OPTIMIZING COMPLETION TECHNIQUES WITH DATA MINING

Robert Balch
Petroleum Recovery Research Center
New Mexico Tech

ABSTRACT

The objective of this project is to use data mining to analyze well completion data to determine if trends or interesting patterns exist between well completion and stimulation methods and subsequent production. While a multitude of completion and stimulation techniques exist there are few objective, systematic examinations of their relative utility. A 370 well subset of basin Dakota wells drilled during 1994-2004 was examined to determine the feasibility of this concept with 58 additional wells drilled during 2004-2006 used for testing. Comparison of companies showed statistically significant differences in average production which could not be resolved by studies of land position via K-means clustering. Data mining was able to determine variances in first years gas production based on fracture fluid gallons, fracture fluid type, fractured interval thickness and sand lbs. A forward model was generated that could accurately estimate first years gas production based on these input parameters.

INTRODUCTION

Data mining is the extraction of hidden predictive information from large amount of Data using a variety of statistical algorithms and methods. The goal of data mining is two-fold: To find unexpected useful results and then create models which allow prediction of future trends. With ~29000 active gas wells in the San Juan Basin, each with logs, scout cards, completion data, and Production histories the San Juan basin is data rich. Such an immense amount of data would be impossible to examine as a whole by individuals or even groups of researchers. Research necessarily focuses on a priori assumptions that certain narrowly defined factors are important, and then data is reduced to manageable units that allow modeling of the pre-defined assumptions. Is it possible that embedded information in this immense dataset can lead to more efficient and effective completion practices? The tight Dakota and Mesaverde sands are an ideal candidate for application of data mining techniques to determine completion/stimulation "best practices" since there are a large number of existing wells and data about those wells, covering a long period of time and a large variety of completion/stimulation techniques. Drilling activity is continuing, and these plays are analogous to other tight gas reservoirs throughout the Rockies. This allows measurable economic advantages if improved completion practices are realized by the study.

METHODS

To establish a proof of concept a pilot study was performed on a small (370 well) subset of non-commingled Dakota wells drilled and completed during the time period 1994-2004[1, 2]. Predictive models were tested using an additional 58 Dakota wells completed during 2004-2006. The project used completion data from IHS Energy and production information from New Mexico's online database (ONGARD). Data studied included geographical Attributes such as Company Name, Completion Date, Location, and depth to Dakota Top as well as non geographical attributes such as Fracture Stages, Fracture Net Thickness, Fracture Gross Thickness, Fracture Fluid Type, Sand Lbs, Sand Type, Sand Size, and Sand Additive.

Data Mining Analysis

It was first necessary to determine f there are measurable differences between companies in well success as measured by gas production in the wells first 12 months of production. First years gas (FYG) was selected to allow comparison of wells with relatively short production histories, to those with many years of production. FYG also isolates production resulting from an initial completion/stimulation from subsequent work-over's. A box plot (Figure 1) was generated for all companies which occurred more than 10 times in the dataset. A visual examination shows that there is noticeable variance between companies and the composite company (35). A 2 sample T-Test was performed with a Null Hypothesis that each company FYG would be the same as the Average FYG of all Companies and it was found that 6 of 8 companies with > 10 wells were statistically different from the null hypothesis. Having observed that there is a real difference between successes of wells by company in the data set the next step was to determine if it was purely a land position which gave certain corporations an advantage. To test this production was exhaustively clustered using location and production data by company[3]. No dominant trends were observed. Companies that had good completions in the fairway were also good out of it, and vice-versa.

Unable to cluster FYG based on company criteria and well location data alone the conclusion was made that additional factors are involved in generating optimal production. The largest unstudied set of variables were all completion/stimulation related and further data mining was deemed necessary to find the best parameters for predicting FYG. Hypothesis-generating approaches were used to discover interesting relationships and patterns in the data. Attribute selection, classification, regression and clustering were performed using various methods and algorithms. Several indicators including the *InfoGainAttributeEval* which evaluates the worth of an attribute by measuring the information gain with respect to the class variable, the *GainRatioAttributeEval* which evaluates the worth of an attribute by measuring the gain ratio with respect to the class, and a *Chi-squared Attribute Evaluator* which evaluates the worth of an attribute by computing the value of the chi-squared statistic with respect to the class all identified similar attributes as important: Fractured Fluid Gallons, Fractured Gross Thickness, Fractured Fluid Type, Sand Lbs, and Acid Gallons.

Predictive Modeling

Several predictive models were built using these data inputs to determine if FYG could be predicted from these variables including decision trees developed using commercial software CART[4] and an in-house neural network package[5]. The CART analysis used a 10 fold cross validation and a 12 leaf node regression tree was generated with a RMS error of 0.12 (perfect RMS error would be 0.0). Primary variables for the CART analysis are shown in figure 2.

CART is proprietary software so a predictive model using it alone is not widely shareable; however the most important parameters for predicting FYG were confirmed to match those determined by other attribute selectors. Neural Networks are a common predictive tool for data mining projects and robust in-house software has been developed at the PRRC. Neural networks are essentially complex multivariate non-linear regression equations. A robustly trained neural network can make predictions given new input data appropriate for its domain. This approach was taken to find whether FYG can be predicted using the numeric attributes selected by the data mining algorithms. The best Architecture consisted on an input layer consisting of Fracture Net thickness, Fracture fluid Gallons, and Sand lbs, an output layer consisting of FYG and 3 hidden layers. The non-linear regression formed by this technique had 87 coefficients and had correlation coefficients of 0.93 and 0.84 for training and testing data, respectively. The model was used to predict wells not included in the analysis that were completed during 2004-2006 with a very high degree of accuracy.

Having found a relationship for predicting FYG from completion variables the next step was to determine an optimal set of completion parameters, based on well lithology, pay thickness, and other factors to optimize the completion process both in terms of expense and projected results. One potential area that has been examined is the relationships between the non-parametric Frac Fluid Type attribute (not useable in a NN analysis) and the numeric attributes Sand Lbs & Frac Fluid Gallons and trends Frac fluid Types and volumes used to complete wells throughout the study period. Interesting observations of that analysis include: Large volumes of sand and X-Link do not correlate to good FYG, lower sand volumes and possibly Slk-WTR volumes correlate to good FYG; good FYG is observed with high Gel Volumes even when sand volumes are reduced, and increasing sand volume appears to improve FYG with Foam.

CONCLUSIONS

Successful determination of primary parametric factors (numerical data) in completions that result in optimal production, were mined and those attributes were then used to construct successful models to predict future results. A robust decision tree using CART with a minimal RMS error was also constructed which allows the use of non-parametric information such as Fracture Fluid type. Trends between Fracture Fluid type and Sand Lbs and Frac Gallons give some indications of potentially useful trends. This preliminary work strongly suggests that completion parameters can be optimized and that it is possible to improve yield while decreasing expenses using data mining. A larger data mining project covering more time and more formations should result in an incremental increase in gas production from tight sands.

REFERENCES

1) Al-Tailji, W., 2006: "Analysis of well completion data with data mining techniques for the Dakota formation, San Juan basin, New Mexico", Masters Thesis, New Mexico Tech, Socorro, New Mexico, 69 pp.
2) Iduri, A. K., 2007: "Analysis of well completion data to Predict First Year Gas Production for the Dakota Formation, San Juan basin, New Mexico", Masters Independent Study, New Mexico Tech, Socorro, New Mexico, 51 pp.
3) Weka Data Mining software: http://www.cs.waikato.ac.nz/ml/weka/
4) CART regression Tree software: http://www.salfordsystems.com/cart.php
5) PredictOnline Neural Network Software: http://ford.nmt.edu:8080/predict.jsp

ACKNOWLEDGEMENTS

A generous donation of data from IHS Energy provided a consistent and high quality data set to base the data mining study on. Two MS students performed portions of the analysis, Ajay Iduri and Wadhah Al-Tailji while earning their degrees in Computer Science and Petroleum Engineering, respectively.

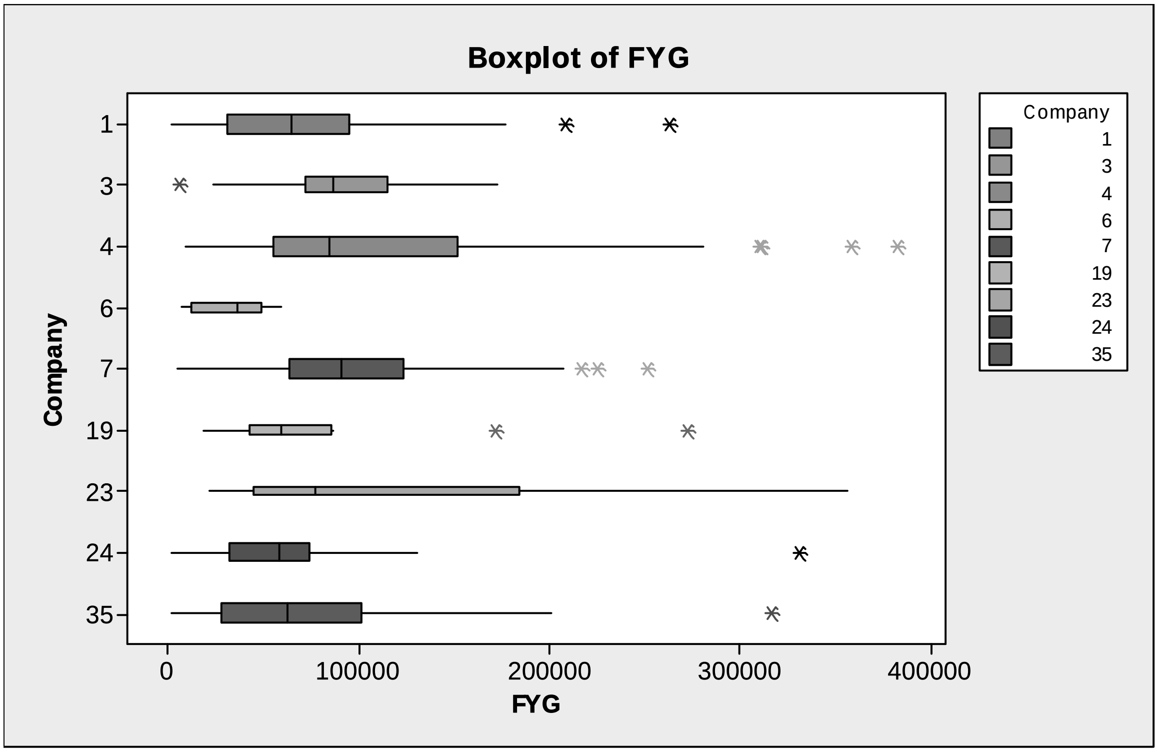
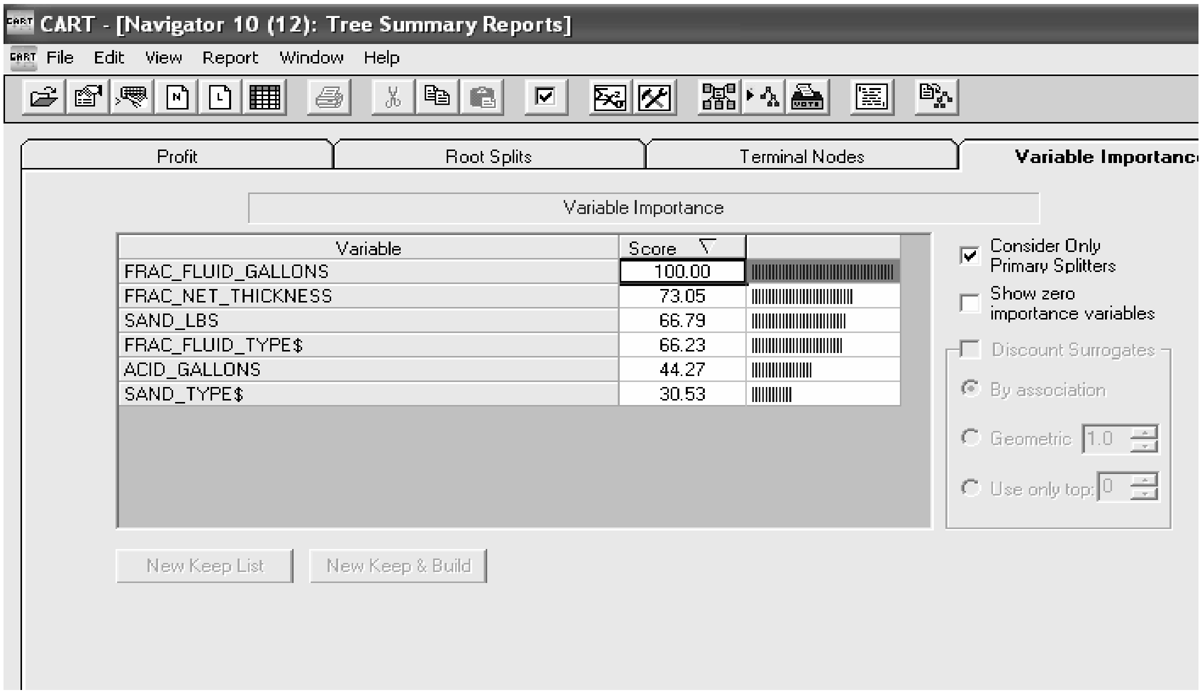


Figure 1 - Comparison to a composite production indicator (company 35) shows statistically significant differences in well success by company.


CART - [Navigator 10 (12): Tree Summary Reports]
File Edit View Report Window Help
Profit
Root Splits
Terminal Nodes
Variable Importance
New Keep List
New Keep & Build

Variable Importance

Variable	Score
FRAC_FLUID_GALLONS	100.00
FRAC_NET_THICKNESS	73.05
SAND_LBS	66.79
FRAC_FLUID_TYPE$	66.23
ACID_GALLONS	44.27
SAND_TYPE$	30.53

☑ Consider Only Primary Splitters

□ Show zero importance variables

□ Discount Surrogates

● By association

○ Geometric 1.0

○ Use only top: 0

Figure 2 - CART regression for significant parametric and nonparametric variables for FYG.

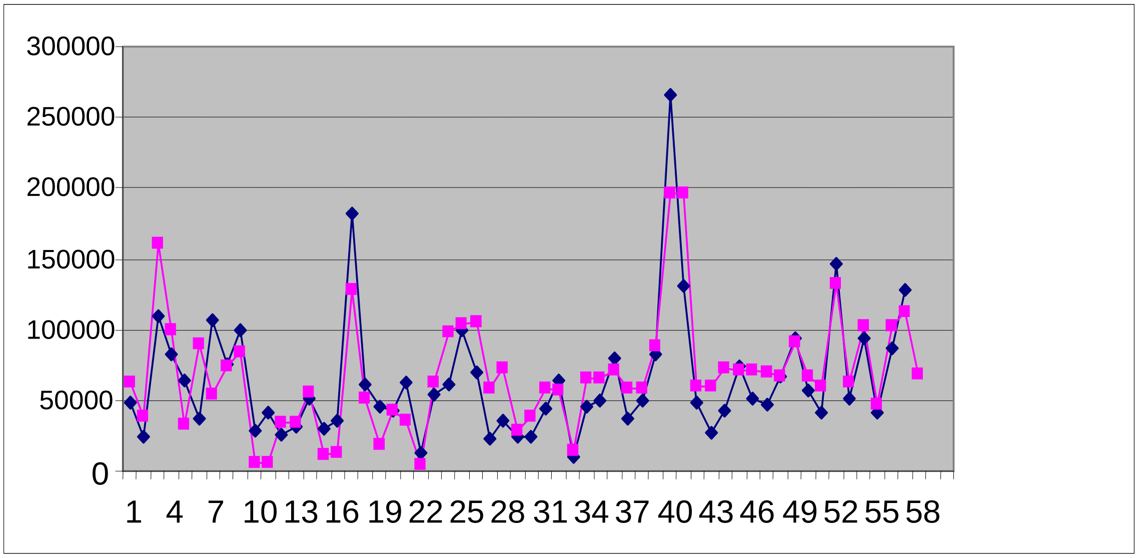


Figure 3 - The network is able to fluctuate between the maximum and minimum values of FYG which indicates a robust solution. Diamonds represent actual production in MCF and squares represent predicted values.