Results in Engineering 18 (2023) 101143

Contents lists available at ScienceDirect

Results in Engineering

journal homepage: www.sciencedirect.com/journal/results-in-engineering

Horizontal well completions using data analytics

A. Alzahabi [a,*], A. Alexandre Trindade [b], A. Kamel [a]

[a] Petroleum Engineering, University of Texas Permian Basin, Odessa, TX, USA
[b] Mathematics and Statistics, Texas Tech University, Broadway and Boston, Lubbock, TX, 79409, USA

ARTICLE INFO

Keywords:
Wolfcamp
Horizontal well completion variables
Data analytics
Supervised machine learning
Multivariate linear regression predictive model

ABSTRACT

There is no quick way to measure completion configurations using simple production data, and assessing them is a challenge for many operators. It has been demonstrated by the O&G industry that completion configurations of horizontal wells influence initial well production potential and long-term performance. The present paper introduces simple, but accurate models to predict completion variables for horizontal wells in the Wolfcamp shale. These include fracture stages, clusters and cluster spacing, and perforations. We believe that these models will help define the optimum completion variables for horizontal wells in unconventional resources. The model development is based on the analysis of hundreds of horizontal wells that include the production history and parameters affecting their production behavior, including but not limited to, well completion configurations, size of proppant, type of fluid, stages, and completed interval of the lateral. We find that six key parameters are essential to precisely predict and optimize the completion variables, namely county, reservoir type, proppant amount, fluid type, rectangular overlap between wells, and initial production. Relationships among these and other parameters and their effect on the production behavior of horizontal wells were evaluated using state-of-the-art data analytics (machine learning). Multivariate linear regression models were devised to predict the four completion variables. Publicly available Well Production Performance data were used as a separate criterion in cross-validating the model predictions out-of-sample. The results of this assessment demonstrate the precision of our models with an absolute relative error of the order of 13%. The true practical advantage of this work is not only in guiding future selective completion variables for horizontal wells in the shale play, but also in providing comparative metrics in assessing different completion styles of various basins using the production history of the offset wells.

1. Introduction

It is well established in the O& G industry that horizontal drilling and hydraulic fracturing have unlocked the hydrocarbons' potential of both tight formations and unconventional resources. Thanks to recent advancements in horizontal wells completion technology, hydrocarbon production has increased drastically over time. Hydrocarbon-independent US is no longer a dream. However, optimizing completion parameters and how to complete these wells are vital keys to enhancing horizontal well performance in unconventional resources.

Several completion challenges are to be addressed. First, completion configurations/variables of the horizontal wells and their profound influence on long-term well performance. Second, need for more effective operations, keeping in mind its economic viability. The two challenges have been proven to have a substantial impact on the performance of horizontal wells and are currently assessed by operators to weigh the well potential, in addition to helping operators compare their assets to a competitive one. The assessment process includes multiple considerations and metrics, Table 1, that may lead to the effectiveness of completion techniques. The assessed metrics include completed interval of the horizontal well, perforations, clusters, cluster spacing, fracturing fluid type, average injection rate, injected volume of fracturing fluid in bbls, normalized fluid volume in gal/ft, Proppant type and size, proppant amount in pounds (lbm); proppant concentration in the injected fluid in lbm/gal. In addition, production data such as Initial Production (IP), Gas Oil Ratio (GOR), liquid Yield, and Water Cut (WCUT). Most of the oil production rate data may be easily measured or estimated from offset wells. Some metrics are publicly available while others are usually obtained through private communications. The current work will help us better understand the most enabling completion variables in the Wolfcamp formation and their influences on driving the company completion strategies, which impact the productivity of the producing

* Corresponding author.
E-mail addresses: alzahabi_a@utpb.edu (A. Alzahabi), alex.trindade@ttu.edu (A. Alexandre Trindade), kamel_a@utpb.edu (A. Kamel).

https://doi.org/10.1016/j.rineng.2023.101143
Received 21 March 2023; Accepted 3 May 2023
Available online 11 May 2023

wells. Four models, new to most of the Oil and Gas industry, are introduced to rapidly forecast the number of fracture stages, the number of clusters and perforations, and spacing between fracture clusters of horizontal wells. In developing the models, we considered a puddle of more than 200 horizontal wells obtained from companies for Wolfcamp formations A through D.

In developing the models to predict the dependent variable of completion patterns (fracture stages, number of clusters and perforations, and spacing between fracture clusters), a group of independent (input) variables was considered (209 variables of specific interest were examined from 201 horizontal wells). The variables are listed in Table 1, which include county, depth, oil Estimated Ultimate Recovery (EUR), IP 30 Oil, IP 60 Oil, the volume of the injected fluid in bbls, etc.

2. Literature review

In absence of offset data, a trial-and-error approach is still widely applied to decide on the number of hydraulic fracture stages and completion variables. Currently, there is no available benchmark in the industry to decide on some key completions variables. Therefore, an essential question remains unanswered. It is the optimum completion-dependent variable needed per well. In addition, there is uncertainty concerning the extent to which each of these variables influences the EUR from a well. For example, some believe that more stages result in higher production. Others believe that long fracture and/or higher fracture conductivity are considered the two major variables for improved productivity. The industry has accepted the trend of the increased number of fracture stages, regardless of completion technique, can achieve larger stimulated reservoir volume (SRV) and in turn, improved hydrocarbon recovery since more natural fractures are activated [1]; JPT [2]; Kazakov, E. et al., 2017)). On the other hand, some noticed a decline in the efficiency of incremental stages. Therefore, beyond a certain number of stages, the cost would surpass the incremental benefits [3].

Recent studies demonstrated an automated workflow that can be executed in real-time used for completion optimization (Rashid et al., 2014). An optimization workflow to pinpoint the number of fracture stages to maximize NPV. While some prefers equally spaced fracture stages, the prediction of optimum perforation clusters and cluster spacing in depleted reservoirs was implemented using specific logging tools [4]. Some of the past work used real-time enablement including data acquisition, visualization, and alerting for completion adjustments during operations to gain informed decisions regarding treating pressure, slurry rates, or proppant concentrations [5]. To understand the impact of analytics on the real-time completion process, (Kuuskraa et al., 2021), some authors used a Simplified Data Analytics methodology to

Table 1
The completion parameters used to develop the model.

No.	Completion metrics	Data Range	Average	Median	Statistical Deviation
1	Completed Feet	3600–11,210	6768	6702	2416
2	Number of Stages	10–58	29	27	13.5
3	Stage Length (ft)	156–404	252	220	68.5
4	Perforations/cluster	3–18	7	6	4
5	Perforations	420-2192	973	792	506
6	Clusters/stage	30–798	182	120	154
7	Cluster Spacing (ft)	13–132	54	53	31
8	Fluid (bbls)	60-655,118	263,397	220,850	141,416
9	Fluid (Gal/ft)	27,400–180,378	1558	1580	415
10	Proppant (lbm)	2,677,278–28,539,969	11,329,804	8,651,855	7,207,262
11	Proppant lbm/ft	724-2839	1575	1364	552
12	Proppant lbm/Cluster	35,100–133,759	74,366	72,066	20,327
13	Proppant lbm/Perfs	3385–21,542	11,469.8	11,994	3079
14	Avg. Prop. Concentration (ppg)	0–1	1	1	0.07
15	Avg. Rate (bpm)	38–98	66	69	7
16	Avg. Pressure (psia)	4208–10,289	6440	6416	808
17	Max. Pressure(psia)	6959–12,567	8200	8119	658
18	Initial Shut-in Pressure (ISIP)/ft	0.91–170	0.82	1	.067
19	Fluid Type	Slickwater, Linear Gel, Composite fluid, XL Borate, Hydrochloric Acid, Mud Acid			
20	Initial production (IP 30 Oil) (BBL/day)	130-3261	617	580	316
21	GOR (scf/STB)	1130-50,000	7814	6275	4777
	WCUT	0.4–0.9	0.78	1	.06
	Ip 30 Yield (bbl)	37-4870	182	146	343
22	ROA (Rectangular Overlap Area) between wells (ft^2)	0-1,590,320	569,052	492,788	502,496
23	Initial Reservoir Pressure (psia)	3900–8900	6559	6748	1183
24	Oil EUR (bbl)	23,859-6,013,040	569,052	569,052	569,052
25	Initial Producing GOR(SCF/STB)	1130–50,000	7422	5263	5720
26	Initial CGR (BBL/MMSCF)	0–320	145	125	71
27	Initial Producing OGR(BBL/MMSCF)	20–885	198	190	114
28	Rate per cluster (Bpm)	5–26	14	14	5
29	Max Prop Concentration	1–6	2	2	1
30	Fluid (Gal/perf)	2770–16,349	11,506	11,403	2770
31	True Vertical Depth (ft.)	879-12,241	10,016	10,066	879
32	County	Culbersson, Reeves, Ward, Eddy, Lea			
33	Reservoirs	Wolfcamp Uay,Wolfcamp Ua, Wolfcamp La,Wolfcamp A Wolfcamp D, Wolfcamp D4, Wolfcamp C,Wolfcamp Ma Wolfcamp La			
34	GOR EUR (SCF)	356,065–13,668,955	5,372,319	4,738,306	2,956,401
35	Initial production (IP 90 Oil) (BBL/day)	129-3116	557	512	298
36	Initial production (IP 180 Oil) (BBL/day)	129-3014	494	449	280
37	IP 90 Yield (Bbl)	44-4567	176	145	321
38	Cumulative 90 Day Oil (Bbl)	3743–123,533	45,850	43,503	21,180
39	Cumulative 120 Day Oil (Bbl)	4218–159,061	60,074	57,677	28,021
40	Cumulative 150 Day Oil (Bbl)	4606–191,261	73,020	70,495	34,662
41	Cumulative 180 Day Oil (Bbl)	4981–222,865	84,810	80,925	41,014
42	IP 30 Yield (Bbl)	37-4870	181	146	343

assess optimum well-completion performances in the Mowry Shale considering geology, well performance, and well completion data of 18 horizontal wells.

Furthermore, the effects of combining spacing with perforation limited entry concept on the hydraulic fractures initiation and propagation of simultaneous jobs were studied [6]. Their work suggested that limited-entry perforation helps not only to match the lengths of fractures growing simultaneously from multiple clusters, but also to level proppant placement of across the entire series of perforation clusters. However, this does not equalize their widths.

A mathematical model was introduced to optimize the number of fracture stages based on production performance and fracture treatment of horizontal wells [7]. In another study. Placement of perforation clusters, Clusters' spacing, stress distribution, and fracture mechanics were found to be the most crucial parameters to control production (Cheng et al., 2012).

In 2013, a case study implemented an integrated engineering approach to optimize multiple sets of lateral horizontal wells in the oil zones of the Eagle Ford formations. Data was used to model the intervals of stimulation-stage and to assess the easy-to-measure injection responses during the fracturing treatments (Stegent et al., 2013). Other studies have established correlations between completion variables and key production factors (Alzahabi et al., 2020).

It is crucial to mention that number of fracture stages has increased over time and industry practice varies widely among basins (Guk et al., 2015). A 12-stage count is common in Barnette, whereas the average number of fracture stages in the Bakken is 30. Fig. 1 shows the increase in intensity in the number of fracture stages with time for Bakken, Eagle Ford, and Marcellus. On the other hand, Haynesville, Barnett, and Fayetteville show no change.

In addition, numerous advances have contributed to the evolution of the completion process. Fig. 2 through 5 show an increasing trend for proppant being used, the volume of fluid, lateral length, and fluid injected/ft.

The proppant in 1bm/ft used in completing horizontal wells has dramatically gone up by 500% (Fig. 2). The lateral length of horizontal wells has also gone up from 3500 to 7000 ft (Fig. 3). The fracturing fluid volumes have risen by 800% (Fig. 4). The cumulative production (MBOE) over 12-month per 4500 ft doubled during the same time interval, as shown in Fig. 5. Table 2 reveals the completion parameters-industry trend in the Permian Basin [8] for well depths (7000–10,500 ft). Many attempts have been introduced to combine data analytics, physics-based completion optimization, development capital optimization, and risk assessment to generate empirically based models for completion optimization [9].

In summary, it is believed that the EUR of horizontal wells usually increases as the fracture stages count increases. However, as the count increases, the efficiency of incremental stages declines resulting in lower long-term productivity.

2.1. Application of data analytics in hydraulic fracturing processes

A volume of publications and studies was published that contained analytics techniques to extract unseen relations and patterns between various parameters and sweet spot locations. Exploratory data analysis of production history was used to characterize fluid flow and parameters controlling it (Jansen et al., 1996). On the other hand, some authors (Grieser et al., 2006) applied data clustering techniques and self-organized maps to feature completion and reservoir data that affect production, mainly for Barnet shale. His data was archived as useful information extracted by statistical noise reduction techniques.

With the use of data from other basins such as Haynesville shale wells, [10]; investigated number of frac stages, average treatment rate, total proppant amount, average proppant per stage, proppant concentration, fluid type, and the total number of clusters. Their work showed that there is a direct relationship between variables such as the number of fracture stages, fracture conductivity, and proppant concentration and production. Also, his work showed that using cross-linked gels has a slight advantage in 12-month production intervals.

Furthermore, data analysis was implemented to group similarly stressed rock for fracture treatment which led to increasing the number of clusters to enhance the production of the Eagle Ford (Slocombe et al., 2013). A boost of 28% was seen in the wells in comparison with their offsets. A method to increase estimated oil recovery via installing smart completions and inflow control devices was introduced through an increase in reservoir contact, despite an expected steep decline in production(Chaudhary et al., 2016). Technology unlock an increase in the number of stages to 20, in a 1000-m lateral section which led to an

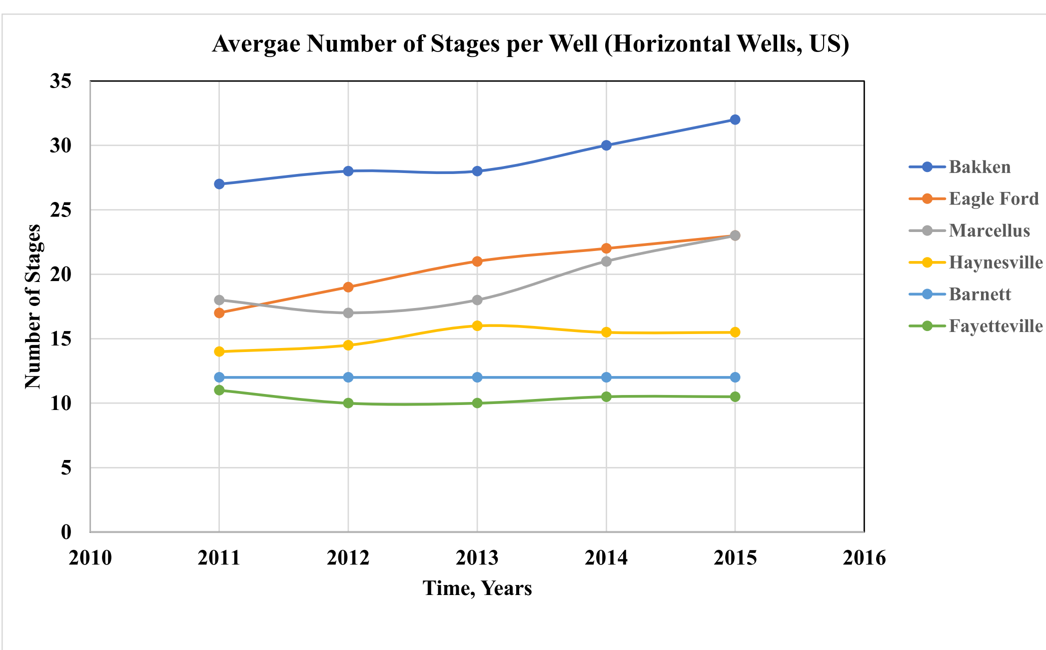
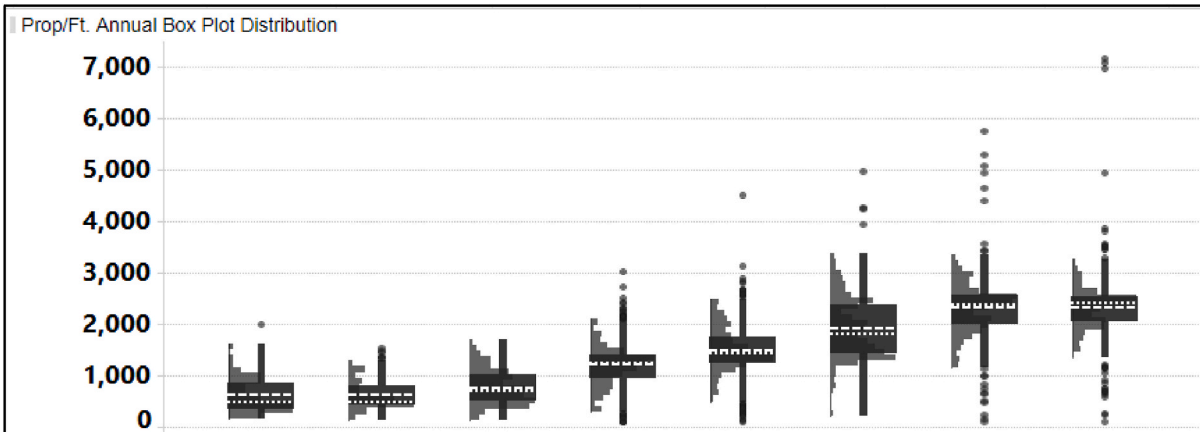


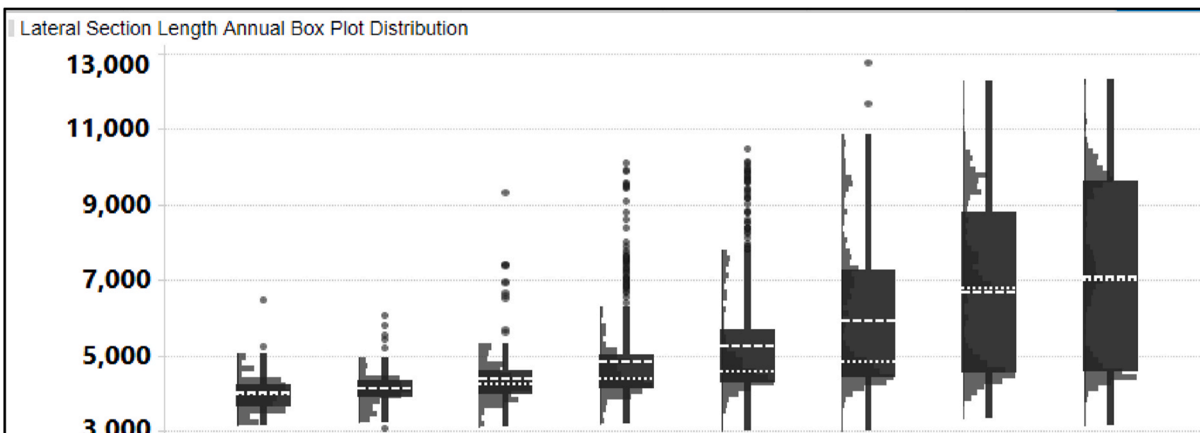
Fig. 1. Average number of stages per well (horizontal wells, US) after Guk et al., 2015.



	2011	2012	2013	2014	2015	2016	2017	2018
Count	79	155	210	440	552	542	1,051	964
Median	477	490	701	1,227	1,414	1,811	2,374	2,393
Avg	603	614	759	1,196	1,488	1,914	2,305	2,316
Min	149	131	116	100	106	201	102	111
Max	1,989	1,534	1,707	3,006	4,511	4,961	5,756	7,152

Fluid/Ft. Annual Box Plot Distribution

Fig. 2. Recent history and advancements in the hydraulic fracturing process(Proppant/ft).



	2011	2012	2013	2014	2015	2016	2017	2018
Count	79	155	210	440	552	542	1,051	964
Avg	4,013	4,124	4,390	4,860	5,249	5,940	6,677	7,086
Min	3,158	3,084	3,123	3,206	3,013	3,014	3,349	3,147
Max	6,476	6,080	9,321	10,130	10,477	12,776	12,294	12,314
Median	3,992	4,118	4,264	4,400	4,584	4,860	6,798	7,022

Fig. 3. Recent history and advancements in the hydraulic fracturing process (Lateral section Length).

increase in production [8]. These efforts, assuming there is no operational or human preference for a certain completions strategy, showed that a benchmark for all shale plays could be established. That benchmark may be linked to production trends and/or main reservoir parameters affecting its performance in shale plays.

Principal Component Analysis was used to build metrics to assess the strategic success of the completion treatments for key profitable shale plays. Thickness and kerogen were identified as the main discriminating parameters for shale well completions [11]. Employing Permian Basin data related to well completions, through a workflow introduced lessons learned for future completions [8]. Data Analytics most commonly used tools for unconventional resources were listed [12].

3. Statement of the problem

For the Permian Wolfcamp, key questions relevant to the optimum variable of completion strategies are not yet answered as many parameters may have to be considered. These parameters show a relationship with the completion process, horizontal well, production history, and constraints or limitations if any. Table 3 summarizes these parameters.

The current paper is a serious attempt to apply data analytics techniques to choose and optimize the number of key paraments affecting

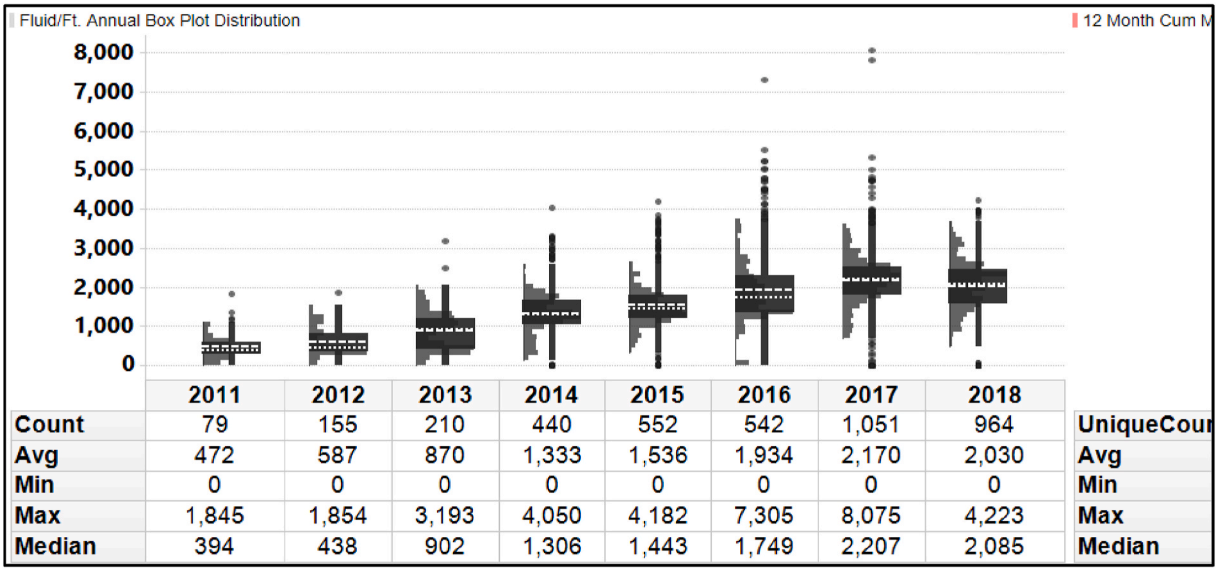


Fig. 4. Recent history and advancements in the hydraulic fracturing process (Fluid/ft).

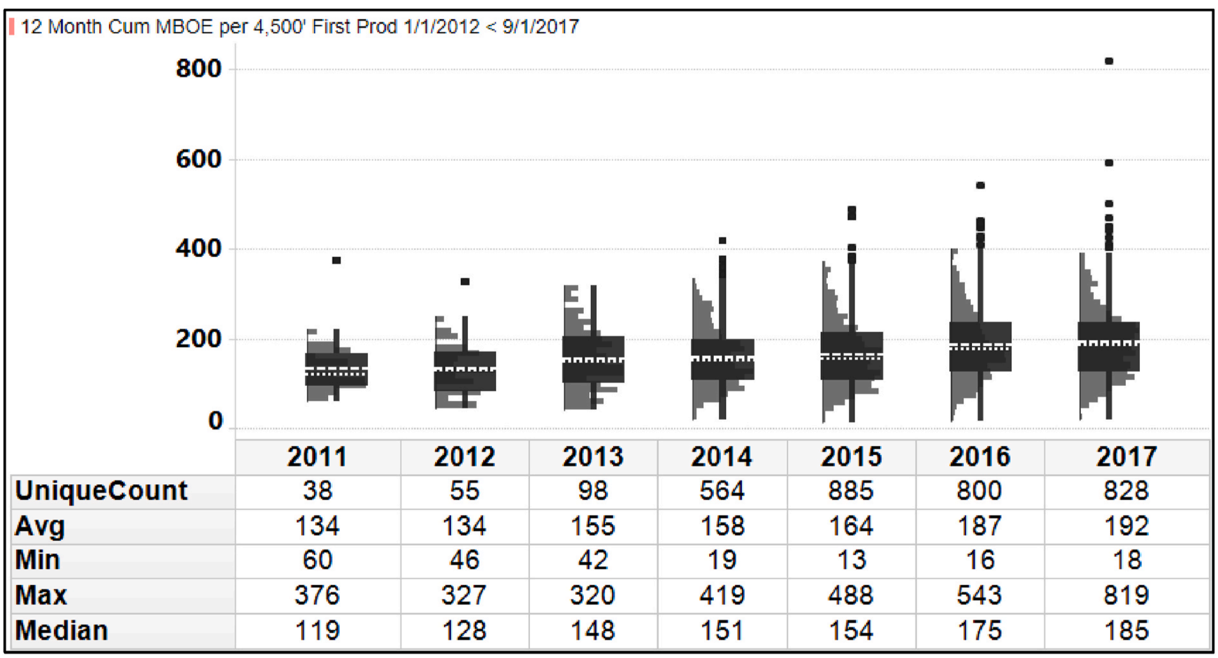


Fig. 5. Recent history and advancements in the hydraulic fracturing process (12-Month cumulative MBOE per 4500 ft).

Table 2
Completion parameters used in the Permian Basin (after [8].

Completion variables	Permian
Proppant Size	40/70, 30/50, 20/50, 20/40, 100 mesh
Proppant Type	White sand vs, Brown sand
Proppant Volume, lb/cluster	30-100,000
Proppant Volume, lb/ft	600–5000
Cluster Spacing (ft)	10–80
Clusters/stage	3–10
Perforations/cluster	30–42
Proppant concentration, ppg	1.5–6
Rate per cluster, bpm	8–30
Rate per hole, bpm	1.75–2.3
Fluid volume, bbl/ft	18–100
Stage length, ft	100–360
Fluid type	Slickwater/Hybrid/Crosslink

completion strategies aimed at production enhancement.

4. Research methodology

In this paper, the workflow and data requirements build on the framework developed in Ref. [12] and others [13,14]; and [15]. This previous preliminary study serves as a foundation for viable prediction techniques, the fact that these models need refinement, and new methods in selecting the input variables to be considered. The relationships among the variables (Table 1) were studied by the use of new advancements in multivariate regression (supervised machine learning with multiple outputs), which incorporate automatic screening of important variables and group-wise inclusion/exclusion of factor variables.

We considered the data from the Permian Basin and developed

Table 3
Investigated parameters for permian wolfcamp.

Set	Parameters
Fracture	• Number of fracture stages • Number of perforations per stage • Number of clusters • Average injection rate • The max shut-in pressure (ISIP/ft and 5′ ISIP/ft)
Completion Fluid	• Fluid type • Injected fluid volume • Injected fluid volume/perf • Average injection rate
Proppant	• Proppant type • Proppant amount • Proppant amount/perf • Proppant amount/cluster • The optimum proppant concentration • Proppant size
Horizontal Well	• Length of completed interval • Minimum and maximum spacing between wells • The optimum number to avoid fracture driven interactions (frac hits)
Production	• IP production variables, • Cumulative production variables • Production ratios (GOR, WC, yield)
Constraints	• Any economic limits • Capacity of the producing wells • Effects of spacing on production profiling

models to effectively select the completion variables. Utilizing state-of-the-art data analytics, these models identify key completion configurations from easy-to-measure field data. The procedure deviates from the orthodox approach of predicting yield/EUR based on completion variables. The pool of predictor variables was chosen according to a local operator's interest in finding models that can be used to compare his completion strategy with other operators' completion practices within the same basin. The question becomes: which of the available variables can feasibly function as predictors? This is the main missing piece of the puzzle. Although still rooted in the industry standard of using engineering techniques to pick the completion variables, these machine-learning-based completion strategies are now beginning to form a bridge between time-consuming conventional methods and complex numerical models. While simulation is still commonly used by petroleum engineering specialists, predictive models are easier to use and are beginning to fill the existing gap.

As far as we are aware, there are no straightforward models in the industry to estimate key completion variables such as fracture stages (stages), spacing, number of clusters (clusters), and perforations of horizontal wells (perfs). Our workflow involves two main steps. The first consists of developing a multivariate regression model for the four outputs (stages, perfs, clusters, spacing) using the group lasso method, a relatively new style for multitasking machine-learning algorithm where strong correlation exists among output variables while incorporating group-wise shrinkage and model selection. The second phase assesses the model's performance on both in-sample and out-of-sample data.

5. Selecting the optimum input parameters

The data from two hundred and one horizontal wells were acquired from private companies to aid in developing the models. The data represent Permian Basin Wolfcamp formations A through D [16] and Alzahabi et al., 2020). Fig. 6 shows the geographical location of the selected wells. The following pool of predictor (or input) variables were considered: oil and gas EUR, IP 30 oil and gas, IP 180 oil, the volume of the injected fluid in bbl and gal/ft., IP 60 gas, GOR (60 days and for the life of the well), oil yield (cumulative oil) for 30 and 60 days, gas yield (cumulative gas), number of days the well on production, number of pounds of proppant in pounds and lbm/ft, TVD (True vertical depth), and rectangular overlap area between horizontal wells (ROA).

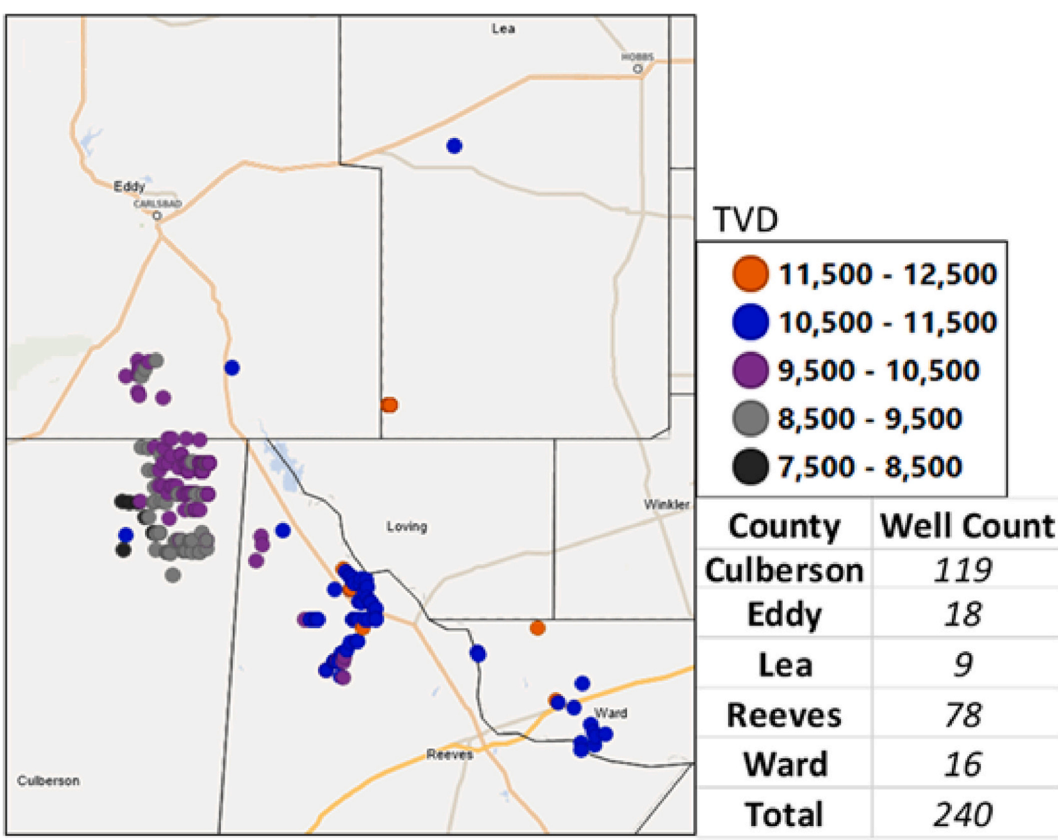


Fig. 6. Geographical coordinates of the wells used to build the model.

An ordered heatmap correlation matrix among these inputs is shown in Fig. 7. The ordering uses correlation as a distance metric, and therefore performs a crude form of clustering. We note a fair amount of correlation among most of these inputs (the correlation coefficient is zero when the random vectors are independent). The generalized pairs plot in Fig. 8 is a sophisticated graphic that provides visual scatterplot information with a mix of parameters, numerical (in this case the 4 outputs), and factor (in this case county and reservoir) [17]. The color-coding scheme here represents the reservoir type. The fact that the density functions from different reservoir types in each of the lower 4 diagonal panels (which represent the distributions of the 4 outputs) are well separated, leading to the conclusion that reservoir type is a strong predictor.The panels above these 4 diagonals represent bivariate distributions, i.e., the joint behavior of the outputs corresponding to the intersection of the appropriate row and column. The fact that the contours in these off-diagonal panels are tightly clustered, which is indicative of the high degree of the correlation that exists among each pair of outputs.

This robust correlation in the output variables suggests the use of a multitask machine learning algorithm so that the resulting prediction can optimally exploit this feature of the data. However, with about one-third of the inputs being categorical (factor or group) variables, model selection techniques for multivariate linear regression would tend to select only some of the components that constitute each group. This and the fact that the pool of predictors is greatly amplified with so many categories to code for led us to choose a “group lasso” model [18]. This fairly recent innovation in machine learning shrinks the coefficients in all the constituent levels of a group together,and by so doing points the way to the regression model that should be fitted. In the final model thus identified, county and reservoir were found to be the most important predictors.

6. Results and discussion

The value of the models developed lies in their simplicity and high accuracy for the private database used in this work. More importantly, it is based on easy-to-measure simple production data. Unlike the common technique of analogy and following other operators’ practices in

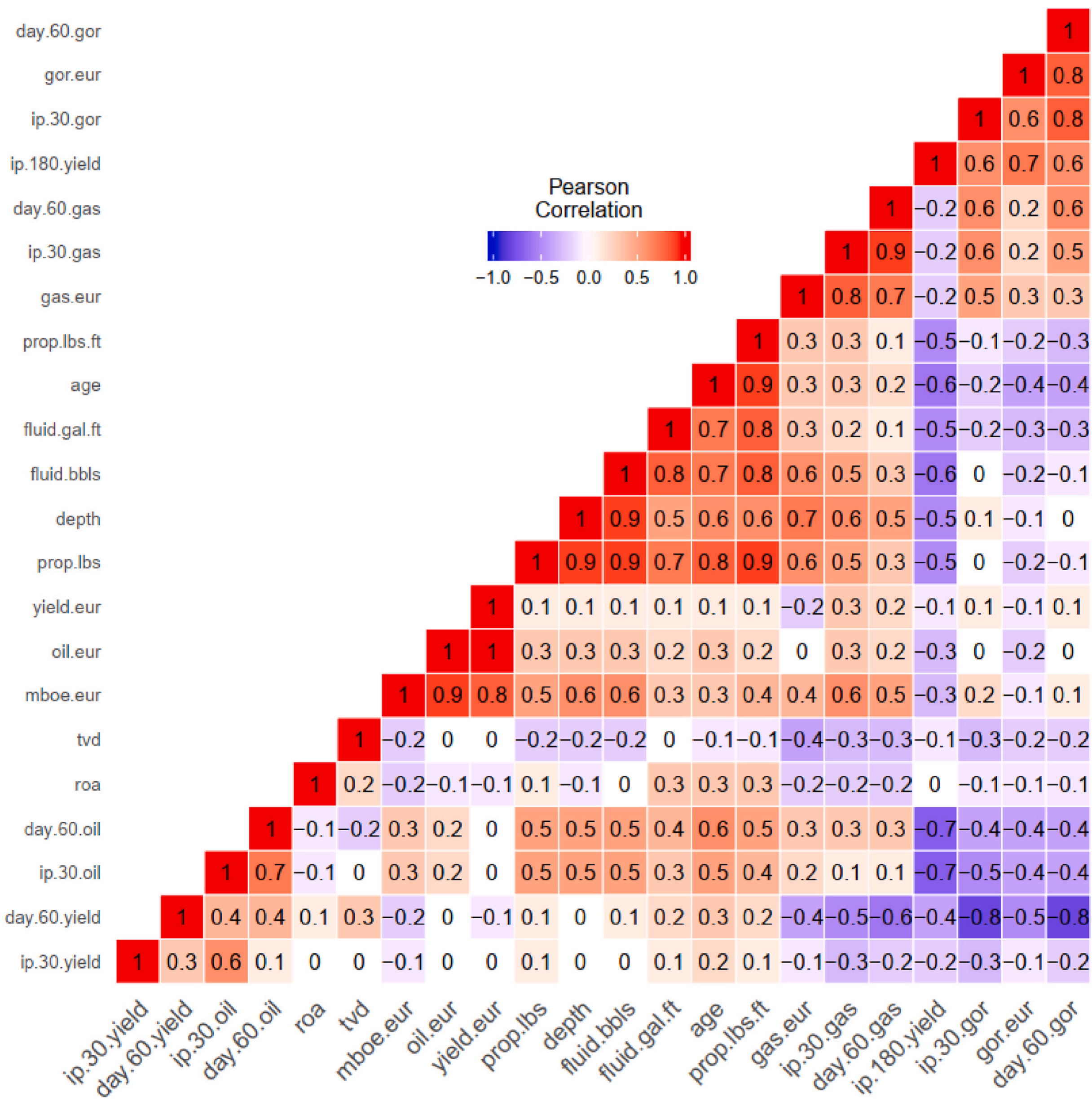


Fig. 7. Heatmap correlation matrix among the input variables.

Fig. 8. Generalized pairwise scatter plot for the categorical predictors' county and reservoir (first 2 diagonal panels), and the 4 completion parameters (last 4 diagonal panels). The coloring scheme distinguishes data from the same reservoir, as indicated in the 2nd diagonal panel.

developing unconventional resources. One needs a few input variables to estimate one of the key completion configurations that drive performance.

The coefficients for the linear regression model fitted to the data are presented in Table 4, starting at"Intercept" and ending at "Day 60 Gas". Note, however, that these values were not obtained via the usual least-squares regression, since a sparsity-inducing algorithm was used (group lasso). To obtain the predicted value for a particular variable, say "Stages", one simply multiplies the coefficients appearing in the column for "Stages" with the appropriate row parameter value. (Since "County" and "Reservoir" are group variables, one simply picks the appropriate coefficient corresponding to the particular county and reservoir names.) The resulting prediction, say Ystd, must then bede-normalized and exponentiated according to the indicated location (μ) and scale (σ) parameter values appearing in the first two rows of Table 4, as follows:

$$Y = \exp\left(\mu + \sigma \times Y_{std}\right)$$

The following example demonstrates how to proceed with the models in order to predict the completion variable Perfs at County Lea and Reservoir W-LA from a particular setting of the production variables (called "production point" in the 6th column of Table 4). Referring to Table 4, proceed as follows: Step 1: Multiply the values in column 3 (Perfs) by those in column 6 (Production Point), and store the results in column 7 (Predicted Perfs). Sum up the values in this last column to obtain Ystd = −1.03. Step 2: Back transform Ystd according to the above equation in order to get the predicted Perfs value (Y), using the correct location and scale values of mu = 6.7897 and sigma = 0.5019:

$$\text{Predicted Perfs} = \exp\left(6.7897 + 0.5019 \times Y_{std}\right) = 484$$

Note that since county "Culberson" and reservoir "Reservoir W" (Wolfcamp) is the referencecategories, their corresponding coefficient values are zero. Variables not appearing in Table 4 (e.g., TVD) were not viewed as being predictive by the group lasso algorithm, and are therefore excluded. An examination of how well the model fits the data can be seen in Fig. 9, which compares the actual observed versus model-predicted values. An examination of how well the model fits the data can be numerically summarized by the R-squared value of 88%. Fig. 9 comparesthe actual observed versus model-predicted values. However, this in-sample diagnostic is of limitedusefulness in practice, as it does not speak to the predictive ability of the model on out-of-sample data. For that, we turn to the machine-learning accepted norm of "K-fold cross-validation (KCV) [18]. This procedure was implemented using K = 5. Thus, in each test/training set split of the data,4/5 were used to fit the group lasso model and generate predicted values for the remaining 1/5 of the data. The percent absolute relative prediction error (ARPE) was used to assess the quality of predictions:

$$ARPE = 100 \times \left[\frac{y_{predicted} - y_{observed}}{y_{observed}}\right]$$

The in-sample and out-of-sample ARPE values are shown in Figs. 10 and 11, respectively, bymeans of boxplots. Note that in the in-sample ARPEs, KCV is not carried out; i.e., the model is fitted to the entire dataset and then used to generate predictions for each of the 201 wells. In In contrast, the model never knows about the existence of the out-of-sample data when it attempts to predict it. Naturally, we would expect

Table 4
Completion parameter model coefficients for perfs, clusters, and spacing models. * Indicates that the tested data belong to that county and reservoir referred to in the tested Production Point.

β Values	Stages	Perfs	Clusters	Spacing	Production Point	Predicted Perfs
Location Parameter	3.3092	6.7897	4.9851	3.7692		
Scale Parameter	0.4735	0.5019	0.7913	0.5509		
Intercept	−1.968E+00	−2.467E+00	−2.661E+00	2.574E+00	−2.467E+00	1
County Culberson	0.000E+00	0.000E+00	0.000E+00	0.000E+00	0	0
County Eddy	−2.650E-02	9.671E-02	2.869E-02	−1.872E-02	0	0
County Lea	−4.716E-02	8.151E-02	−2.003E-01	3.367E-01	1*	8.15E-02
County Reeves	1.051E-02	−9.253E-02	1.079E-02	−3.460E-02	0	0
County Ward	4.266E-01	2.531E-01	4.097E-01	−4.898E-01	0	0
Reservoir W	0.000E+00	0.000E+00	0.000E+00	0.000E+00	0	0
Reservoir W-A	1.207E-01	8.649E-02	1.496E-01	−7.855E-02	0	0
Reservoir W–C	5.928E-02	−1.021E-01	−2.998E-03	1.121E-01	0	0
Reservoir W-D	4.887E-02	−1.252E-01	−2.543E-02	1.401E-01	0	0
Reservoir W-D4	1.688E-01	−6.036E-02	8.969E-03	1.319E-01	0	0
Reservoir W-LA	1.004E-01	−6.157E-03	5.758E-02	7.895E-03	1	−6.16E-03
Reservoir W-MA	−1.419E-01	8.916E-02	2.424E-01	−1.626E-01	0	0
Reservoir W-UA	9.610E-02	1.315E-01	9.183E-02	−3.268E-02	0	0
Reservoir W-UAY	1.600E-01	1.643E-01	1.538E-02	1.756E-01	0	0
ROA	1.824E-07	2.047E-08	3.412E-07	−4.669E-07	282,120	5.77E-03
Age	−9.198E-05	−1.035E-04	−4.547E-05	−1.518E-04	5400	−5.59E-01
Depth	2.222E-04	2.204E-04	1.838E-04	3.196E-05	10,923	9.58E-01
Fluid (bbl)	1.552E-06	1.446E-06	−4.629E-07	5.314E-07	144,758	2.09E-01
Fluid (gal/ft)	3.408E-04	1.463E-04	1.217E-04	9.375E-05	1558	2.28E-01
Proppant (lbs)	−5.778E-09	2.132E-08	9.978E-09	−2.308E-08	3,370,443	7.19E-02
Proppant (lbm/ft)	9.060E-05	2.676E-04	8.562E-04	−1.191E-03	724	1.94E-01
Oil EUR	2.955E-09	−3.315E-10	−4.837E-09	7.423E-09	268,260	−8.89E-05
Gas EUR	−7.081E-11	−2.034E-10	−2.740E-10	−6.054E-11	1,608,303	−3.27E-04
GOR EUR	−7.068E-07	1.761E-06	1.058E-06	−1.636E-06	5995	1.06E-02
MBOE EUR	6.095E-11	−7.395E-11	−1.934E-10	1.452E-10	536,311	−3.97E-05
IP 30 Oil	4.499E-07	1.852E-06	7.296E-07	9.173E-07	399	7.39E-04
IP 30 Gas	3.944E-07	−7.884E-08	5.039E-07	−8.997E-07	1552	−1.22E-04
IP 30 GOR	7.239E-06	−1.923E-06	6.672E-06	−9.464E-06	3890	−7.48E-03
IP 30 Yield	1.137E-05	1.720E-05	4.572E-06	1.158E-05	229	3.94E-03
IP 180 Yield	−1.296E-04	7.706E-05	−8.190E-05	1.013E-04	3162	2.44E-01
Day 60 Gas	1.963E-07	4.327E-08	2.540E-07	−4.063E-07	63,128	2.73E-03

Y_{std} = Sum the values = -1.03, Predicted Perf = 485.

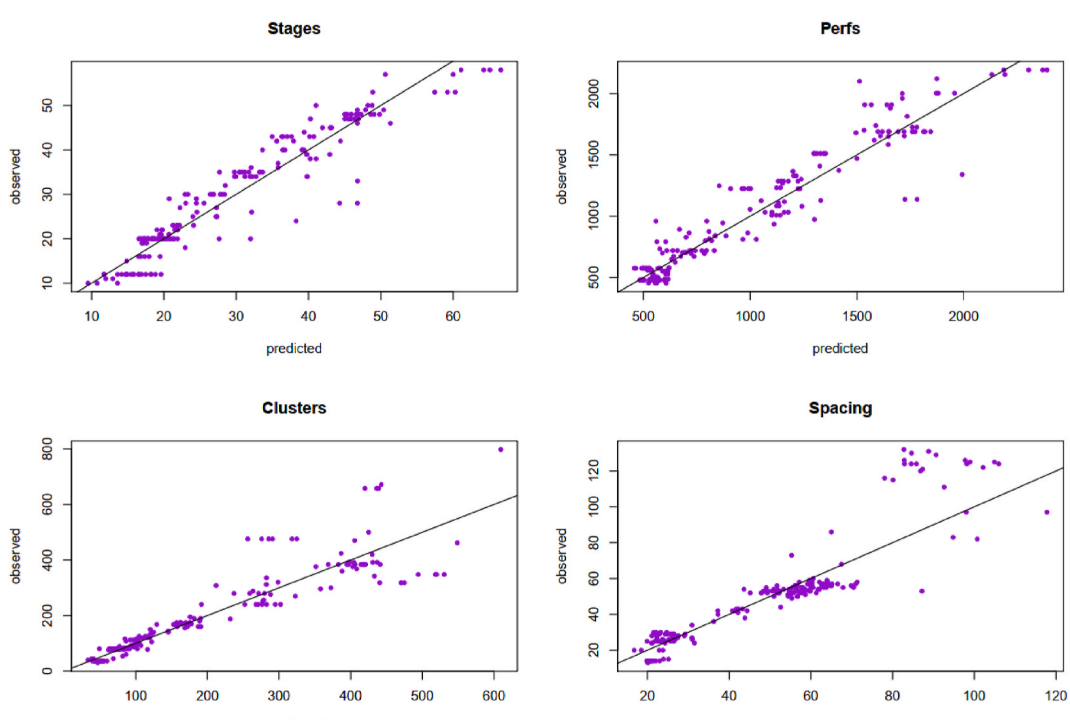


Fig. 9. Predicted versus actual number of stages, perfs, clusters and spacings.

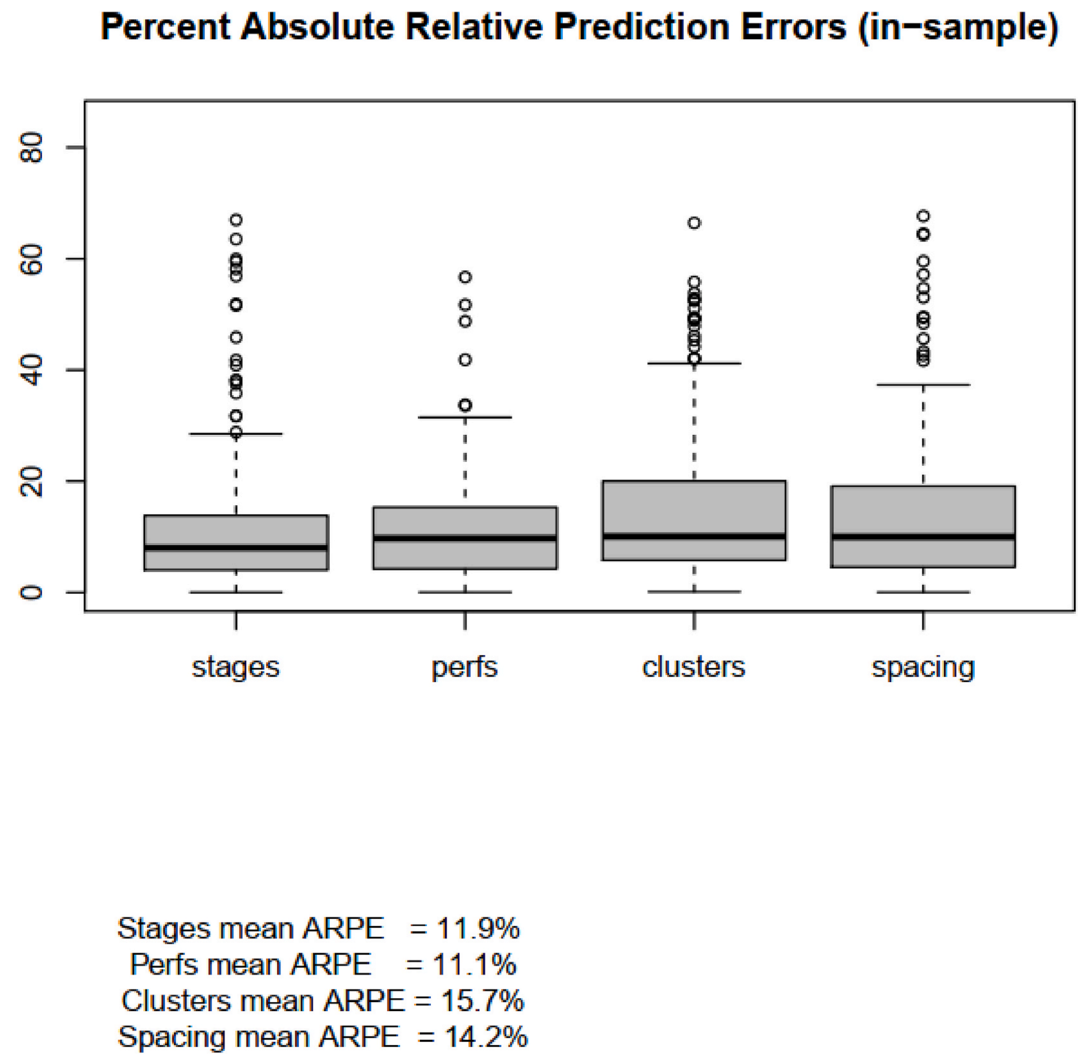


Fig. 10. Percent absolute relative prediction errors (in-sample).

in-sample predictions to fare better, which is indeed the case. Compare for example the Stages in-sample and out-of-sample means of 11.9% and 12.9%, respectively. We also note that predictions for Perfs have smaller ARPEs, with out-of-sample means of the order of 13%, while those for Clusters and Spacing have out-of-sample means of the order of 18%. (There are a few extreme outlying values, particularly for Spacing where the largest is 80%.)

7. Conclusions and recommendations

In this paper, we proposed new stages, spacing, clusters, and perforation models based on completions of horizontal wells data. We used outputs of actual data on wells of the Permian Basin to guide us through input selection for the models. In effect, we are introducing fast-to-use models that the O&G industry can use to estimate stages, cluster spacing, and the number of clusters and perforations. The models can be used in real-life operations. Our work progressed as follows:

- Wide-ranging consulting with professionals and review to find all completion variables and define industry practice in the Wolfcamp formations
- Generation of predictive models using privately owned data to be able to determine optimal completion design from industry practice
- The building of new models using data analytics to choose optimum completion parameters, after the exclusion of worthless variables such as average injection rate (bpm) from the given group of independent variables
- Analysis and validation of the models based on arbitrarily selected data sets from the public database.

The following are the main conclusions that can be drawn from the paper.

1. The final models identified the county and reservoir to be the most vital predictors of the completion parameters. This finding is to be expected due to the spatial clustering of wells.
2. The proposed models may be utilized to forecast the initial production of horizontal wells using a few input completion variables with sufficient precision.
3. Lessons learned from the completion practice in the Permian Basin will guide future practice for successful completion. The models proved to be accurate when tested on publicly available data.
4. Predictions for perfs have smaller ARPEs, with out-of-sample means of the order of 13%, while those for clusters and spacing have out-of-sample means of the order of 18%. The testing showed Stages in-sample and out-of-sample means ARPE of 11.9% and 12.9%, respectively.
5. It is expected that the predicted variables from this work may be used for several purposes. First, the variables may be used as a performance comparative measure to evaluate the well performance of a

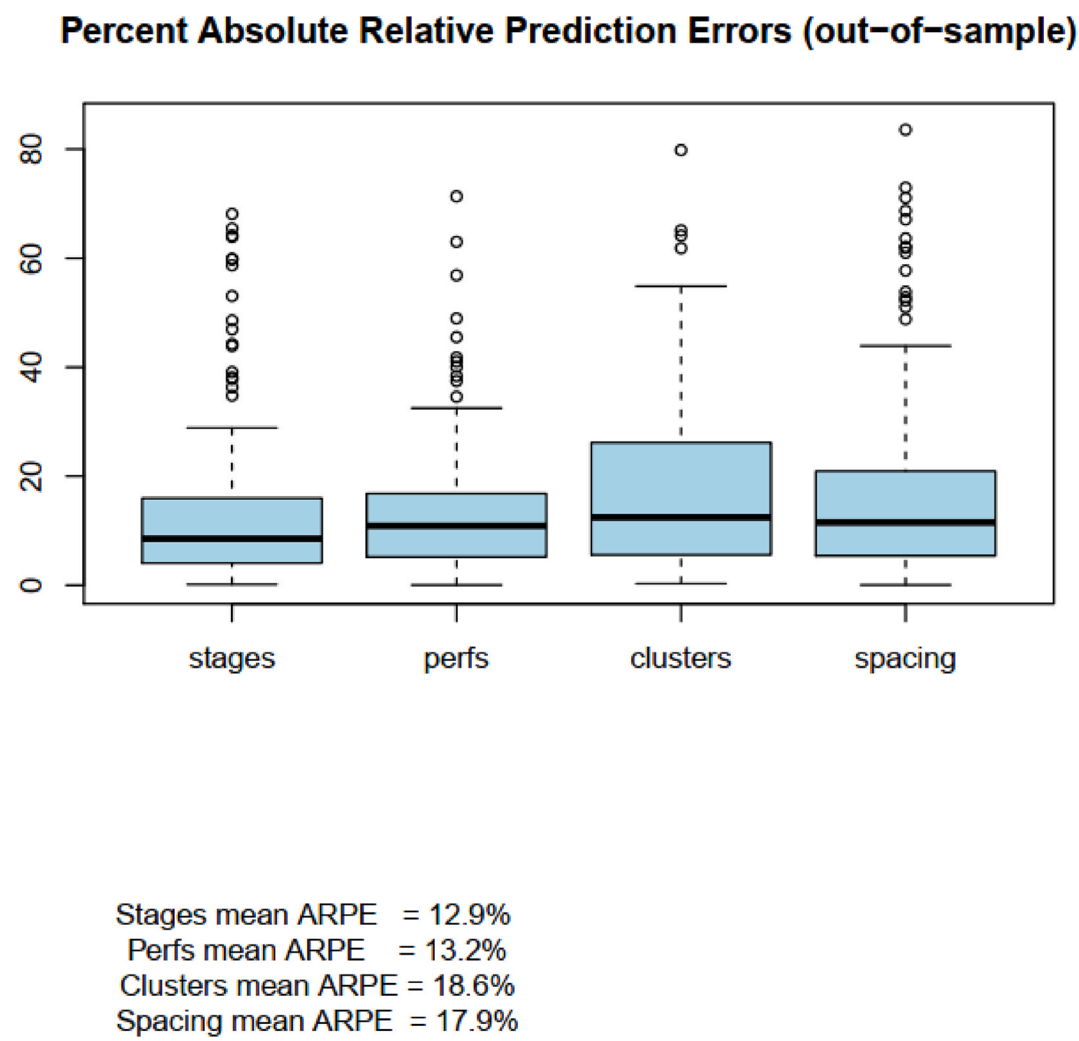


Fig. 11. Percent absolute relative prediction errors (out-of-sample).

specific play versus that of a competitor. Second, it may be used as a tool to enable quick decisions on certain completion variables needed for horizontal wells, which can facilitate further estimation of fracture stages, clusters, cluster spacing, perforations, and other important configurations. Third, in cases of multiple leases with a variety of completion configurations, the models may work as a tool to optimize the production of a lease. Estimation of the completion variables will be used to discriminate against different reservoir categories (Wolfcamp LA, B).

6. The significance of the new models suggested in the paper lies in their lack of sophistication for a practicing engineer to optimize oil production from horizontal wells. Most substantially, it is based on completion configurations that may be chosen reasonably promptly.
7. In multilayered formations, the models should be applied separately for each reservoir type.
8. As a recommendation, additional models could be suggested in forthcoming work with the collection of additional data, and intensive selection may begin on completion's designs in the Wolfcamp of the Delaware Basin.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

The authors would like to thank Wade Baustian and the support provided by CIMAREX for providing the data used in this work.

Nomenclature

Reservoir W the formation W in which the lateral was drilled
ROA rectangular overlap area: the area of overlap from wells in the same section with one another determined by their legal spacing location, assuming an Xf of 770′ and Hf of 200′ and rectangular drainage area
Age number of days between the time the well was completed and January 01, 2001
Fluid (bbl) the volume of fluids pumped downhole to initiate fracture and place proppant
Fluid (gal/ft) fluid volume in gallons per foot
Proppant (lbs) the pounds of proppant pumped per fluid gallon pumped
Proppant (lbm/ft) the pounds of proppant pumped per foot
Yield Condensate yield, MMSCF/STB
TVD the furthest true vertical depth drilled

WCUT Water cut (%)
Stage length the length of each fracture stage (ft.)
Reservoir Variable a numeric variable distinguishing between reservoirs Wolfcamp C-D and Wolfcamp A
Oil, Gas, MBOE EUR the estimated ultimate recovery of oil and gas. MBOE = Oil + Gas/6
GOR amount of gas produced per oil produced. (Scf/Stb)
IP initial production rates

References

[1] T.R. Roe, A. Seale, D. Snyder, Design and deployment of a new fracturing port technology to increase stage number capability of openhole multistage systems, Proceedings of the SPE (2011), https://doi.org/10.2118/146790-MS.

[2] J.P.T. staff, Stage multiplier technology provides ultrahigh stage numbers, Petrol. Tech. 64 (2012) 44–47, https://doi.org/10.2118/0612-0044-JPT, 2012.

[3] B. Ran, M. Kelkar, Fracture stages optimization in bakken shale formation, in: Proceedings of the Unconventional Resources Technology Conference, 2015, https://doi.org/10.15530/URTEC-2015-2154796 paper URTEC-2015-2154796, 20–22.

[4] S. Zaker, S.M. Nafchi, M. Rastegarnia, et al., Prediction of new perforation intervals in a depleted reservoir to achieve the maximum productivity: a case study of PNN logging in a cased-well of an Iranian oil reservoir, Petroleum 6 (2) (2020) 170–176.

[5] H. Stephenson, T. Nguyen, G. Murrell, R, Artificial intelligence for real-time monitoring of fracture driven interactions and simultaneous completion optimization, in: Proceedings of the SPE Canada Unconventional Resources Conference, 2020, https://doi.org/10.2118/199967-MS. Virtual.

[6] G. Izadi, R. Settgast, D. Moors, et al., Fully 3d Hydraulic Fracture Growth within Multi-Stage Horizontal Wells, 13th International Congress of Rock Mechanics, 2015, 978-1-926872-25-4.

[7] J. Guo, F. Gu, J. Zhou, Optimizing the fracture numbers and predicting the production performance of Hydraulically fractured horizontal wells, in: Proceedings of the SPE ATCE, 1997, https://doi.org/10.2118/97-108.

[8] O.A. Jaripatke, I. Barman, J.G. Ndungu, et al., Review of Permian Completion Designs and Results, Proceedings of the Society of Petroleum Engineers ATCE, Dallas, Texas, 2018, https://doi.org/10.2118/191560-MS.

[9] G. Voneiff, P. Bastian, Data to Decision: A Unified and Rapid Workflow for Unconventional Reservoirs Blending Data Analytics, Physics-Based Completion Optimization, and Investor-Oriented Economics, Proceedings of the at the SPE/AAPG/SEG URTC, Houston, Texas, USA, 2021, https://doi.org/10.15530/urtec-2021-5454.

[10] N. Modeland, D. Buller, K.K. Chong, Statistical analysis of the effect of completion methodology on production in the Haynesville shale, in: Proceedings of the North American Unconventional Gas Conference and Exhibition, 2011, https://doi.org/10.2118/144120-MS. The Woodlands, Texas, USA, June 2011.

[11] A. Alzahabi, M.Y. Soliman, G. Thakur, et al., Horizontal Completion Fracturing Techniques Using Data Analytics: Selection and Prediction, Proceedings of the 51st U.S. Rock Mechanics/Geomechanics Symposium, San Francisco, California, USA, 2017.

[12] A. Alzahabi, A. Kamel, A.A. Trindade, W. Baustian, Data Analytics Quickly Predict Number of Fracture Stages in Horizontal Wells, Proceedings of the American Rock Mechanics Association, New York City, New York, 2019. ARMA-2019-0475.

[13] A. Syed, M.F. Azman, Z. Awang, et al., Defining the optimum well completion for marginal field development – an approach, in: Proceedings of the Offshore Technology Conference Asia, Virtual and Kuala Lumpur, Malaysia, 2022, https://doi.org/10.4043/31552-MS, 2022.

[14] N. Al-Hajri, M. Javed, A. Barghouti, H. Al-Shuwaikhat, " big data analytics maximizes value from smart well completions,", in: Proceedings of the the Abu Dhabi International Petroleum Exhibition & Conference, 2021, https://doi.org/10.2118/207623-MS. Abu Dhabi, UAE.

[15] J. Emerson, W. Green, B. Schloerke, J. Crowley, D. Cook, H. Hofmann, H. Wickham, The generalized pairs plot, J. Comput. Graph Stat. 22 (79) (2012), 91.

[16] R. Sorkhabi, P. Palash, Data analysis of the Permian Basin Wolfcamp and bone spring leads to better understanding of production sweetspots, in: Proceedings of the SPE Annual Technical Conference and Exhibition, Virtual, 2020, https://doi.org/10.2118/201730-MS.

[17] T. Hastie, R. Tibshirani, M. Wainwright, Statistical Learning with Sparsity: the Lasso and Generalizations, CRC Press, Boca Raton, 2015.

[18] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, second ed., Springer, New York, 2009.