SHALI DESCRIP'TIVE ANALYTICS: WHICH PARAMETERS ARE CONTROLLING PRODUCTION IN SHALE

Mohagheg behaviour, S. D., Intelligent Solutions, Inc. & West Virginia University

ABSTRACT

Descr iptive Analytics is the first step of a three-step data-driven analytics workflow used for managing and optimizing completion, production and recovery of shale wells. The comprehensive data-driven analytics workflow for the unconventional resources is called Shale Analytics (Mohaghegh 2017). The key behind Shale Analytics is the incorporation of all field measurements that contribute to the productivity of shale wells. There are workflows in the market that claim to be data analytics related but do not make use of all the available field measurements when performing their analyses. These workflows are mainly based on traditional statistical algorithms rather than Artificial Intelligence and Machine Learning. Such approaches represent different versions of Decline Curve Analysis.

Shale Descriptive Analytics takes into account seven categories of field measurements; (i) well construction and trajectory, (ii) well spacing and stacking, (iii) reservoir characteristics, (iv) completion design, (v) hydraulic fracturing implementation, (vi) operational constraints, and (vii) well productivity. Each of the above categories of field measurements include several parameters. Shale Descriptive Analytics provides two types of insight on the contribution of all the field measurements to well productivity. The first type of insight compares and quantifies the contribution of the different categories of field measurements to well productivity. The second, more detailed type of insight compares and quantifies the contribution of each of the parameters of the first six categories to the final category that is well productivity and then compares all the parameters to one another. The Shale Descriptive Analytics presented in this article demonstrate the results of more than 800 shale wells in one of the most productive shale plays in Texas.

Two conclusions have been achieved as the result of this study. (a) In the early life of a shale asset, when the wells are NOT too close to one another (when Frac-Hit is not an issue), using well productivity indices (such as initial production, initial decline rate, first 30, 60, 90, 120, 180 and 365 days of cumulative production, etc.) can provide realistic insight for completion optimization, well productivity and recovery. (b) Once the number of wells in a given asset increases, resulting in the reduction of the distances between parent and child wells (Frac-Hit impacts production and recovery), well productivity indices will no longer be able to provide the required insight for modeling and analysis of field measurements. This is because as the number of wells increases in a given shale asset, the fracture-driven interaction between wells (also known as Frac-Hit) takes over the overall productivity of all the wells in the field. Frac-Hit not only negatively influences parent and child wells productivity and recovery, it completely undermines all the
existing techniques (traditional techniques such as RTA and numerical simulation as well as all the
existing techniques based on Data Analytics) for completion and production optimization of shale wells.
At the conclusion of this article, a new approach to overcome this specific problem is introduced.

TECHNOLOGY USED FOR SHALE DESCRIPTIVE ANALYTICS

The Descriptive Analytics involves no modeling. The Descriptive Analytics provides the foundation upon
which successful predictive data-driven modeling can be performed. It mainly covers the processes
through which the practitioner develops a deep and valuable understanding of the collected data from the
field. The physics and the geology associated with unconventional plays are quite complex. Because of
this complexity, when traditional statistical algorithms are used to observe the behavior and the correlation
of the field measurements as they related to well productivity index, no patterns and/or trends can be
detected.

Figure 1 and Figure 2 show the chaotic relationship between twelve formation and completion related
parameters with well productivity for a large number of wells in the Marcellus shale. Lack of any trends
and patterns from data collected from shale wells is a common occurrence. Therefore, it becomes hard to
make any general judgement on the contribution of the measured parameters to the well productivity. To
simplify the process and to avoid analysis of the complexities that are associated with the measured data,
many engineers and geoscientists end up concentrating on production data without any attention to what
has caused the production to behave in a certain fashion.

This is one of the main reasons for the traditional Decline Curve Analysis (DCA) to become the most
popular technique for the analysis of the production from shale wells. Traditional Decline Curve Analysis
is the plot of production data as a function of time, drawing of a curve using a simple mathematical
equation (hyperbolic, harmonic, or exponential decline), and then trying to match the two by modifying a
couple of coefficients in the mathematical equation. DCA shows the well productivity as a function of
time without providing any scientific and/or realistic information about the contribution of tens of
parameters that control the physics and the geology of the shale plays. These parameters that are the main contributors to, and the controllers of, the well productivity are related to the physics of the formation, completion, operation, and production in shale plays. However, it seems that lack of incorporating physics in the models that are now controlling all decision makings in the shale plays, apparently does not bother the scientists and engineers in the operating and the service companies.

![Graphs showing field measurements](image)

*Figure 2. This figure shows how field measurements such as treatment rate and pressure, proppant per stage, clean volume, proppant concentration, and slurry volume correlated with well productivity (180 Days Cum. Gas Production) for a large number of wells in the Marcellus shale. The lack of patterns and trends represents the complexity of the physics and the geology associated with shale plays.*

Traditional Decline Curve Analysis lets you look at the results without any information on how such results were achieved. In traditional Decline Curve Analysis parameters that control well productivity such as well construction and trajectory, well spacing and stacking, reservoir characteristics, completion design, hydraulic fracturing implementation, and operational constraints, are completely ignored. Judgment on the contribution of such parameters to well productivity is left to the engineers and practitioners and it ends up becoming a non-scientific and purely guesswork approach to decision making.

Shale Descriptive Analytics provides a unique alternative to traditional statistical approaches for the analysis of all the field measurements. Shale Descriptive Analytics can help engineers and geoscientists to discover any existing trends and patterns in the collected data. To perform any kind of scientific analyses and in order to make any reasonable completion and production related decisions, it is important to learn the contribution of different parameters to the overall productivity of the shale wells. The objective of the Shale Descriptive Analytics is to:

- Discover trends and patterns in the contribution of measured data to well productivity,
- Provide visual representation to demonstrate the contribution of each parameter in a given category to well productivity,
- Provide visual representation to demonstrate the contribution of each category of parameters to well productivity.

**Fuzzy Sets**

Shale Descriptive Analytics uses fuzzy set theory (Mohaghegh 2000) in order to address the complexity of the field measurements in the shale plays. Using multi-valued logic instead of Aristotelian two-valued
logic, fuzzy set theory mimics the human brain on how it makes sense of the non-exact (fuzzy) observations and communicated information. In fuzzy set theory, all variables are divided into multiple fuzzy sets with non-crisp overlapping boundaries. This type of classification of variables results in each incident of a given variable to be defined by its partial membership in multiple sets instead of being defined (categorized) only by two values (0 and 1 – yes and no – black and white – off and on). For example, if we decide to categorize well quality (the variable) based on its productivity (180 days Cum. Oil Production) and call them poor, average, and good wells, using fuzzy set theory we can qualitatively define each well and then translate the qualitative definition into quantitative characteristics using the mathematics of fuzzy set theory for the Descriptive Analytics.

Figure 3 shows an example of how a given well that has produced 56,981 barrels of oil in its first 180 days of production can be defined using fuzzy set theory. When the granularity of the well productivity (180 days of cum. oil production) is defined using three fuzzy sets (poor, average, and good), then this well is partially average and partially poor (Figure 3, graph on the left), having a fuzzy membership of 0.33 in the set of poor wells and a fuzzy membership of 0.67 in the set of average wells. When the granularity of the well productivity is defined using four fuzzy sets (poor, average, good and very good), then this well is average (Figure 3, graph in the middle), having a fuzzy membership of 1.00 in the set of average wells. When the granularity of the well productivity is defined using five fuzzy sets (poor, average, good, very good and excellent), then this well is partially average and partially good (Figure 3, graph on the right), having a fuzzy membership of 0.35 in the set of good wells and a fuzzy membership of 0.65 in the set of average wells.

![Figure 3. Fuzzy set theory qualitatively categorizes well productivity and uses mathematical definitions for their quantification.](image)

**Supervised Fuzzy Cluster Analysis**

Shale Descriptive Analytics incorporates a new and innovative version of soft cluster analysis that has been developed specifically for the purposes of identifying the contribution and the impact of large number of field measurements on well productivity for the shale wells. This technology is called “Supervised Fuzzy Cluster Analysis” and its details have been covered in couple of previously published documents (Mohaghegh 2015 and Mohaghegh 2017). “Supervised Fuzzy Cluster Analysis” addresses three major issues associated with cluster analysis when it is used for Shale Descriptive Analytics. The issues associated with cluster analysis are:

I. the optimum number of clusters that can maximize the release of information content (minimize entropy) from the data set,

II. the optimum number and the combination of parameters to be used in order to perform cluster analysis, and

III. how to make sense of the results generated by the cluster analysis in a manner that will be useful in the decision making process.

Before explaining the details of “Supervised Fuzzy Cluster Analysis”, it is necessary to explain the differences between the two main types of classification algorithms namely, the traditional (hard) cluster
analysis such as k-mean clustering and the fuzzy (soft) cluster analysis. What is common between these two clustering algorithms are (a) identification of the optimum number of clusters, and (b) selection of the cluster centers, both of which are done by the algorithm to maximize the efficiency of the cluster analyses. “Supervised Fuzzy Cluster Analysis” modifies these two items in order to accomplish the objectives of Shale Descriptive Analytics.

Figure 4 shows the difference between the two main types of classification algorithms namely, the traditional (hard) cluster analysis such as k-mean clustering (Figure 4, graph on the left), and the fuzzy (soft) cluster analysis (Figure 4, graph on the right). This figure demonstrates that the soft (fuzzy) cluster analysis provides much better information regarding the status of the data when it is compared with the traditional hard cluster analysis.

The problem presented in this figure demonstrates a simple, two dimensional data set. In this example, the type and the amount of information that is released (shown at the bottom of the figure) and communicated by the traditional (hard) cluster analysis algorithm (without looking at the picture/plot) indicates that points “A” and “B” are very similar since both have a membership (\( \mu \)) of 1.0 in Cluster 1, and a membership of 0.0 in Cluster 2. Furthermore, it shows that points “C” and “D” are very similar since they both have a membership (\( \mu \)) 0.0 in Cluster 1, and a membership 1.0 in Cluster 2.

However, when the actual plot of the data set is revealed and the cluster centers and the locations of these points (“A”, “B”, “C”, and “D”) are observed, it becomes obvious that the traditional (hard) cluster analysis have not done a good job of releasing information from the data. Discovering and releasing information from the data set (that is usually unobservable, due to large number of dimensionality) is the
main objective of the clustering algorithms. When fuzzy (soft) cluster analysis is applied to the same data set, while the cluster centers and the location of these points (“A”, “B”, “C”, and “D”) are the same, the released information by the fuzzy (soft) cluster analysis is quite different.

In fuzzy (soft) cluster analysis since the membership of each point in each cluster is a function of its distance to the cluster center, then the results clearly describe (without looking at the picture/plot) that points “A” and “D” should be far away from one another and points “C” and “B” should be quite close to each other. This is due to the fact that points “A” and “D” have large membership in one of the clusters and small membership in the other cluster. Point “A” has a membership ($\mu$) of 0.35 in Cluster 1, and a membership of 0.65 in Cluster 2 while point “D” has a membership ($\mu$) of 0.68 in Cluster 1, and a membership of 0.32 in Cluster 2.

Furthermore, this algorithm shows that points “C” and “B” are very similar to each other because of their membership in each of the clusters. This is due to the fact that points “C” and “B” have almost similar memberships in both clusters. Point “C” has a membership ($\mu$) of 0.51 in Cluster 1, and a membership of 0.49 in Cluster 2 while point “B” has a membership ($\mu$) of 0.48 in Cluster 1, and a membership of 0.52 in Cluster 2. Therefore, when the actual plot of the data set is revealed and the cluster centers and the locations of these points (“A”, “B”, “C”, and “D”) are observed, it becomes obvious that the fuzzy (soft) cluster analysis has done a great job of releasing information from the data.

Supervised Fuzzy Cluster Analysis refers to a unique approach where the number of clusters and the cluster centers in a given data set are defined and managed by domain experts (the practitioners) that are following a specific objective in their analyses, instead of being found (or allocated) by an algorithm. In Supervised Fuzzy Cluster Analysis, the number of clusters and the cluster centers are assigned by domain experts in order to accomplish a certain objective. The objectives such as impact and influence of formation quality on well productivity will drive the Supervised Fuzzy Cluster Analysis by identifying the formation quality as a function of specific field measurements.

For example, in the context of Shale Descriptive Analytics, the number of clusters and the location of the cluster centers for each variable (parameter) are determined with the objective of identifying the impact and the influence of this specific category of field measurements on the well productivity. When this approach is applied to the reservoir characteristics in order to discover its impact and influence on the well productivity, following steps need to be taken by the practitioner. Reservoir Characteristics as a category of field measurements will include multiple parameters such as porosity, pay thickness, water saturation, clay content, and TOC. These parameters will be used in order to identify and compare the impact and the influence of Poor, Average, and Good formation qualities on well productivity. Furthermore, these analyses can help the practitioner to compare the impact and influence of formation quality on well productivity with the impact and influence of other categories of parameters on well productivity.

In Supervised Fuzzy Cluster Analysis, the range of each of the parameters (the five field measurements, porosity, pay thickness, water saturation, clay content, and TOC), are used in order to classify reservoir characteristics into poor, average, and good formations. For example, a portion of the formation where a specific well has been drilled in, is classified as a “Poor” formation as long as it has “Low” porosity, “Thin” pay thickness, “High” water saturation, “Large” clay content, and “Low” TOC. On the other hand, a part of the formation where a specific well has been drilled in, is classified as a “Good” formation as long as it has “High” porosity, “Thick” pay thickness, “Low” water saturation, “Small” clay content, and “High” TOC.
Figure 5 shows an example of the projection of formation qualities on two reservoir characteristics in order to classify “Poor”, “Average”, and “Good” formations. In this figure each of the small white points represent hydrocarbon saturation and net thickness for an individual well. The distances from each of the small white points from the cluster centers determine the degree of the membership of this particular well in the cluster of “Poor”, “Average, and “Good” formations. Now imagine a five dimensional graph of reservoir characteristics (porosity, pay thickness, water saturation, clay content, and TOC), then each well will have a certain membership in each of the three formation categories as a function of all five reservoir characteristics. These membership values determine the relative quality of the formation for the specific well as it is compared with all other wells in the field.

**APPLICATION OF SHALE DESCRIPTIVE ANALYTICS**

When Shale Descriptive Analytic is applied to a certain field it will identify the impact and the influence of each category of parameters as well as each individual parameter (field measurement) on the well productivity. The analyses that are presented in the next several figures represent the application of Shale Descriptive Analytic to a large number of shale wells in a well-known shale play in the state of Texas.

**Well Quality Analysis (WQA)**

In this section application of Shale Descriptive Analytics to a large number of wells in the state of Texas is demonstrated. The first step of this analysis that is known as Well Quality Analysis (WQA), well productivity index (180 Days Cum. Oil Production) is classified using fuzzy set theory. The fuzzy classification of the well productivity is performed using three different granularities. As shown in Figure 3 the well productivity is classified into three, four, and five different qualities. Each time, all the
parameters from all five categories of field measurements are applied to the membership of the well productivity indices as they are identified in each of the granular classifications.

Plotting the average value of the parameter being analyzed as a function of the well productivity index in each category of well quality, provides a view of any potential trend and patterns that correlates this particular parameter to the well productivity index. Figure 6 through Figure 10 demonstrate that application of this part of the Shale Descriptive Analytics to five of the parameters in this specific shale in Texas.

These figures (Figure 6 through Figure 10) include five plots. The three plots in the bottom represent the first part of Shale Descriptive Analytics with the categorization of the well productivity as shown in Figure 3 with the granularities of 3, 4, and 5 well qualities. Each of these bar charts include a single bar on the left with a gray background and then 3, 4, and 5 bars on the right of each plot. The bar with the gray background represent the overall average of the field measurement for the entire field (all the wells being analyzed) while the bars with the white background represent the average value of the given parameter (field measurement) for each specific quality of the well based on their productivity.

![Figure 6. Application of part one of the Shale Descriptive Analytics to Clay Volume.](image)

The two plots on the top of each figure show the Cartesian plot of the field measurement versus well productivity (the top plot on the left hand side with each red dot representing a single well), and the largest possible granular classification that results in a continuous trend (pattern) discovery of the behavior of the given field measurements as it correlates with the well productivity. Please note that the plot on the top right (blue curve) is not generated using traditional statistical algorithms such as regression or windowed-averaging of the data. As mentioned above, this plot represents the discovered trend from the actual data using the largest number of possible granularity that is demonstrated in the discrete fashion in the bottom three plots.

Figure 6 displays the Well Quality Analysis (WQA) performed for “Clay Volume” (formation characteristic) for hundreds of shale wells in a specific shale play in the state of Texas. The Cartesian plot (top-left) shows a general trend that is clearly identified by the optimum granularity of fuzzy pattern recognition that is shown with the continuous blue curve (top-right). The fuzzy pattern recognition with three well quality (poor, average, and good) that is shown in the bottom left, indicate a clear pattern that wells with poor productivity are located in the part of the field with the highest clay volume when they
are compared with well with average and good productivity. As the granularity increases to four well qualities (bottom-middle) and to five well qualities (bottom-right) the trend is maintained and more detail is revealed.
Influence of Different Parameters on Well Productivity

For the purpose of this article 18 different parameters (field measurements) divided into five different categories from a given shale asset in the west Texas were used for analyses. Following is the list of categories and the parameters (field measurements) that were used in this analysis:

- **Group 1: Formation Characteristics**
  - Porosity
  - Net Thickness
  - Hydrocarbon Saturation
  - TOC
  - Clay Volume

- **Group 2: Well Spacing and Stacking**
  - Spacing Distance
  - Spacing Vertical Separation
  - Stacking Distance
After performing Well Quality Analysis for all the 18 field measurements, examples of which were shown in Figure 6 through Figure 10, impact of each category of parameters were calculated. Figure 11 shows the comparison of the impact of each category of parameters on well productivity. This figure clearly shows that Well Spacing and Stacking is the most impactful and influential set of parameters that control the production and recovery in this asset. This mainly has to do with the fact that Frac-Hit has been influencing the production of the shale wells in this shale play in Texas more than any other set of parameters. Hydraulic Fracture implementation is the second most influential category of parameters that controls the well productivity in this asset followed by operational constraints, formation quality and finally the completion design. Figure 12 shows the Key Performance Indicators (KPI) for this asset at the individual parameters level. The well productivity index that was used to perform these analyses is 180 days cumulative oil production. Well Quality Analyses that were demonstrated in Figure 6 through Figure 10 were main foundation of generating the results that are shown in Figure 11 and Figure 12.
Based on the fact that Well Spacing and Stacking has shown to be the most influential category of parameters that controls well productivity, it would make sense to demonstrate how we have calculated these parameters. Figure 13 demonstrates how the parameters that represent well spacing and stacking are calculated for this study. Providing more information regarding the impact of Well Spacing and Stacking on well productivity, Shale Descriptive Analytics allows the practitioners to observe such impact in three dimensional graphs. Figure 14 shows how well productivity is impacted by different parameters of well spacing and stacking in three dimensions. The trend shown in this figure is discovered using fuzzy pattern recognition technology that are shown in the top-right-hand plot of Figure 6 through Figure 10. These trends were used in order to generate the behavior demonstrated in Figure 14.

The next step in Shale Descriptive Analytics is the analyses that demonstrate the influence of different qualities of multiple categories of parameters (field measurements) on well productivity in much more detail. In order to perform this analysis, first each of the categories of the parameters (field measurements) are divided into multiple qualitative classifications. For example, formation quality category is divided into three categories if (a) Good Quality Formation, (b) Average Quality Formation, and (c) Poor Quality Formation.
Figure 15 shows the five field measurements (Porosity, TOP, Net Thickness, Hydrocarbon Saturation, and Caly Voluem) that have been used in order to qualitatively categorize formation quality into three categories of Poor, Average, and Good formations. This figure shows that Good Quality Formation is identified by high maturity (has a higher value of TOC), higher Thickness, more hydrocarbon saturation, lower clay volume and higher porosity values when it is compared to average and poor formation qualities. Imagine a five dimensional graph as each of the dimensions represent one of the five field measurements.
measurements that are shown in Figure 15. Then each of the formation quality classes (clusters), having
the values that are shown in this figure represent a point in this five dimensional graph, representing the
center of the soft cluster. In order to show some details three separate two-dimensional projection of the
five dimensional formation quality are shown in Figure 16.

Figure 15. Five different field measurements are used in order to qualitatively classify the formation quality in this shale asset.

Figure 16. Projection of the five dimensional formation quality soft clustering on three different two-dimensional plots.

In this figure, examples of fuzzy cluster membership for several wells are demonstrated. At the end of the
process, each well in this field has a membership in all three categories of formation quality. Then, the formation quality of each well is correlated with the quality of the well productivity in order to identify any potential correlation that exist between formation quality and well productivity.

Figure 17. Five different field measurements are used in order to qualitatively classify the Completion Design quality in this shale asset.

Figure 18. Projection of the five dimensional completion design quality soft clustering on a two-dimensional plot.

Figure 17 shows the five field measurements (lateral length, stage length, number of clusters per stage, cluster spacing, and shot density) that have been used in order to qualitatively categorize completion
design quality into two categories of small and large completion designs. This figure shows that Large completion design is identified by lengthy laterals, short stage length, large number of clusters per stage, small cluster spacing, and high density of perforations. Imagine a five dimensional graph as each of the dimensions represent one of the five field measurements that are shown in Figure 17.

Then each of the completion design quality classes (clusters), having the values that are shown in this figure represent a point in this five dimensional graph, representing the center of the soft cluster. In order to show some details, a two-dimensional projection of the five dimensional completion design quality is shown in Figure 18. In this figure, examples of fuzzy cluster membership for couple of wells are demonstrated. At the end of the process, each well in this field has a membership in the both two categories of completion design qualities. Then, the completion design quality of each well is correlated with the quality of the well productivity in order to identify any potential correlation that exist between completion design quality and well productivity.

Figure 19. Three different field measurements are used in order to qualitatively classify the Hydraulic Fracture implementation quality in this shale asset.

Figure 19 shows the three field measurements (job size [lbs/ft], amount of cross-linked gel used [%], and proppant concentration [lbs/gal]) that have been used in order to qualitatively categorize hydraulic fracture implementation quality into two categories of small (with high cross-linked gel and high proppant concentration) and large (with low cross-linked gel and low proppant concentration) hydraulic fracture implementation. Please note that when Shale Descriptive Analytics is being performed in a new field, the practitioner gets to choose the parameters that represent each category, as well as number and quality of the clusters for the analysis. Figure 19 shows that Large hydraulic fracture implementation is identified by large job sizes, small amounts of cross-linked gell and small proppant concentrations. Imagine a three dimensional graph as each of the dimensions represent one of the three field measurements that are shown in Figure 19.
Then each of the hydraulic fracture implementation quality classes (clusters), having the values that are shown in this figure represent a point in this three dimensional graph, representing the center of the soft cluster. In order to show some details, a two-dimensional projection of the three dimensional hydraulic fracture implementation quality is shown in Figure 20. In this figure, examples of fuzzy cluster membership for couple of wells are demonstrated. At the end of the process, each well in this field has a membership in the both two categories of hydraulic fracture implementation qualities. Then, the hydraulic fracture implementation quality of each well is correlated with the quality of the well productivity in order to identify any potential correlation that exist between hydraulic fracture implementation quality and well productivity.

![Figure 20. Projection of the three dimensional hydraulic fracturing implantation quality soft clustering on a two-dimensional plot.](image)

Figure 21 shows the four field measurements that have been used in order to qualitatively categorize well spacing and stacking quality into two categories of short and long well spacing and stacking. As shown in Figure 13, the field measurements used for this clustering are space [distance] and vertical distance between the parent and child wells located in the same formation as well as space [distance] and vertical distance between the parent and child wells located in the different (stacked) formations. Figure 21 shows that long well spacing and stacking is identified by large distance and vertical spacing between all the four categories while short well spacing and stacking is identified by small distance and vertical spacing between all the four categories. Imagine a four dimensional graph as each of the dimensions represent one of the four field measurements that are shown in Figure 21.
Figure 21. Four different field measurements are used in order to qualitatively classify the well spacing and stacking quality in this shale asset.

Figure 22. Projection of the four dimensional well spacing and stacking quality soft clustering on a two-dimensional plot.

Then each of the well spacing and stacking quality classes (clusters), having the values that are shown in this figure represent a point in this four dimensional graph, representing the center of the soft cluster. In order to show some details, a two-dimensional projection of the four dimensional well spacing and
stacking quality is shown in Figure 22. In this figure, examples of fuzzy cluster membership for couple of wells are demonstrated. At the end of the process, each well in this field has a membership in the both two categories of well spacing and stacking qualities. Then, the well spacing and stacking quality of each well is correlated with the quality of the well productivity in order to identify any potential correlation that exist between hydraulic fracture implementation quality and well productivity.

**RESULTS AND DISCUSSIONS**

In the previous section, four different categories of parameters (formation, completion, hydraulic fracturing, and well spacing and staking) were classified into two or three classes. Next, soft cluster algorithm is applied to each of the shale wells in the field in order to indentify the well quality based on its productivity index. The final step of the Shale Descriptive Analytics will be to analyze the quality of formation, completion design, hydraulic fracturing implementation, and well spacing and stacking in order to see if they correlate with well productivity. The objective of this analysis is to find out if there are patterns and trends that can be discovered in field measurements that correlate well productivity to:

a. The quality of the formation (reservoir) where the well has been drilled and completed,

b. The design of the completion that was applied to the well,

c. The implementation of the hydraulic fracturing, and finally

d. The distances of the well (child well) from the parent wells in the same formation and the stacked formation.

To define the well productivity in the same manner that other four category of parameters are classified each shale well in the field is classified based on its 180 days cumulative oil production as its productivity index. Figure 23 shows how wells in this field are classified as “Poor”, “Average”, and “Good” wells. Each well can have full membership in one of the classes or have membership in two classes, based on its productivity index.

![Figure 23. Using fuzzy set theory to classify the well quality based on the productivity index (180 days’ cumulative oil production).](image)

The next step is to cross-mach the productivity of each individual shale well in this particular asset in Texas, with the quality of the four category of parameters. Figure 24 summarizes the application of this Shale Descriptive Analytics algorithm, while the more comprehensive version of the results of this analysis is shown in Figure 25. As it can be seen in this figure, when it come to the shale wells with “Poor” productivity, 53% of them have been drilled and completed in the portion of the reservoir that has “Poor”
formation quality. 35% of the “Poor” wells have been drilled and completed in the portion of the reservoir that has “Average” formation quality, and only 12% of them have been drilled and completed in the portion of the reservoir that has “Good” formation quality. Furthermore, when it comes to the shale wells with “Average” productivity, a very tiny percent of them (1%) have been drilled and completed in the “Poor” formation quality. 29% of the “Average” wells have been drilled and completed in the “Average” formation quality, and 70% of them have been drilled and completed in the “Good” formation quality.

Finally, the shale wells with “Good” productivity have never been drilled and completed in the “Poor” formation quality, while 28% have been drilled and completed in the “Average” formation quality, and 73% of them have been drilled and completed in the “Good” formation quality. All the results shown in this analyses, makes perfect sence. It emphasizes the contribution of formation quality to the well productivity. However, it must be noted that the trend (pattern) that has been discovered and the valuable information that has been released are from a set of highly chaotic data (field measuremets). In other words, no conventional technique has been able to extract such information from the data that has been gathered from this shale reservoir in Texas.

Next in this figure, the category of parameters that represent Completion Design is analyzed. Figure 24 shows that 100% of the shale wells with “Poor” productivity have had a “Small” completion design, while none of them had “Large” completion design. Furthermore, it shows that 58% of the shale wells with “Average” productivity have had a “Small” completion design, while 42% had “Large” completion design. Finally, only 3% of the shale wells with “Good” productivity have had a “Small” completion design, while 95% of the well with “Good” productivity had “Large” completion design. These results make it obvious that the Large Completion designs have considerable contribution to better well productivity.

Hydraulic fracturing implementation is the next category being analyzed. Figure 24 shows that 79% of the shale wells with “Poor” productivity have been fractured using “Small” job sizes with high volumes of cross-linked gel and high proppant concentration, while 21% of them have been fractured using “Large” job sizes with low cross-linked gel volumes and low proppant concentration. Furthermore, it shows that 34% of the shale wells with “Average” productivity have been fractured using “Small” job sizes with high volume of cross-linked gel and high proppant concentration, while 66% of them have been fractured using “Large” job sizes with low cross-linked gel volumes and low proppant concentration. Finally, 5% of the shale wells with “Good” productivity have been fractured using “Small” job sizes with high volumes of cross-linked gel and high proppant concentration, while 95% of them have been fractured using “Large” job sizes with low cross-linked gel volumes and low proppant concentration. These results make it clear that the “Large” job sizes with low cross-linked gel volumes and low proppant concentration have considerable contribution to better well productivity.
The last category of parameters being analyzed is Well Spacing and Stacking. Figure 24 shows that 63% of the shale wells with “Poor” productivity have “Large” well spacing and stacking, while only 37% of them have “Small” well spacing and stacking. Furthermore, it shows that 45% of the shale wells with “Average” productivity have “Large” well spacing and stacking, while 55% of them have “Small” well spacing and stacking. Finally, 35% of the shale wells with “Good” productivity have “Large” well spacing and stacking, while only 65% of them have “Small” well spacing and stacking.

### Table 3: Impact of Formation Quality, Completion Design, and Frac Design on Well Productivity

<table>
<thead>
<tr>
<th>Formation Quality</th>
<th>Poor Wells</th>
<th>Poor/Average Wells</th>
<th>Average Wells</th>
<th>Average/Good Wells</th>
<th>Good Wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Poor Formation</td>
<td>53%</td>
<td>20%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>Average Formation</td>
<td>35%</td>
<td>29%</td>
<td>29%</td>
<td>29%</td>
<td>28%</td>
</tr>
<tr>
<td>Good Formation</td>
<td>12%</td>
<td>52%</td>
<td>70%</td>
<td>71%</td>
<td>73%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Completion Design</th>
<th>Poor Wells</th>
<th>Poor/Average Wells</th>
<th>Average Wells</th>
<th>Average/Good Wells</th>
<th>Good Wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Completion</td>
<td>100%</td>
<td>92%</td>
<td>58%</td>
<td>36%</td>
<td>3%</td>
</tr>
<tr>
<td>Large Completion</td>
<td>0%</td>
<td>8%</td>
<td>42%</td>
<td>64%</td>
<td>98%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Frac Design</th>
<th>Poor Wells</th>
<th>Poor/Average Wells</th>
<th>Average Wells</th>
<th>Average/Good Wells</th>
<th>Good Wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small Frac - High (xg:pcn)</td>
<td>79%</td>
<td>74%</td>
<td>34%</td>
<td>21%</td>
<td>5%</td>
</tr>
<tr>
<td>Large Frac - Low (xg:pcn)</td>
<td>21%</td>
<td>26%</td>
<td>66%</td>
<td>79%</td>
<td>95%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Well Spacing &amp; Stacking</th>
<th>Poor Wells</th>
<th>Poor/Average Wells</th>
<th>Average Wells</th>
<th>Average/Good Wells</th>
<th>Good Wells</th>
</tr>
</thead>
<tbody>
<tr>
<td>Large Spacing &amp; Stacking</td>
<td>63%</td>
<td>56%</td>
<td>45%</td>
<td>47%</td>
<td>35%</td>
</tr>
<tr>
<td>Small Spacing &amp; Stacking</td>
<td>37%</td>
<td>44%</td>
<td>55%</td>
<td>53%</td>
<td>65%</td>
</tr>
</tbody>
</table>

Figure 25. The more comprehensive version of the results of the cross-matching well productivity with formation, well spacing and stacking, completion design and hydraulic fracturing implementation quality.

### CONCLUSIONS

The results shown in Figure 24 and Figure 25 clearly demonstrate the impact and the influence of the formation quality, completion design, and hydraulic fracturing implementation on the productivity of shale well in this particular asset in state of Texas. Based on these results it is obvious that the poor wells have been mainly drilled and completed in poor to average quality parts of the shale play, have been completed with small designs and small frac jobs that have included high volumes of cross-linked gel and high proppant concentration. While the impact of poor formation quality, small completion design and small frac job size may sound quite reasonable for wells with poor productivity, it should be interesting to learn that the impact of high volumes of cross-linked gel and high proppant concentration have made wells to produce poorly even when they are drilled and completed in the average quality reservoirs.

Another interesting finding of this analysis has to do with the impact and the influence of the well spacing and stacking on well productivity. While well spacing and stacking has shown to be the most influential category of parameters in this shale asset (as shown in Figure 11), the results of the impact of this group of parameters are hard to be interpreted as it is shown in Figure 24 and Figure 25. The trend for well spacing and stacking shown in Figure 24 and Figure 25 is not as clear and reasonable as the trends that are shown for other three categories of the parameters. What made this lack of discovery of reasonable trends and patterns (when it comes to well spacing) had to do with the fact that similar analyses had been performed for a large number of shale assets throughout the United States and such behavior regarding the well spacing was never encountered.

Upon completion of the analyses shown above, and in order to develop a better understanding of the reasons behind this specific behavior of well spacing and stacking, we went back to the data in order to see if any reason can found. Looking at the production profiles of a large percentage of the wells in this...
asset, it became obvious that the interferences between parent and child wells in this field (the phenomenon known as Frac-Hit) is highly impacting the production and recovery of the hydrocarbon from this field. Figure 11 clearly shows that well spacing and stacking is highly impacting well productivity in this asset, while the lack of clear trend in Figure 24 and Figure 25 shows that details of such impact is hard to be discovered, when the productivity index (a snap-shot in time) is used as the output of the analyses.

Therefore, the main conclusion of this analyses can be mentioned as follows:

- Once the number of wells in a shale asset increases, resulting in the shortening of the distances between parent and child wells; Frac-Hit becomes a major issue in hydrocarbon production from shale wells.
- Once Frac-Hit becomes an issue, using productivity index (a snapshot in time) as the model output will no longer be able to provide meaningful and useful modeling that would be able to predict, mitigate, and/or manage Frac-Hit.
- This is true even when Artificial Intelligence and Machine Learning are used as the main tool for analyses.
- The solutions to address this problem is the development of full field reservoir simulation that is capable of realistically modeling and history matching the entire production profile of every individual shale well in a given asset.

As a final note, it must be mentioned that traditional numerical simulation and modeling is not the correct tool for building such model for realistic and fact-based modeling and analyses of shale assets. This fact become even more true once the hydrocarbon production from shale wells are impacted by Frac-Hit.

**Summary of Shale Analytics**

Shale Analytics includes three major steps: Descriptive, Predictive and Prescriptive Analytics. In this Article Descriptive Analytics was covered. Shale Descriptive Analytics extracts much information in terms of patterns and trends form the collected data from a shale asset with multiple shale wells. The information that is extracted from the data during Shale Descriptive Analytics builds the foundation for the next step that is Shale Predictive Analytics (Mohaghegh 2017c). Multiple data-driven models are trained, calibrated and validated via blind wells during the Shale Predictive Analytics. Characteristics of Shale Predictive Analytics includes validation of the predictive data-driven models through a long and complex process.

The validation process includes predicting the productivity of a series of wells that have not been used during the training and calibration of the data-driven models, also known as blind wells. The key to the validation of the predictive models in the context of oil and gas industry is developing a series of Type Curves for individual wells, group of wells, and the full field that are well-behaved and honor the known physics of fluid flow in the porous media. Furthermore, the predictive model must have the ability to explain its predictions on a well-by-well basis. Once Shale Predictive Analytics is completed and the
practitioners develop confidence on its predictive capabilities, the data-driven models developed during the Shale Predictive Analytics are used in an inverse fashion for the purposes of Shale Prescriptive Analytics.

Shale Prescriptive Analytics reverse engineers the validated data-driven predictive models in order to assist management in making key decisions. During the Shale Prescriptive Analytics the management can look-back at all the prior completion and hydraulic fracturing implementations in order to find out how well those decisions were made and how they can be improved. The completion design and the hydraulic fracturing implementations can be modified in order to maximize production and recovery from every individual shale well in the asset (completion optimization).

Using the Shale Prescriptive Analytics the management of the operating companies can compare the contribution of multiple different service companies in implementing completion and hydraulic fracturing. The operating company’s management will be able to identify the best, most effective service company that has been providing services. Finally, Shale Prescriptive Analytics provides the tools for the operating company’s management to make fast and accurate financial decisions on how to operate the shale asset in order to maximize company’s profit margin.

**NEXT STEP**

Application of Artificial Intelligence and Machine Learning in the upstream oil and gas industry goes back to early 1990s. Continuous and serious research and development associate with Artificial Intelligence and Machine Learning in the oil and gas industry was limited to only one university and one or two companies until a few years ago. In the past several years, a large number of startup companies started using the application of this technology in the oil and gas industry. When it comes to the application of Artificial Intelligence and Machine Learning to production of hydrocarbon from shale wells, all practitioners have one thing in common: They use well productivity index as their model output.

In other words, all data-driven activities that have something to do with predicting hydrocarbon production from shale, use well productivity indices as the source of all their correlations. This is because of the fracture-driven interaction between wells (also known as Frac-Hit). Frac-Hit takes over the overall productivity of all the wells in the field. Frac-Hit not only negatively influences parent and child wells productivity and recovery, it completely undermines all the existing techniques (traditional techniques such as RTA and numerical simulation as well as all the existing techniques based on Data Analytics) for completion and production optimization of shale wells.

Our research and development in the past decade on the application of Artificial Intelligence and Machine Learning to production of hydrocarbon from shale wells have resulted in one major finding. When Frac-Hit becomes an issue in shale assets, using well productivity index (that is a snapshot in time when it comes to well productivity) will no longer contribute to any type of predictive and prescriptive analytics. We have learned that in order to overcome this shortcoming, a detail, and dynamic reservoir simulation and modeling process that attempts to model the entire production profile of every individual well in a given shale asset must be developed. A new technology that will address this shortcoming is currently under development using Artificial Intelligence and Machine Learning. This new technology is called Dynamic Shale Analytics. Dynamic Shale Analytics is a completely data-driven, full field, reservoir...
simulation and modeling process that will use data-driven reservoir modeling (Mohaghegh 2017b) as its main source. Dynamic Shale Analytics history matches the entire production profile of every shale well in an asset (there are no limitations in the number of wells that can be used in this modeling process) in order to be able to predict, mitigate, and manage Frac-Hit.

REFERENCES


Mohaghegh, 2015; Formation vs. Completion: Determining the Main Drivers behind Production from Shale; A Case Study Using Data-Driven Analytics. Mohaghegh, S.D., West Virginia University & Intelligent Solutions, Inc. URTeC 2147904, Unconventional Resources Technology Conference. San Antonio, Texas, USA, 20-22 July 2015.

