulative 6-month production)-, and high (greater than 100 MMbbl cumulative 6-month production)-productivity subareas (Figure 1). The

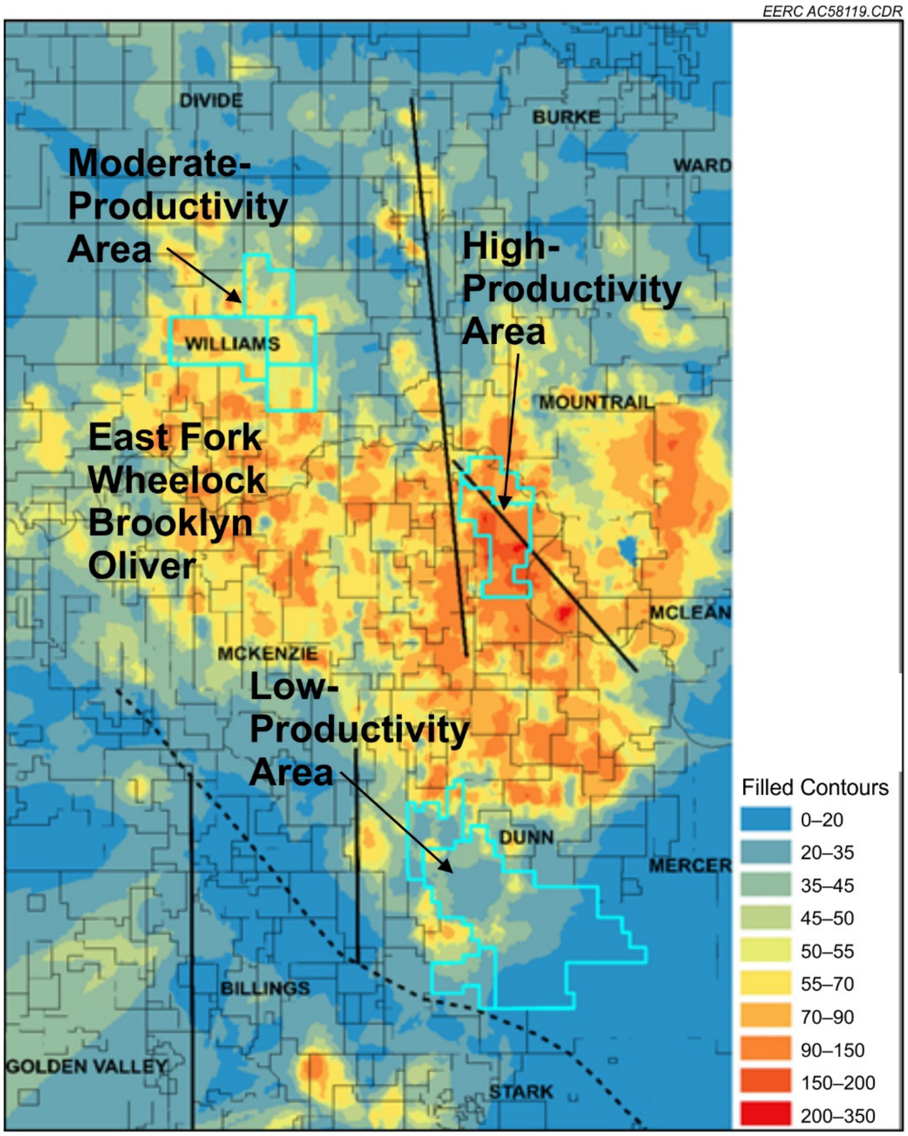


Figure 1. Heat map of BPS showing cumulative 6-month oil production (MMbbl) and identifying subareas used in current study for low-, moderate-, and high-productivity areas (light blue outlined areas). The heat map indicates well performance, with the lowest values in blue and the highest values in red.

variability in geologic and reservoir properties within these subareas is likely much less than the variability encountered across the entire BPS. Therefore, the optimization work presented in this study omitted geologic variables and focused solely on completion parameters to estimate more refined recommendations for completion design optimization for each of the three subareas.

Stated differently, the subareas were used to constrain the completion optimization calculations, which were conducted on wells within each individual area, i.e., one predictive model for each area (three models in total). The three geographically distinct subareas were associated with specific geologic characteristics (Nordeng and Helms, 2010; EERC internal database). The high-productivity area, which comprises eastern McKenzie County, has the highest average oil production performance, the greatest reservoir depth, and the highest reservoir temperature and pressure. These geologic settings translate into higher maturity levels of source rocks (Upper and Middle Bakken shales), as indicated by the highest T_{max}

and lowest HI (or higher transformation ratio) values. The bulk hydrocarbon properties impacted by higher thermal maturity are characterized by the highest API gravity values, the lowest paraffin and sulfur contents, and dryer gas composition. In contrast to the high-productivity area, the low-productivity area located to the south of the BPS core area, has the lowest average oil production performance, Bakken and Three Forks thickness, reservoir porosity, and total organic carbon (TOC) content in the source rocks (Upper Bakken Shale [UBS] and Lower Bakken Shale [LBS]) and slightly heavier and more viscous oils. The moderate-productivity area in Williams County has lower maturity levels based on T_{max}, lowest transformation ratio of organic matter estimated using HI, and highest TOC content in the UBS and LBS. The depth, thickness, and temperature of the Bakken Formation within the Moderate Productivity Area are toward the lower end of the range, while porosity and permeability are the highest among the three subareas.

High-Productivity Subarea: Two adjacent oil fields (Antelope and Elm Tree) located in McKenzie County were chosen to represent the high-productivity subarea. The wells in this subarea corresponded to the red region of Figure 2, which represents wells in approximately the top 15% of production within the BPOP Analytics Database. The total number of wells included in the high-productivity subarea was 311 (Table 1).

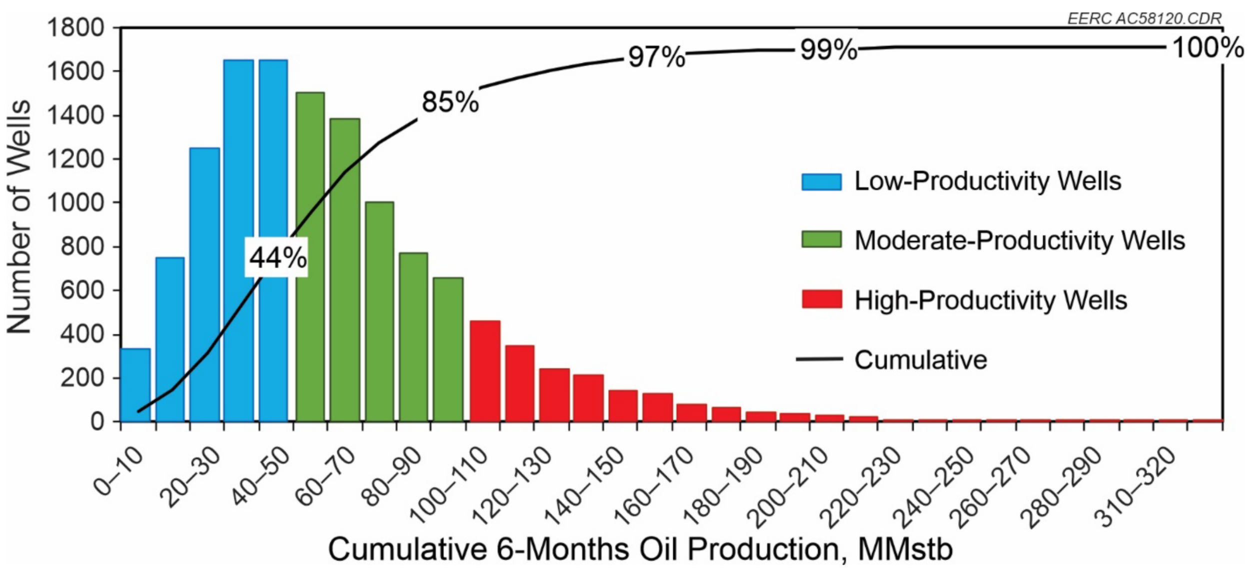


Figure 2. Histogram and cumulative frequency of cumulative 6-month oil production for wells in BPOP Analytics Database.

Moderate-Productivity Subarea: The moderate-productivity subarea included four fields (Oliver, Brooklyn, Wheelock, and East Fork) located in Williams County. The wells in this subarea correspond to the green region of Figure 2, which represents wells in approximately the 50th (median) to 85th percentile of production within the BPOP Analytics Database. The total number of wells included in the moderate-productivity subarea was 295 (Table 1).

Low-Productivity Subarea: Three adjacent fields in Dunn County (Manning, Murphy Creek, and Jim Creek) were selected to represent the low-productivity subarea. The wells in this subarea correspond to the blue region of Figure 2 and represent approximately the bottom 50% of production in the BPOP Analytics Database. The total number of wells included in the low-productivity subarea was 301 (Table 1).

Table 1. Field Names, Field Well Counts, Total Well Counts, and Average Cumulative 6-month Oil Production for Each Productivity Subarea

Productivity Subarea	Field Name	Field Well Count	Total Well Count	Average Cumulative 6-month Oil Production (MMstb) per Well
High	Antelope	226	311	115
	Elm Tree	85		103
Moderate	East Fork	136	295	66
	Wheelock	26		53
	Brooklyn	90		58
	Oliver	43		65
Low	Manning	36	301	29
	Murphy Creek	194		35
	Jim Creek	71		59

Eight of the completion features were continuous variables: total fluid (bbl), total proppant (lb), proppant concentration (lb/bbl), proppant (lb/ft), perforated interval (ft), maximum treatment rate (bbl/min), and maximum treatment pressure (psi). Three of the treatment features were categorical variables: current operator, treatment type, and proppant type. The current operator was the name of the well operator listed in the NDIC database at the time the well information was accessed (March 2019). The treatment type variable included nine categories as reported in publicly available sources: 1) crosslink, 2) crosslink HC, 3) energized fluids, 4) friction reducer HC, 5) linear gel HC, 6) hybrid crosslinked–linear gel HC, 7) slickwater, 8) slickwater HC, and 9) unknown or missing. The proppant type included five categories: 1) ceramic; 2) ceramic, sand; 3) resin-coated, sand; 4) sand; and 5) unknown or missing.

Exploratory Data Analysis: Prior to SPM-GB analysis, exploratory data analysis was conducted to evaluate correlations and statistical distributions of the target variable and completion features. Table 2 provides a correlation matrix of the target variables and ten completion features for wells located in the high-productivity subarea. The correlation matrix showed almost zero correlation between target variable and current operator name (r = 0.07), suggesting that this feature would not provide information of value to the analysis. Based on this result, current operator was omitted from subsequent SPM-GB analysis. Three of the completion feature parameters—total proppant, total fluid, and proppant per foot—showed modest positive correlations with the target variable, with r values of 0.59, 0.46, and 0.61, respectively. As expected, total fluid and total proppant (expressed as bbl/lb, bbl/lb per bbl, and bbl/lb per foot) were highly correlated. In addition, maximum treatment rate was positively correlated with total fluid (r = 0.60). Despite showing little correlation with the target variable, treatment type (r = 0.03) and proppant type (r = –0.02) were retained in the analysis, as were the remaining completion features. Figure 3 shows histograms of the target variable and ten completion features for the 311 wells located in the high-productivity subarea. The broad range of values for the completion features shows that operators have employed various designs and completion strategies even within the relatively small spatial extent of the subarea. SPM-GB analysis was used to determine whether the variation in these features affected production and to identify optimal values of each feature within each subarea.

Statistical Modeling: SPM-GB analysis was conducted separately for each of the three subareas. Across all three subareas, the model R^2 values were approximately 0.60 or higher, which was determined to be an acceptable goodness-of-fit. Figure 4 shows the feature importance diagrams for the high-, moderate-, and low-productivity subareas. Each of the completion features was determined to have a significant effect on oil production, with the lowest score for any single feature being 20 on a 100-point scale. Across all three

Table 2. Correlation Matrix for Target Variable and Ten Completion Features for Wells Located in High-Productivity Subarea

Variable	Current Operator	Treat-ment Type	Prop-pant Type	Total Fluid, bbl	Total Prop-pant, lb	Prop-pant Concen-tration, lb/bbl	Prop-pant, lb/ft	Perf. Inter-val Gross, ft	Max. Treat-ment Rate, bbl/min	Max. Treat ment Pres-sure	Cum. 6-month Oil
Current Operator	1										
Treatment Type	−0.03	1									
Proppant Type	0.04	−0.01	1								
Total Fluid, bbl	0.14	0	−0.06	1							
Total Proppant, lb	−0.04	0.05	−0.1	0.87	1						
Proppant Concentration, lb/bbl	−0.24	0.11	−0.17	−0.15	0.27	1					
Proppant, lb/ft	−0.06	0.06	−0.09	0.73	0.91	0.35	1				
Perf. Interval Gross, ft	0.07	−0.02	−0.12	0.35	0.3	−0.03	−0.03	1			
Max. Treatment Rate, bbl/min	0.3	−0.03	0	0.6	0.42	−0.29	0.5	−0.06	1		
Max. Treatment Pressure	0.43	0.01	0.02	0.44	0.4	−0.05	0.36	0.23	0.43	1	
Cum. 6-month Oil	0.07	0.03	−0.02	0.46	0.59	0.26	0.61	0.16	0.29	0.38	1

subareas, the top three most important features always included total proppant and total fluid, which supports previous data analytics approaches. The differences in feature importance were most pronounced between the high- and low-productivity subareas. For example, in the high-productivity subarea, the top two features (total proppant and treatment type) had nearly equal weighting (scores) and the next most important feature had a score of more than 20 points lower. In contrast, in the low-productivity subarea, the top five features all had scores above 80, suggesting more equal importance for a broader set of features.

Partial dependency plots were used to identify optimal values for each feature (well completion parameter) based on SPM-GB results. These plots are presented in two figures, one for total proppant, total fluid, stages, and perforated interval (Figure 5) and another for proppant (lb/ft), proppant type, proppant concentration (lb/bbl), treatment type, maximum treatment pressure (psi), and maximum treatment rate (bbl/min) (Figure 6). In the partial dependence plots, the x-axis shows the value of the feature from the minimum to maximum value in the data set and the y-axis shows the predicted value of the target variable (cumulative 6-month oil production) while holding all other features in the model constant. The red arrow in each panel shows the optimal value for each feature, defined as the value of the feature that maximizes the target variable while minimizing the feature, i.e., greatest oil production for least investment (cost) for the completion feature. For example, the upper left-hand panel in Figure 5 shows the partial dependence plot for total proppant. Predicted oil production dramatically increased as total proppant increased from 2.5 to 10 million lb per well. However, additional total proppant beyond 10 million lb resulted in no additional oil production, as shown by the horizontal blue line, i.e., no observed increase in well production with increasing total proppant. In this example, the optimum value for total proppant would therefore be 10 million lb for wells completed in the high-productivity subarea. This process was repeated for each feature and for all three subareas. The optimal result for each feature was extracted and summarized in Table 3.

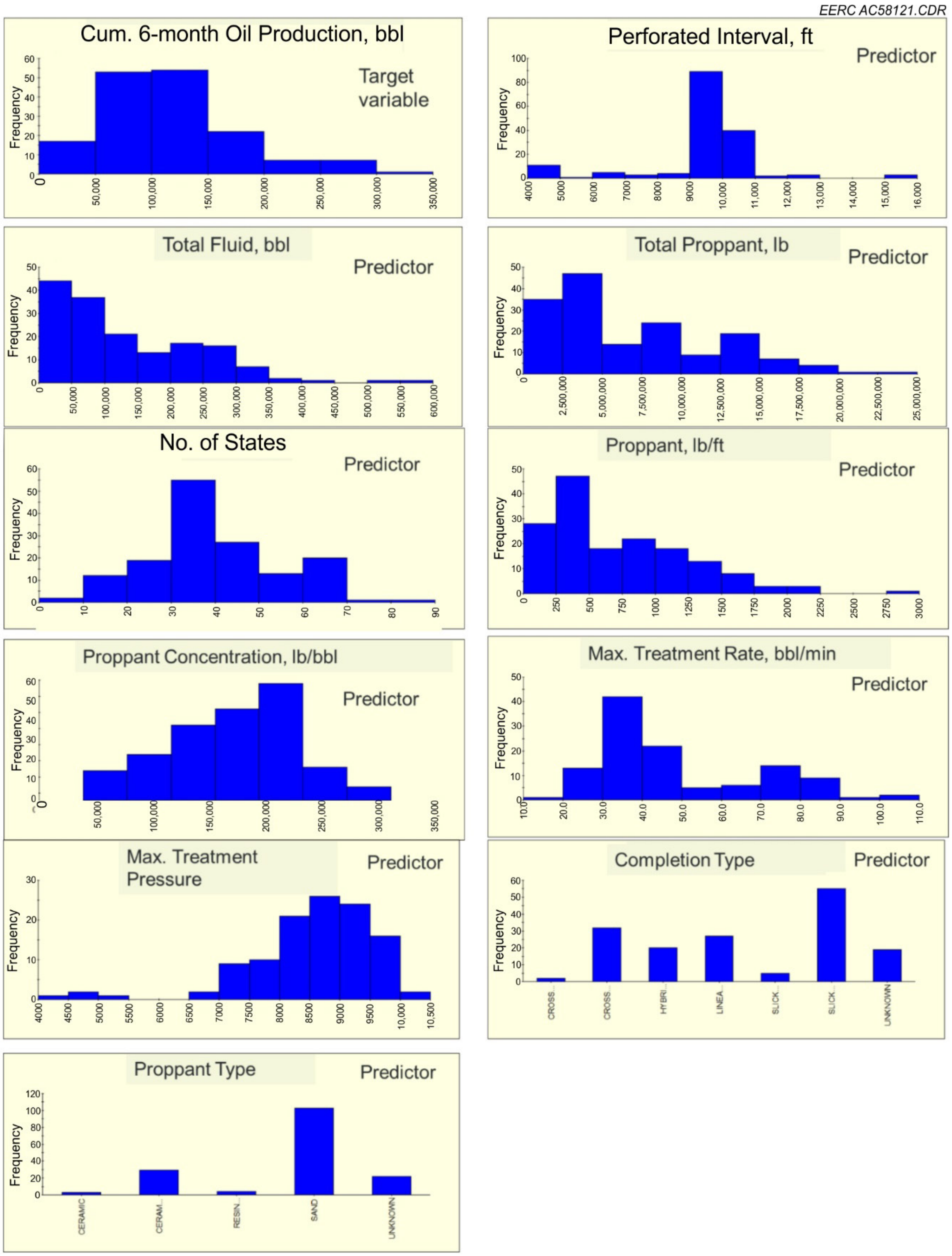


Figure 3. Histograms for target variable and ten completion features used in SPM-GB analysis for 311 wells located in high-productivity subarea.

EERC AC58122.CDR

Completion Parameter	Score	Magnitude
High-Productivity Subarea		
Total Proppant, lb	100	
Treatment Type	93	
Total Fluid, bbl	70	
Perforated Interval, gross ft	60	
Stages	59	
Proppant Type	55	
Proppant, lb per ft	52	
Proppant Concentration, lbs. per bbl	50	
Max. Treatment Pressure, psi	41	
Max. Treatment Rate, bbl per min	36	
Moderate-Productivity Subarea		
Completion Parameter	**Score**	**Magnitude**
Total Proppant, lb	100	
Stages	67	
Total Fluid, bbl	64	
Perforated Interval, gross ft	39	
Treatment Type	36	
Proppant, lb per ft	27	
Proppant Type	25	
Proppant Concentration, lb per bbl	25	
Max. Treatment Rate, bbl per min	23	
Max. Treatment Pressure, psi	20	
Low-Productivity Subarea		
Completion Parameter	**Score**	**Magnitude**
Total Proppant, lb	100	
Total Fluid, bbl	97	
Stages	94	
Treatment Type	89	
Perforated Interval, gross ft	81	
Proppant Concentration, lb per bbl	75	
Proppant, lb per ft	66	
Max. Treatment Rate, bbl per min	64	
Proppant Type	59	
Max. Treatment Pressure, psi	51	

Figure 4. Feature importance diagrams for high-, moderate-, and low-productivity subareas as determined through SPM-GB analysis.

EERC AC58123.CDR

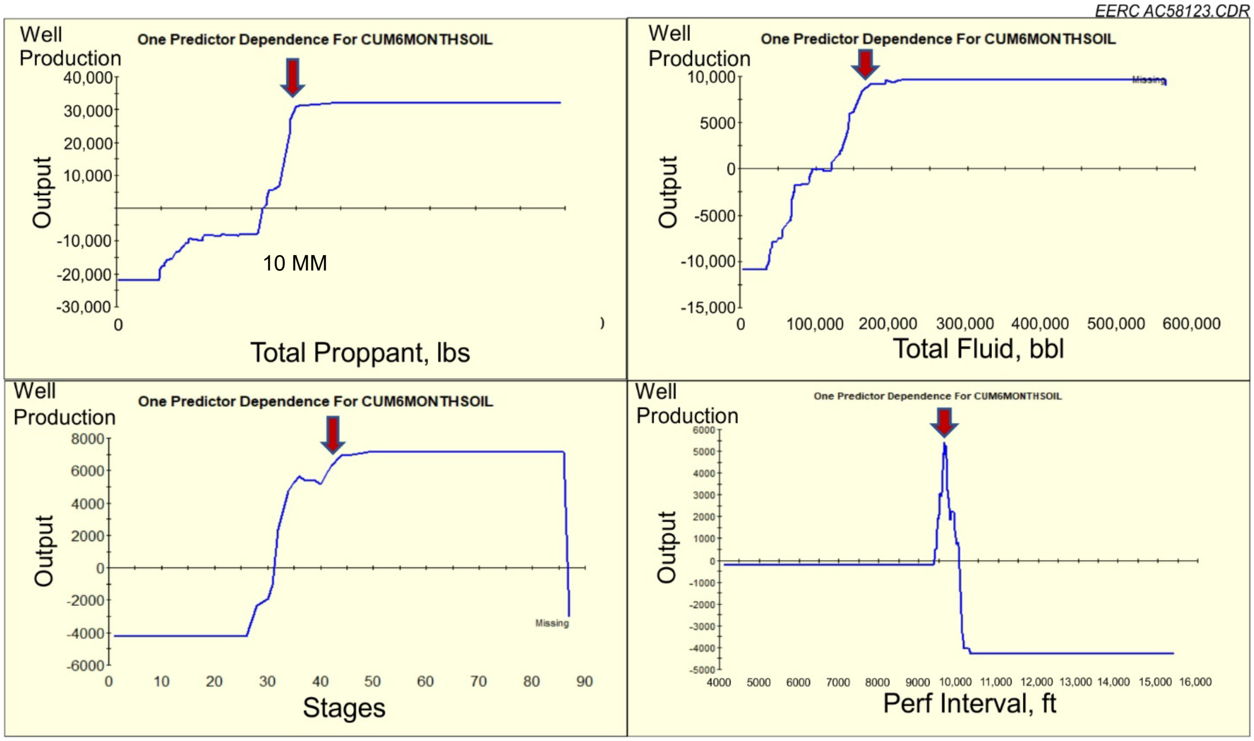
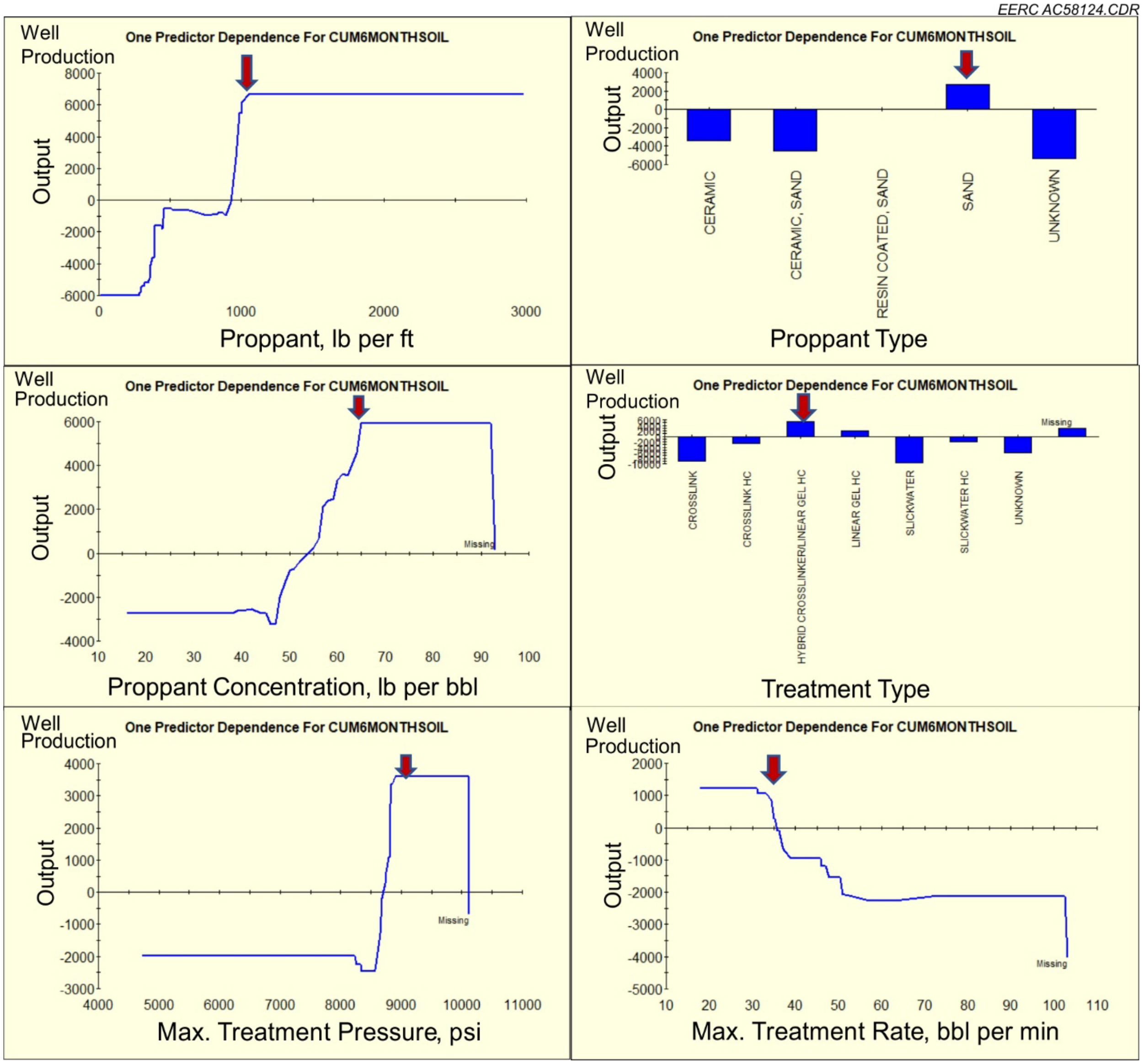


Figure 5. Partial dependence plots generated from SPM-GB for wells located in high-productivity subarea showing optimal values for total proppant (lb), total fluid (bbl), stages, and perforated interval (ft). The red arrows show the maximum value for each completion variable beyond which there was no measurable production performance observed and therefore a possible “optimum value.”

Based on optimal values for each feature shown in Table 3, the high-productivity subarea would benefit from injection of significant volumes of proppant up to 10 million lb; the moderate- and low-productivity subareas could optimize production using less proppant of 6.5 and 5.8 million lb, respectively. However, the moderate- and low-productivity subareas would require more injected fluid of 240,000 and 260,000 bbl, respectively, as compared to the 175,000 bbl needed for the high-productivity subarea. In other words, the low- to moderate-productivity areas would benefit from slickwater fractures (higher fluid amounts and lower proppant concentrations). The results also suggest that the moderate- and low-productivity subareas would benefit from a higher number of stages (up to 45). Less proppant per ft of completed lateral and lower treatment pressures of 7200 to 7800 psi were other optimal treatment parameters in the moderate- and low-productivity subareas as compared to 9700 psi in high-productivity subareas.

Interpretations for the proppant type and treatment type should be treated with caution because many wells were missing data for these features and operators reporting on these completion parameters were often inconsistent. Nevertheless, the results suggest that relatively inexpensive sand worked well in the high-productivity subarea, whereas both the moderate- and low-productivity subareas would benefit from ceramic proppant (or resin-coated sand). Treatment types maximizing performance in the three subareas were different, with crosslinked–linear gel working best for the high-productivity subarea and other treatment types, such as slickwater fractures, working best for the moderate- and low-productivity subareas.

EERC AC58124.CDR



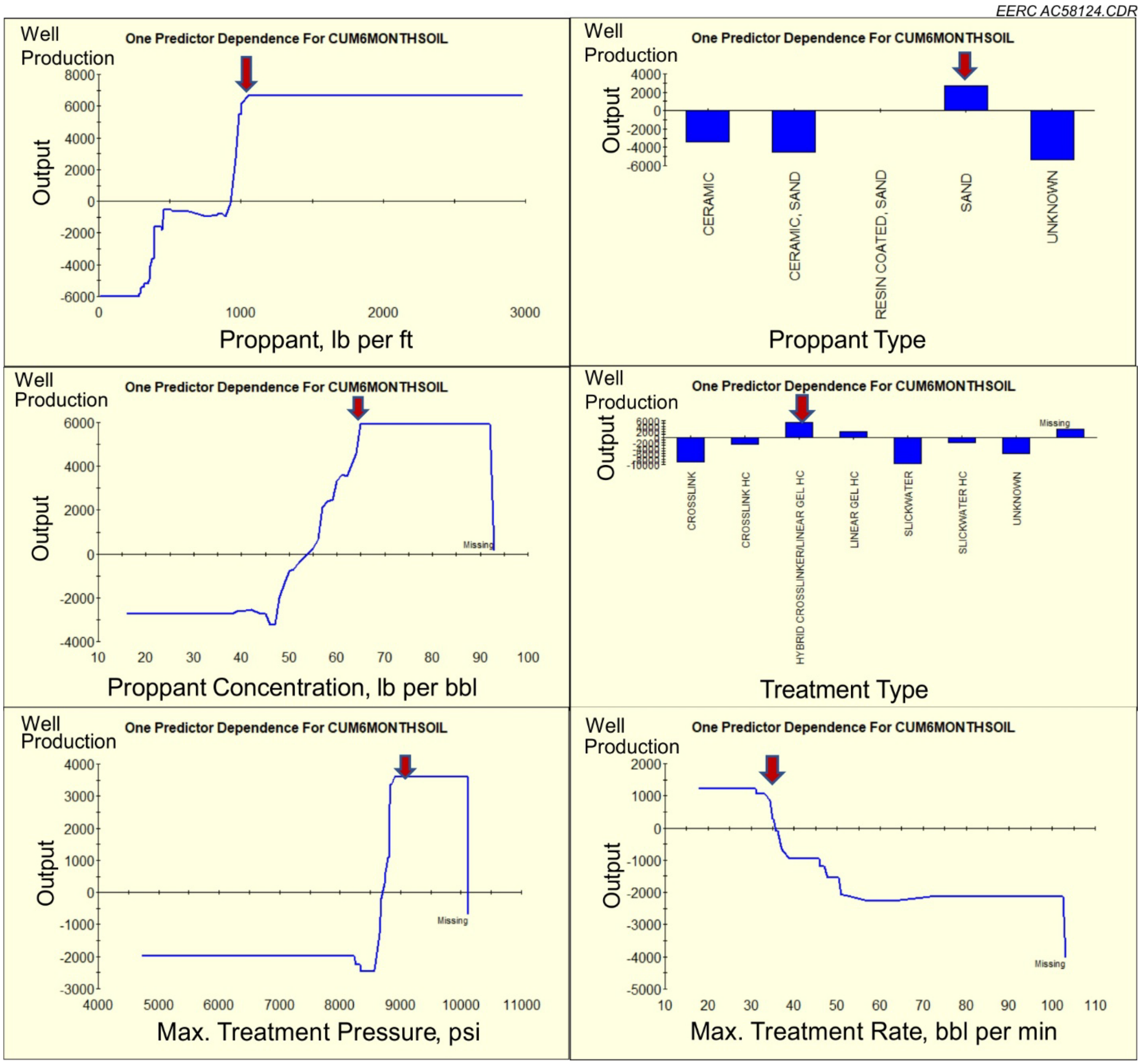


Figure 6. Partial dependence plots generated from SPM-GB for wells located in high-productivity subarea for proppant (lb/ft), proppant type, proppant concentration (lb/bbl), treatment type, maximum treatment pressure (psi), and maximum treatment rate (bbl/min). The red arrows show the maximum value for each completion variable beyond which there was no measurable production performance observed and therefore a possible "optimum value."

Table 3. Optimal Values of Ten Completion Parameters Based on SPM-GB Applied to Wells Located in Three Subareas of the BPS

Completion Parameter	High-Productivity Area	Moderate-Productivity Area	Low-Productivity Area
Total Proppant, lb	10,000,000	6,500,000	5,800,000
Total Fluid, bbl	175,000	260,000	240,000
Perforated Interval, ft	9700	10,250	10,200
Number of Stages	30–35	35	45
Proppant, lb/ft	1100–1600	640	650
Proppant Concentration, lb/bbl	65	53	37
Maximum Treatment Pressure, psi	9700	7800	7200
Maximum Treatment Rate, bbl/min	25–30	NA[1]	NA
Proppant Type	Sand	Resin-coated	Ceramic, coated sand
Treatment Type	Cross-linked–linear gel	Energized fluids, high-viscosity friction reducer HC, linear gel HC	High-viscosity friction reducer HC, slickwater

[1] Not applicable.

Conclusions

To reduce the variation from the geologic or reservoir heterogeneity on the results, completion design optimization analysis using SPM-GB was conducted on wells located in three subareas of the BPS representing low-, moderate-, and high-productivity regions. The target variable for all analyses was cumulative 6-month oil production, and each of the wells included values for ten completion design parameters (features).

Each of the completion features was determined to have a significant effect on oil production, with the lowest score for any single feature being 20 on a 100-point scale. Across all three subareas, the top three most important features always included total proppant and total fluid, which supports other work that showed these features to be significantly related to oil production.

Partial dependence plots were used to identify optimal values for each feature for the three subareas, defined as the value of the feature that maximized the target variable while minimizing the feature, i.e., the greatest oil production for the least investment (cost) for the completion feature. The results suggest different optimal configurations for the subareas, with the high-productivity subareas benefitting from higher total proppant, lower total fluid, and higher maximum treatment pressure and the middle- and low-productivity subareas maximizing oil production with less total proppant, greater total fluid, higher number of stages, and lower maximum treatment pressure.

This work highlights the value of ML techniques like SPM-GB for identifying optimal design configurations and provides information that can be used by operators to optimize completion practices and significantly reduce completion costs.

References

Bhattacharya, S., Ghahfarokhi, P. K., Carr, T. R., and Pantaleone, S., 2019. Application of Predictive Data Analytics to Model Daily Hydrocarbon Production Using Petrophysical, Geomechanical, Fiber-Optic, Completions, and Surface Data: A Case Study from the Marcellus Shale, North America. *J. Pet. Sci. Eng.* **176**: 702–715.

Chakhmakhchev, A. V., Azzolina, N. A., Kurz, B. A., Yu, X., Dalkhaa, C., Glazewski, K. A., Sorensen, J. A., Gorecki C. D., Harju, J. A., and Steadman, E. N., 2020. Advanced Analysis of Bakken Data to

Optimize Future Production Strategies. Report for North Dakota Industrial Commission Oil and Gas Research Program and Members of the Bakken Production Optimization Program (BPOP), Contract No. G-040-080, EERC Publication 2020-EERC-06-14, Energy & Environmental Research Center, Grand Forks, North Dakota (June 2020).

Dalkhaa, C., Azzolina, N. A., Pekot, L. J., Kurz, B. A., Kalk, B. P., and Harju, J. A., 2019. Bakken Production Evaluation Using Multivariate Statistical Analysis. Final Report for North Dakota Industrial Commission Oil and Gas Research Program and Members of the Bakken Production Optimization Program, EERC Publication 2019-EERC-04-16, Energy & Environmental Research Center, Grand Forks, North Dakota (April 2019).

Enverus, 2020. www.enverus.com/ (accessed May 2020).

Flowers, J. R., Guetta C. J., Stephenson P., and d'Arco J. N., 2014. A Statistical Study of Proppant Type vs. Well Performance in the Bakken Central Basin. Presented at the SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, February 4–6, 2014, SPE 168618.

Friedman, J. H., 2000. Greedy Function Approximation—A Gradient Boosting Machine. IMS 2000 Retz Lecture, Stanford University.

Friedman, J. H., 2001. Greedy Function Approximation—A Gradient Boosting Machine. *Ann. Stat.:* 1189–1232.

Gao, C. and Gao, H., 2013. Evaluating Early-Time Eagle Ford Well Performance Using Multivariate Adaptive Regression Splines (MARS). Presented at the SPE Annual Conference and Exhibition, New Orleans, Louisiana, September 30 – October 2, 2013, SPE 166462.

Griffin, L. G., Pearson, C. M., Strickland, S., McChesney, J., Wright, C. A., Mayer, J., and Weijers, L., 2013. The Value Proposition for Applying Advanced Completion and Stimulation Designs to the Bakken Central Basin. SPE Annual Technical Conference and Exhibition, New Orleans, Louisiana, SPE-166479-MS.

Huckabee, P., Jin, M., Lund, R., Nasse, D., and Williams, K., 2010. Tight Gas Well Performance Evaluation with Neural Network Analysis for Hydraulic Propped Fracture Treatment Optimization. Presented at the SPE Annual Technical Conference and Exhibition, Florence, Italy, September 19–22, 2010, SPE 135523.

LaFollette, R. F., Holcomb, W. D., and Aragon, J., 2012. Practical Data Mining—Analysis of Barnett Shale Production Results with Emphasis on Well Completion and Fracture Stimulation. Presented at the SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, February 6–8, 2012, SPE 152531.

Lolon, E., Hamidieh, K., Weijers, L., Mayethofer, M., Melcher, H., and Oduba, O., 2016. Evaluating the Relationship Between Well Parameters and Production Using Multivariate Statistical Models—A Middle Bakken and Three Forks Case History. Paper presented at the SPE Hydraulic Fracturing Technology Conference, The Woodlands, Texas, February 9–11, 2016, SPE 179171-MS.

Luo, G., Tian, Y., Bychina, M., and Ehlig-Economides, C., 2018. Production Optimization Using Machine Learning in Bakken Shale. Presented at the Unconventional Resources Technology Conference, Houston, Texas, July 23–25, 2018.

Male, F., Chastity, A., and Duncan, I., 2018. Using Data Analytics to Access the Impact of Technology Change on Production Forecasting. Paper presented at the 2018 SPE Annual Technical Conference and Exhibition, Dallas, Texas, September 24–26, 2018, SPE 191536-MS.

Nordeng, S. H. and Helms, L. D., 2010. Bakken Source System—Three Forks Formation Assessment. North Dakota Department of Mineral Resources, p. 22.

Pearson M. C., Griffin L. G., Strickland S. L., and Weddle P. M., 2018. Twelve Years and Twelve Thousand Multi-Stage Horizontal Wells in the Bakken—How Is Industry Continuing to Increase the Cumulative Production per Well. Paper presented at the SPE International Hydraulic Fracturing Technology Conference and Exhibition, Muscut, Oman, October 16–18, 2018, SPE-191455-18IHFT-MS.

Pekot, L. J., Dalkhaa, C., Musich, M. A., and Martin, C. L., 2016. Bakken Production Analysis. Energy & Environmental Research Center report, 35 p.

Swindell G., 2012. Eagle Ford Shale – An Early Look at Ultimate Recovery. Paper presented at the SPE Annual Technical Conference and Exhibition, San Antonio, Texas, October 8–10, 2012, SPE 158207.

Theloy, C. and Sonnenberg, S. A., 2013. Integrating Geology and Engineering—Implications for Production in the Bakken Play, Williston. SPE URTEC 1596247-MS.