Using Data-Driven Analytics to Assess the Impact of Design Parameters on Production from Shale

Esmaili, S., West Virginia University and Mohaghegh, S.D., Intelligent Solution Inc. and West Virginia University

Abstract

The importance of production from Shale and its impact on the total US energy equation has focused much attention on this prolific source of hydrocarbon. Consequently, research related to unconventional reservoirs has increased significantly in order to better understand the inherent complexities of their behavior.

Analytical, numerical and statistical analyses have been applied to large multi-variable data set from Shale assets with different degrees of success. The notion that shale is a “statistical play” may be attributed to the fact that many of our preconceived notions on storage and flow mechanisms in shale are not supported by facts. Therefore, we set out to examine the possibility of learning from the data in order to be able to answer some of the questions that rise during the production process. Data Driven Analytics, having roots in pattern recognition and machine learning, have proven to be capable of extracting useful information from large data sets and are extensively used in many industries. Their application to multi-variable data sets from Shale assets, in order to extract understandable structure in the data, is the subject of the work being presented here.

This paper presents a Data Driven Analytics study of design parameters such as well trajectories, completion, and hydraulic fracturing variables for a large number of horizontal wells in Marcellus Shale. The data set from the Shale assets is so complex that use of conventional statistical analysis does not results in understandable trends and patterns. On the other hand, when advanced pattern recognition tools are used, certain (previously hidden) patterns emerges from the data with unmistakable trends.

In this article impact of parameters such as up-dip versus down dip deviation of wells, stimulated lateral length and cluster spacing, etc. on production from wells in a Shale asset is analyzed using an advance pattern recognition algorithm. The analyses are performed using production from multiple time intervals throughout the life of wells.

Introduction

Data Mining is the practice of searching through a database in order to discover useful patterns or relationships among the data. It employs artificial intelligence, machine learning, pattern recognition, statistical methods to discover and present knowledge in a form which is easily understandable to humans. Data Mining has been widely used in different areas such as financial, communication and marketing organizations, where enables them to determine relationships among internal factors and external factors as well as the investigating the impact of different parameters on external factors such as economic indicators and etc.

Data Mining can be applied to data sets of any size. However, while it can be used to uncover hidden patterns in data that has been collected, obviously it can neither uncover patterns which are not already present in the data, nor can it uncover patterns in data that has not been collected.

The interest in Data Mining in petroleum engineering area dates back to the early 90s where the companies have started to
realize the value of this technique for identification of best practices in their operations. Despite having substantial data in the oil and gas industry, data mining has been utilized to optimize solutions to many issues and challenges in this industry such as identifying key drivers in well productivity, maximizing the recovery factors, investigating the best practices in depletion strategy, optimizing the reservoir performance and etc.

Due to the major improvements in shale reservoir development and the imminent need for understanding their complex behavior in order to be able to predict the future performance, data mining has also attracted a great deal of attention in shale assets industry in recent years. Given all the facts about complexity of the shale reservoirs, the physics of production from these reservoirs are not fully understood and therefore making decisions and reservoir management based on the available tools (such as analytical and numerical) could be challenging. Therefore data mining may be a good alternative which gives a hand to the engineers and operators to make better decision when it comes to the shale assets.

Despite the vibrant and fast growing literature related to application of data mining in shale assets, it has also frequently been showed that the traditional and conventional statistical analysis may not be helpful in understanding the complex behavior of shale and using such analysis to make business decision is not feasible and most of the researchers attempted to apply some advanced data mining techniques. For instance, Shelley et al (2008) found that the large number of variables and scatter within the production data in Barnett shale required the use of data clustering techniques to extract useful information. Advanced data mining techniques were used including self-organizing maps (SOM) to reduce the statistical noise and highlights completion and reservoir parameters directly affecting production.

In the context of data mining in shale assets, Lafollette et al (2011) performed a data mining study on Barnett shale by using Geographical Information System (GIS) pattern recognition techniques in conjunction with more traditional statistical techniques to interpret hidden trends in otherwise scattered data sets. In this work they also realized that the conventional cross-plots of peak gas as a function of different parameters may not show an acceptable regression coefficient while the geographic grouping of wells may possibly lead to more homogenous groups of wells when attempting to define which parameters may be changed to increase well productivity. In similar work done by the same authors (Lafollette et al, 2012), the boosted tree models have been applied in order to evaluate the impact of completion and stimulation parameters on the production from Barnett wells. In this study it was also shown that the impact of individual variables on the production outcome is often difficult to interpret with any degree of confidence when traditional linear regression methods are used because of the impacts of missing data, erroneous data, non-linear data and subtle interrelationships among variables.

Given the complexities associated with the production behavior in shale assets it should be evident that conventional statistics would have little to offer in analysis and modeling the data that represent such behavior. Nevertheless, the first part of this article is dedicated to the conventional statistical analysis in order to demonstrate that the behavior of shale (here Marcellus Shale) is too complex to be analyzed and possibly modeled using conventional statistics. The conclusion will be the necessity of more complex types of analyses that makes use of sophisticated technologies in order to deduce patterns and trends from the seemingly chaotic data.

The advanced data-driven analytics and data mining technology being used here is called “Fuzzy Pattern Recognition” since it is based on Fuzzy Set Theory” (Mohaghegh, 2000). This technology is applied to large number of horizontal wells in Marcellus shale in order to evaluate the hydraulic fracturing practices in this asset. This collection of algorithms do not predict a target value, but focus more on the intrinsic structure, relations and interconnectedness of the data. In other words, during the implementation of this technology, data is not manipulated or modified in any shape or form; rather it is presented in a new light that makes recognition of existing trends and patterns possible.

**Database Preparation**

The dataset presented here is gathered from multiple operators. It includes more than 160 horizontal wells (from more than 60 pads) completed in Marcellus shale. These wells contain approximately 2000 stages or 6000 clusters of hydraulic fractures. The dataset includes detail completion and stimulation information such as injection pressure, injection rate, volumes of fluid and proppant as well as total number of stages/clusters, perforated intervals and stimulated lateral length.

Since the production is available on a per well basis, the volumes of fluid and proppant for multiple hydraulic fracture stages performed on the same well were summed while the rates and pressures for these cases were averaged. In addition, matrix porosity and net thickness (derived from logs) were available on a “per well” basis and were used to normalize production from each well to “per foot of thickness” basis.
The impact of completion and stimulation parameters were studied against the production indicators such as the best 3 months and the best 12 months of cumulative rich gas production. Figure 1 shows the frequency of the wells with the best 3 months and 12 months of cumulative rich gas production in this dataset.

Figure 1: Frequency of the wells with best 3 months and best 12 months of cumulative rich gas production

Figure 2: Cross plot of best 3 months of cumulative rich gas production versus total injected proppant for different deviation types and BTU regions
Conventional Statistical Analysis

Due to highly complex nature of the production from hydraulically fractured shale wells in Marcellus, the linear and instinctive behavior should not be expected. In this section, the complexity of the system will be shown by using conventional cross plots of various parameters versus the best 3 months and 12 months of cumulative rich gas production.

The first plot is the cross plot of best 3 months cumulative rich gas production versus total injected slurry (bbl.) for different well deviation types (Down-Dip, No-Dip and Up-Dip) and for different BTU regions (Figure 2). Based on the content of condensate, the Marcellus shale area was divided into four different BTU areas varying from Dry (low BTU) to wet (high BTU) regions. Note that the production result versus total injected slurry doesn’t show any apparent trend when they are clustered based on either the deviation type and/or BTU regions.

Figure 3 is demonstrating the scatter plot of best 3 months cumulative rich gas production which is normalized based on the porosity * thickness (%-ft.) versus total injected proppant for different well deviation types and different BTU regions in a logarithmic scale. The broad spread of production for all types of deviation in this plot clearly shows the lack of correlation.

Figure 3: Cross plot of Normalized 3 Months cumulative rich gas production per porosity * thickness versus total injected proppant in Semi-log scale for different deviation types and BTU regions.
This spread behavior of data points can be observed consistently through the entire cross plots made, illustrating in Figure 4 in which the normalized 3 months cumulative rich gas (based on stimulated lateral length and total clusters) was plotted as a function of porosity*thickness for various well deviation types and BTU regions.

As shown in Figure 4, the correlation between the factors that typically contribute to production performance is weak even though the significance becomes more obvious when the wells are clustered based on different deviation types and different BTU regions. There are weak correlations between the best 3 months cumulative rich gas production and different reservoir characteristics such as porosity and thickness as wells different stimulation parameters such as injected proppant and injected slurry volume. Since a clear trend cannot be observed using conventional statistical technique production behavior of the wells, the advanced data mining techniques should be used in order to analyze such a complex behavior.

![Cross plot of Normalized 3 months cumulative rich gas production per Stimulated Lateral Length per Total Clusters vs. Porosity*Thickness (%-ft) for Different Deviation Types](image1)

![Cross plot of Normalized 3 months cumulative rich gas production per Stimulated Lateral Length per Total Clusters vs. Porosity*Thickness (%-ft) for Different BTU Regions](image2)

**Figure 4**: Cross plot of Normalized 3 months cumulative rich gas production based on the stimulated lateral length and total clusters versus porosity*thickness for different deviation types and BTU regions

**Fuzzy Pattern Recognition Analysis**

The result of conventional statistical analysis showed that production from shale is too complex to be meaningfully analyzed using this technique. To address these complexities advanced fuzzy pattern recognition technology is used in order to disclose any hidden pattern in dataset.

Pattern Recognition is a branch of artificial intelligence concerned with classification or description of observations. Pattern Recognition aims to classify data (patterns) based on either a priori knowledge or on statistical information extracted from the pattern. The patterns to be classified are usually groups of measurements or observations, defining points in an appropriate multidimensional space. When fuzzy set theory is used to determine the appropriate multidimensional space that
would provide optimum separation of overlapping classes, the result is known as “Fuzzy Pattern Recognition”. When Fuzzy Pattern Recognition is applied to a limited number of classes of wells the process is called the “Step Analysis” or “Well Quality Analysis (WQA)”. When a similar analysis is performed while every single well in the dataset is treated as a potential unique well quality the result is a continuous curve (rather than a discrete set of steps), called a “Fuzzy Trend Analysis (FTA)”.

It is important to note that the result of “Fuzzy Trend Analysis” is usually a non-linear two-dimensional line, as is shown in the following figures in this paper. An extension of this algorithm exist that can provide multi-dimensional versions of these trends. Following figures show a three dimensional view of the “Fuzzy Trend Analysis”. In these figures a well behaved trend on of the production indicator (best 3 months) is shown as a function of two other parameters (Left graph in Figure 5 and Figure 6). Please compare these plots with those plots in right side of these two figures where the actual data is presented. Note that in Figure 5, the best 3 months of cumulative rich gas production is plotted as a function of porosity and total injected proppant while in Figure 6, the production indicator is plotted as a function of total clusters and total injected slurry.

Figure 5: Three dimensional fuzzy trend analysis for best 3 months production of rich gas as a function of porosity and total injected proppant (left graph)- three dimensional plot of actual 3 months production of rich gas as a function of porosity and total injected proppant (right graph)

Figure 6: Three dimensional fuzzy trend analysis for best 3 months production of rich gas as a function of total clusters and total injected slurry (left graph)- three dimensional plot of actual 3 months production of rich gas as a function of total clusters and total injected slurry (right graph)

In “Well Quality Analysis”, three different well qualities, identified as Poor, Average and Good wells were defined based on the production indicators which are the best 3 months and the best 12 months of cumulative rich gas production.

Figure 7 and Figure 8 shows the fuzzy sets that were defined for each of the time intervals. As shown in these figures, there are several regions that wells with different qualities overlap and because of this definition, some of the wells have membership in more than one class of wells. These class memberships are referred to as fuzzy membership function and it is

1 The name reflects the shape of the resulted graphs that are in the form of ascending or descending steps.
2 Multiple wells with similar production indicator will be classified similarly.
3 To protect the confidentiality of the information that is presented in this paper, the production values have been modified. While the numbers that are seen as production values in this paper are fictitious, the trends and relative values have been preserved.
a very unique feature of this technology. For example in Figure 8, a well belongs to fuzzy set of “Poor Wells” with a degree of membership equal to 0.39 and to fuzzy set of “Average Wells” with a degree of membership equal to 0.61.

To check the validity of the process, first the best 3 months and then the best 12 month cumulative rich gas production are plotted. Figure 9 clearly demonstrates a trend in the production indicators. This is to check and see if the intended classification is indeed honored. The information presented in this figure (left) shows that the average best 3 months of production is a bit less than 250MM scf. This is indicated with the first bar (red background). The three bars (yellow background) represent the average best 3 months of production for Poor, Average and Good Wells. These values are 125MM, 275MM, and 475MM, respectively. The bar graph on the right provides similar check for the best 12 month cumulative rich gas production. The average for the entire data set is 650MM. The three bars (yellow background) represent the average best 12 months of production for Poor, Average and Good Wells. These values are 380MM, 700MM, and 1.3 BCF, respectively.

Figure 7: Fuzzy Clusters plots and values for the best 3 months of cumulative rich gas production

Figure 8: Fuzzy Clusters plots and values for the best 12 months of cumulative rich gas production

Figure 9: Well Quality Analysis for the best 3 months and the best 12 months of cumulative rich gas production
Once all wells are defined using this qualitative definition, the design parameters of the dataset is plotted based on the fuzzy membership function of the well qualities in order to see if any specific trend can be observed. The detailed well quality analysis results and fuzzy trend analysis results for each design parameters involved in this study is discussed as following:

**Well Location** - By looking at the production performance versus location of the wells (Easting) in this area (Figure 10 and Figure 11), it clearly shows a preferential trend towards gas production in both time intervals, as we move to the east part. Many factors might play role in observing such trend for production behavior towards the east which increasing the pay thickness is one of them. In 2004 D.G. Hill presented as isopach map for the Marcellus shale which shows the thickness of reservoir becomes thicker to the east.

Note that in Figure 10, the grey dots refer to the y-axis in the right hand side of the graph. These represent the actual rich gas data while the purple points (line) represent the discovered hidden trend line that exist in the actual data, but cannot be deciphered without the use of this particular algorithm. As mentioned before (Esmaili 2013) this algorithm that is called “Fuzzy Trend Analysis” is used in order to deduce understandable trends form seemingly chaotic behavior. In this analyses each of the parameters (for example the Easting-End [longitude] in Figure 10) is coupled with the production indicator (Best 3 months [left] and best 12 month [right]) and then is divided into an optimum number of two-dimensional fuzzy classes (the optimum number is identified using an optimization routine) that together will help unravel the hidden trends and patterns that exists in the raw data. The result of such analysis is the continuous curves rather than a discrete set of steps.

Similar to Figure 9, in Figure 11, the left plot which is the best 3 months of cumulative rich gas production, the first bar in left with red background shows the average location (Easting) of the wells in dataset which is about 570,000 m. The three bars (yellow background) represent the average location (Easting) for Poor, Average and Good Wells. These values are 560K, 572K, and 594K, respectively. The bar graph on the right provides similar check for the best 12 month cumulative rich gas production. The average location (Easting) for the entire data set is 564K while the three bars represent the average location for Poor, Average and Good Wells which the values are 559K, 564K and 577K, respectively.

![Figure 10: Fuzzy Trend Analysis for Easting in two time intervals of 3 months and 12 months](image)

![Figure 11: Well Quality Analysis for Easting in two time intervals of 3 months and 12 months](image)
**Well Deviation Type** - In this area of study, wells was drilled with different type of configuration and deviation, as it was discussed in chapter five, mainly three types of deviation, identified as Up-Dip, Down-Dip and No Dip can be observed. The well quality analysis (Figure 12) shows that in this part of Marcellus shale, the well performance in both early and late time has been impacted by the well deviation favoring the Down-Dip wells.

The Down-Dip wells are either completely or partially hydraulically fractured in Lower Marcellus which is more prolific in terms of reservoir characteristics and net thickness. Note that Marcellus shale is usually subdivided into tow subunits; Upper Marcellus and Lower Marcellus which are separated by a thin bed limestone unit know as Purcell. The Lower Marcellus is thicker and contains the higher levels of TOC and higher gas saturation.

![Figure 12: Well Quality Analysis for Deviation Type in two time intervals of 3 months and 12 months](image)

**Stimulated Lateral Length** - Drilling the horizontal wells with long lateral length became a successful strategy in production from shale and Marcellus shale is not an exception.

Figure 13 and Figure 14 and show the trend of production behavior of the field in different time intervals as a function of stimulated lateral length. These figures present that those wells with longer stimulated lateral show the better performance in either early or late time of production.

![Figure 13: Fuzzy Trend Analysis Stimulated Lateral Length in two time intervals of 3 months and 12 months](image)

**Total Number of Clusters** - the common practice in developing the Marcellus shale is hydraulic fracturing the wells with multi stages multi clusters. Number of clusters in this dataset changes widely, some wells have experienced up to 100 clusters while there some wells with number of clusters as low as 8.

Figure 15 shows the increasing trend of production (for both time intervals) as a function of total number of clusters. This behavior is also observed in Figure 16 where the wells were categorized into fuzzy clusters.
Figure 14: Well Quality Analysis for Stimulated Lateral Length in two time intervals of 3 months and 12 months

The FTA curves in Figure 15 for the best 3 months cumulative rich gas production shows a non-linearly increasing pattern in the production performance with the total number of clusters that starts with the sharp slope and decreases in steepness as the number of total clusters increases beyond 65. Note that few wells in this study do not have enough production history (less than 12 months) and therefore the number of wells in time interval of 12 months is less than the number of wells in time interval of 3 months. Because of this the scale of fuzzy trend curves (the range of the parameter) in FTA analysis might be different for these two time intervals as shown in Figure 15.

Figure 15: Fuzzy Trend Analysis for Total number of clusters in two time intervals of 3 months and 12 months

Cluster Spacing- Increasing the number of clusters and stimulated lateral length is always in favor of more rich gas production (as it shown), but inadequate cluster spacing can actually lead to lower ultimate recovery. Cheng (2012) showed that decreasing the cluster spacing so as to increase the total number of fractures may significantly reduce gas production when the cluster spacing is reduced to an adequately small size, where the width growth of fractures is strongly inhibited
because of the mechanical interaction.

Figure 17 and Figure 18 illustrate the result of fuzzy trend analysis and well quality analysis for cluster spacing in both time intervals. From these figures, it is observed that those wells with closer clusters (cluster spacing around 75 feet) have significantly better production performance than those with farthest clusters.

![Figure 17: Fuzzy Trend Analysis for Cluster Spacing in two time intervals of 3 months and 12 months](image)

![Figure 18: Well Quality Analysis for Cluster Spacing in two time intervals of 3 months and 12 months](image)

**Minimum Distance to the Offset Well** - Most of the laterals in Marcellus shale are spaced from 1000 ft. to 2000 ft. apart even though some operators experimented drilling with decreasing spacing (around 500 ft.) (Figure 19).

Figure 19 and Figure 20 show the fuzzy trend analysis and well quality analysis results for the impact of minimum distance to the offset well on production behavior of the wells in different time intervals. As illustrated in these figures, the production performance of the wells decreases by increasing the minimum distance between the laterals.

The closer distance between the laterals might be translated to more so-called stimulated reservoir volume because of more interaction between the hydraulic fractures and existing natural fractures and as a result higher productivity.

It should be noted that increasing the minimum distances between lateral does not necessarily increase the well performance as well produces more, therefore there should be an optimum distance to first establish enough connectivity to the matrix and second avoid well interference.

Moreover, some operators prefer to drill wells closer and compensate the production lose (by having the overlap in stimulated reservoir volume) by having longer stimulated lateral and more clusters.

In this dataset, there are some wells with about 100 clusters which are drilled very close to each other (minimum distance of 500 ft.) and they have very high rates (especially at the beginning) which categorized them as good wells with the other wells with higher minimum distance and less number of clusters that also shows high rates. This makes the conclusion of optimum minimum distance between the laterals, difficult. On the other hand, the effects of minimum distance between the laterals
may change by having longer production history (as the flow regime changes as a function of time) and may not be very evident during the initial first years of production (As shown in Figure 20, the average minimum distance for good wells in early time is different from the average minimum distance for these wells at later time; i.e. 12 months).

![Figure 19: Fuzzy Trend Analysis for Minimum Distance to the offset well in two time intervals of 3 months and 12 months](image)

**Figure 19**: Fuzzy Trend Analysis for Minimum Distance to the offset well in two time intervals of 3 months and 12 months

![Figure 20: Well Quality Analysis for Minimum Distance to the offset well in two time intervals of 3 months and 12 months](image)

**Figure 20**: Well Quality Analysis for Minimum Distance to the offset well in two time intervals of 3 months and 12 months

**Injected Slurry Volume**: The common practice in Marcellus is pumping water, sand and other additives under high pressure into the formation to create massive fracture networks. As shown in Figure 21 and Figure 22, a clear increasing trend can be observed in production performance by increasing the injected slurry volume.

![Figure 21: Fuzzy Trend Analysis Injected Slurry Volume in two time intervals of 3 months and 12 months](image)

**Figure 21**: Fuzzy Trend Analysis Injected Slurry Volume in two time intervals of 3 months and 12 months
Figure 22: Well Quality Analysis for Injected Slurry Volume in two time intervals of 3 months and 12 months

**Injected Proppant** - As illustrated in Figure 23 and Figure 24, the production performance of the wells has a direct correlation with the amount of proppant pumped in well. These figures show that better wells were treated with more proppant than wells with lower quality.

Figure 23: Fuzzy Trend Analysis for Injected Proppant in two time intervals of 3 months and 12 months

**Days between Completion and Production** - As discussed by Cheng (2012), the damages caused by injected water can be different in cases of having delay in bringing the well to production because in this case the injected water may reduce the formation permeability in the invaded zone because of clay swelling, scaling and fines migration [Holditch, 1979]. This process along with relative permeability and capillary pressure changes can potentially cause severely adverse impacts on long-term production.

Figure 25 clearly shows the inverse effect of delay in well performance where the good wells are those which came to
production with less delay after completion.

![Figure 25: Well Quality Analysis for Days between completion and production in two time intervals of 3 months and 12 months](image)

**Conclusions**

In this article, the production performance of the large number of wells in Marcellus shale was analyzed using state of the art in Advanced Data Mining. The final results of this study provides a tool to the operators for designing the optimum multi-stage frac jobs for planned wells based on the data-driven, non-biased completion and stimulation practices in Marcellus shale.

The following conclusions can also be drawn from the above analysis:

1. Drilling the new wells in eastern part of the reservoir can result in better performance of the well due to better reservoir quality (thickness, more open natural fractures)

2. The wells with deviation type of down-dip have shown more cumulative production in early and late time of production. The wells with down-dip deviation type which are partially completed in Upper and Lower Marcellus are the most productive wells.

3. The longer stimulated lateral and more clusters are always in favor of more gas production although given the same stimulated lateral length for a horizontal well, increasing the number of clusters results in reducing the cluster spacing which does not necessarily improve well performance and it should be optimized.

4. Having an optimum distance between the laterals in order to first establish enough connectivity to the matrix and second avoid well interference is very important.

5. This study also revealed that bringing the well to the production after completion with delay is not in favor of production performance. Those wells with longer delay have shown poor performance.

6. The more injected proppant per stage and also injected slurry per stage will improve the well performance.

**Acknowledgments**

The authors would like to thank Intelligent Solutions Inc. for supporting us with Improve software package.

**References**


Intelligent Solution Inc. Website- Oil Field Data Mining Overview


