AADE-15-NTCE-37

Statistical Time Estimates for Deepwater Completions in the Gulf of Mexico: Integration of Neural Networking and Distribution Modeling

Paul Robinson, Stone Energy; Saeed Salehi, University of Louisiana at Lafayette

Abstract

Predicting completion time in deepwater wells is an imminent necessity in the modern well construction cycle. The primary objective of this paper is to present a novel integrated approach of statistical analysis and neural network models to identify well characteristics and their impact on total time to complete a well.

Using a neural network, fifteen crucial attributes from the Dodson Database were used in this study and analyzed for relative impact with respect to time. These attributes included primary parameters such as well depth and interval number. Wells in the database were assigned a value, depending on their fifteen attributes, that correlated to length of time to complete.

The program designated prospect wells a value using the same time weighted impact system, as well as the same impact parameters. Wells within the database with most similar values to the prospects were used in the statistical analysis for total completion time.
Actual data were used for fifteen parameters in the program for "Dark Star," "Liberty," and "Terrapin" to test the reliability of the statistical analysis. Estimates for "Dark Star" and "Liberty," which were completed in 2014, were within 5% of field completion time. "Terrapin" is yet to be completed; however, the programs estimate was within 3% of the Approval For Expenditure (AFE). With access to the data provided by Stone Energy as one of the active operators in GOM, this paper presents a valuable methodology to estimate completion time.

Introduction

Quest for hydrocarbons and rich oil reservoirs have made offshore environments a lucrative target for energy industry. Data provided by Bureau of Ocean Energy Management shows a huge increase for drilling in deep water from early offshore drilling in 1960s to present. Water depths often can go up to more than 10,000ft. Drilling in these offshore environments is a complex and risky process. Regardless of advancement in technologies to extract, drill and produce hydrocarbons, uncertainties are part of important factors in risk assessments and safety procedures. Huge cost associated with deep water projects is another key factor to be considered in pre-planning phase.

Often times, wells are drilled without encountering any hydrocarbons. Other times, hydrocarbons are encountered, but the targets are not profitable enough to cover the expense for drilling and completion operations. Needless to say, companies who are invested in a well want to have an estimate of how much it would cost to complete it. Thus, it is important to analyze costs prior to initiating any well plan.

Completion time estimates could be a determining factor as to whether or not to drill a well. The problem is that deepwater operations are relatively new and there is little completion data from which to base the estimates. Stone Energy recently developed a deepwater department, thus, even less data was available to estimate completion costs for deepwater wells in the Gulf of Mexico. Improving the completion time estimates was crucial for the newly developed Stone Energy deepwater department.

The number one factor affecting drilling and completion costs is time; the longer it takes, the more it costs. Therefore, the primary objective of this study was only to improve Stone Energy's statistical time estimates for deepwater completions. Previous deepwater completion time estimates were based on experience, rather than statistics. This was due to a minimal amount of data in the archive, as the Stone deepwater team was only recently developed.

Statistical time estimates for operations are typically performed at the preliminary stage of the total cost estimate for a well (drilling and completions). In addition, front-end time estimates are expected to be within ±40% of the actual completion field days. These estimates are fine-tuned as well planning progresses and more details are available, but there is the potential need for multiple preliminary statistical time estimates at once.

Stone Energy, an independent operator in Gulf of Mexico has recently received 40 prospect wells that required preliminary statistical cost estimates in a two-week period. With the accuracy only needing to be within 40% and the potential to have multiple prospect wells in a short period of time, spending substantial amounts of time on every well estimate would be a poor use of company resources. Thus, developing a user-friendly program that could rapidly perform statistical time estimates within the allotted accuracy range for deepwater completions was essential.

Dodson Datasystems

Dodson Datasystems® is a company that stores enormous amounts of drilling and completion data from a variety of operators in the Gulf of Mexico. Operators submit specific well parameters to the database, as well as time segments in two different stages: after drilling and after completion. The purpose of the database is to benchmark operators in the system.

Operators, who are subscribed to Dodson Datasystems, are allowed to view data from all the other operator's wells, provided they are subscribed as well. It is a useful tool for comparing performance of operators with each other, but the data can also be used to estimate the time to drill and complete a well. While Dodson Datasystems stores time data for a substantial amount of wells, it also considers the difficulty of the wells for benchmarking purposes.

Completion risk index (CRI) is Dodson Datasystems' approach to numerically measure the difficulty of a well. Therefore, if it takes a significant amount of time to complete a well with a higher complexity, it will be normalized when compared to easier wells. CRI is calculated using a Dodson Datasystems copy written formula and is included with every completed well that is submitted to the database. The CRI of every well considers 16 variables obtained from well attributes in the formula, all of which are required to be submitted by the operator upon completion of the well (Table 1). Figure 1 displays a consistent relationship between the CRI to completion time when plotted (IHS, 2013). Thus, the CRI alone can be a useful tool to estimate completion times.

The primary objective of this study was to improve statistical time estimates and create a program that simplifies the process. CRI is calculated using an algorithm, not statistically. If the CRI was used in the study, predictions would be made from predicted data, which would result in further error. Therefore, the CRI was not considered in this study. However, the study has the same premise as the CRI, a way to relate well parameters to completion time.

Data for 168 wells are stored in the Dodson Datasystems deepwater completion database. This data was obtained from wells that used deepwater rigs only: semi-submersible, platform, or drillship. As discussed earlier, this small number is due to deepwater operations being relatively new to the oil and gas industry. In addition to that, completions data are significantly smaller in the deepwater database in comparison to drilling, which has data for over 1,000 wells in the Dodson Datasystems database.

This study focused on the completion data within the Dodson Datasystems database, specifically, time data in relation to completion characteristics of wells. The drilling database defined total drilling time as spud to TD; total completion time was not this simple. Dodson Datasystems completion times were divided into nine different segments, and the non-productive time (NPT) associated with those times (Table 2).

All times had clear definitions of when they begin and end. For example, run and cement casing time started when the last logging tool was laid down and ended just prior to picking up the completion string to begin displacement (IHS, 2014).

The explanation for the nine time segments, as opposed to having set start dates and stop dates, was due completions having different operational procedures. For instance, not every well in the database ran and cemented casing. For that reason, the completion start date could not be defined solely by the commencement of running and cementing casing. This created consistency between wells within the database.

Additionally, there were wait on weather (WOW) times and rig failure times. Dodson Datasystems subtracted these times from the total completion time, as they do not accurately portray an operator's performance.

Having a clear definition of these time segments was crucial to this study, so that all the well times in the database were consistent with one another. Different operators will have different definitions of when a particular procedure starts and stops; Dodson Datasystems eliminated the potential for this variability.

Additionally, blow out preventer (BOP) certification times were not included for the wells in the Dodson Datasystems database. The British Petroleum (BP) oil spill in 2010, also known as the Macondo blowout, is responsible for the discrepancy in BOP certification times. Before the Macondo blowout, BOPs were not required to be certified before each new well. Post-Macondo wells were required to have BOPs certified prior to being latched, which added on about two weeks to total completion time on average. A database containing wells with BOP certification times and without BOP certification times would severely skew the total completion time statistical estimate.

Neural Networking

Methods

With only 168 wells in the deepwater completions database, it was difficult to filter out wells with similar characteristics. If a prospect well is greater than 25,000 feet and the wells were filtered by this criteria, the sample set would not be robust enough to perform a statistical analysis.

Approach based on using every well in the database would not give an accurate statistical representation of completion time. Furthermore, it cannot also provide an average time for all of the wells. If a prospect well with a total depth of 25,000 feet was to be analyzed for total completion time, using a well with a depth of 5,000 feet in the analysis would typically not be practical. Therefore, using this type of wells in the statistical analysis would give erroneous results.

An important step in data analysis is data normalization. By normalizing the data,any combination of wells could be used in the statistical analysis, regardless of whether or not the well characteristics were the same.

Well parameters, used in the CRI formula, were plotted against the different time segments obtained from Dodson Datasystems. This method was obtained from performing the theoretical background (Dawson et al., 1987). For instance, completion time previously was thought to increase with interval length. This process was repeated for every applicable well parameter against every Dodson Datasystems time segment, which resulted in over 750 plots. Scatter in the data made it difficult to decipher any true correlation between the parameters and completion time. Results of these plots were not significant enough to base any statistical time estimations. Additionally, there were two important explanations for not taking this approach:

1) Using every well in the database would not produce as accurate results as using a subset of wells with similar completion times.
2) Applying a trend line to all the data to fit our empirical equations for well parameters would be more representative of approaching the solution from an answer.

It was decided that another method of analysis was needed as the foundation of the time estimates.

Neural Network

Neural networks, or artificial neural networks, are mathematical constructs that were inspired by biological neural networks (Heaton, 2010). They are effective when it comes to accomplishing simple tasks, but they are particularly useful for recognizing patterns. The network structure is composed of input layers, hidden layers, and output layers.

The input layers are the data which are utilized in the analysis to predict the output layer. Within a dataset of six fields, five of the fields will be used to predict values within the sixth via pattern detection. These five fields may contain continuous or discrete variables, all of which are on different scales (Heaton, 2010). With the original input values, these fields cannot be compared to one another because they are not normalized.

Neural networks take these inputs of various scales, and convert them all on a scale of zero to one (Microsoft, 2014). This process is called scoring. These values are then normalized, so that they can be analyzed for their effects on the sixth field. The hidden layer accounts for independent and dependent variables. If two input parameters affect the output together, but not individually, they still need to be accounted for. For instance, if one of the five fields was removed from the analysis, the scores for all the variables would likely change.

Neural network models were used in this study to normalize the well characteristics, so that the properties were weighted relative to their impact on completion time. Thus, it was not important whether all the wells utilized in the analysis had similar water depths, as the other well properties would compensate for the differences.

For a statistical analysis to be valid, a minimum sample size of 30 must be used (Hubele, 2011). Thirty wells (30), from the database of 168, were used in each analysis. However, these sample wells require a combination of properties that would impact completion time similarly to the prospect well, regardless of whether the properties were the same or not. Therefore to execute this plan, completion properties of the wells needed to be analyzed with respect to completion time.

The background study conducted revealed that the primary driving factor for total completion time was well depth. However, the review did not elaborate on how much of an impact in comparison to other well properties. While it was easy to say that deeper wells typically took longer to complete, it was not as easy to explain why some deeper wells were completed faster than shallow wells. Therefore, other factors must have been contributing to the total well time, like the number of intervals.

The goal of this study was to identify which characteristics impact time, but primarily to determine the extent. As discussed earlier, Dodson Datasystems used the CRI value to numerically evaluating the difficulty of a well's completion. This value was calculated using the inputs outlined in Table 1. A discussion over the CRI was had with Stone's completion engineers to determine which of the input parameters were applicable to this study. Few of the parameters were omitted, but the majority was to be used in the neural network analysis.

When data is exported from the Dodson Datasystems database, it is exported with information for well intervals (Table 3). Table 3 displays a two interval well with parameter and time data; more data is displayed when exported from Dodson Datasystems, but not necessary to convey the message. The blank cell in the interval column is a summary of the total well, however, the data cannot be exported as the well summary only. Thus, wells with more than one interval needed to be combined for this study. It was decided that the deeper the interval, the longer the completion would take. The deepest interval was recorded for wells with more than one interval, likewise, the longest interval was recorded. This process was repeated for all 15 applicable well parameters. Once all interval data had been combined into well data, the neural network analysis could commence.

Microsoft Excel, in combination with SQL Server, was the program used to perform the neural network analysis (Microsoft, 2014). SQL Server Data Mining Add-Ins for Office needed to be installed in Microsoft Excel to perform any of the neural network analysis. While this powerful software can perform multiple functions, the Prediction Calculator was the only tool used in this study.

Prediction calculator uses the Microsoft Logistic Regression algorithm, but it is simplified to be user friendly. Logistic regression is a method of determining the contribution of multiple factors to a pair of outcomes. As discussed earlier, the multiple factors are converted on the same scale and designated scores depending on their effect on the outcome pairs. The factors used in the analysis can be either discrete or continuous variables. Microsoft calls this process of converting these values the Z-Score normalization (Microsoft, 2014).

Step one of this process was to choose a target column within the dataset to define the pair of outcomes in the analysis. The objective of performing a neural network analysis was to compare well parameters for impact on total completion time, thus, the target column was total completion time. A maximum value, a minimum value, and a median were obtained from the total completion time data and were 10.2 days, 34.1 days, and 210 days, respectively. The pair of outcomes selected for the analysis was the time data in the range of 10.2 days to 34.1 days and 34.1 days to 210 days. The theory behind this method was to completely divide the dataset in half, and compare 50% of the wells with faster completion times to wells with 50% of the wells with longer completion times. The explanation for this was that wells in the 50th percentile to the 100th percentile would have attributes that impact completion time more than those in the 1st percentile to the 50th percentile (Figure 2).

Once the pair of outcomes were selected, well characteristic categories used in the analysis were selected. These were the 15 parameters that were obtained from the Dodson CRI calculation. Columns of data for each of the 15 parameters for all 168 wells were selected to be used in the analysis and compared to completion time.

When the program was run, it went through iterations using Equations 6 – 9 to determine the scores of every potential input for every category. Impact results for every possible input were obtained (Table 4).

Results were all on the same relative impact scale, regardless of the category. Now the number of intervals could be compared to completion depth to determine, which typically impacts time more. The higher the value of the relative impact of each parameter, the more likely a well with that characteristic was to be in the 50th to 100th time percentile range.

Results from the first iteration of this process were inconsistent. For instance, as interval depth increases, so should the impact on completion time. However, this was not the case in Table 4. Results indicated that interval bottom depths within the range of 14,469 feet to 17,800 feet would likely take less time than interval bottom depths within the range of 10,999 feet to 14,469 feet. This is likely attributed wells to having inconsistent steps during the completion process.

As discussed earlier, time data in Dodson Datasystems was broken down into segments, as opposed to simply a start date and a stop date. This was due to the wide variety of steps that may or may not be performed during the completion process (IHS, 2014). This subject was discussed with the Stone completion team and it was decided that three time segments would not always be performed: Run and Cement Casing, Well Test, and Temporarily Abandon (TA) Re-Entry. That being said, wells that performed all of these processes, but had characteristics that would typically not have a significant impact on time, were skewing the data. To remedy this, the neural network analysis needed to be performed against total time with every possible combination of these three segments subtracted out. This resulted in eight different time columns (Table 5) and eight different impact scales (Table 6).

When the analysis was run against the eight different time columns, there were eight different impact values for every single parameter. Results were still skewed for all the potential impact columns (Table 6), for the same reason stated earlier. Some variables had a significant impact in one column, but a minor impact in another. However, if a variable had a significant impact in any column, it could be argued that it had an impact on completion time but it was skewed from breakdown of the time datasets. In order to portray this, the maximum value of each variable needed to be taken from all the relative impact time columns (Table 7).

This eliminated any skewed data from the inconsistent completion times. Maximum values obtained from this analysis were to be used as the scaled weight system of well parameters with respect to time. With the new scaled weight system, it was not as important that wells with exact or similar properties as the prospect well were in the database.

Every well in the database was assigned a score, determined from the well parameters for the fifteen categories. Each well attribute was linked to the weighted scale, and that corresponding value was applied to the well. Impact values for all 15 parameters were assigned to every well in the database, depending on the well characteristics (Figure 4).

For each well, the relative impact values were summed to give the well a time score: the higher the score, the higher the likelihood of increased completion time (Table 8).

The program was designed to have user inputs for prospect wells that were linked to the same weighted scale system. Thus, the prospect well would be assigned a score using the exact same scale as every well in the database. The relative impact values of the prospect well were summed to designate a score as it relates to time. Wells with similar scores were more likely to have similar completion times, regardless of how similar the well properties were. This was the foundation for how the 30 wells were selected for the sample set to be used in the statistical analysis.

Scores of every well in the database were linked to the score of the prospect well. The absolute value of the difference between the well scores in the database and the prospect well score were taken. These were the values used to sort the data for wells with the most similar completion times. The smaller the difference, the closer completion time between wells. However, there were still some adjustments that needed to be incorporated into the program in order to produce accurate results.

Statistical Analysis

Stone engineers decided that they wanted to filter the data depending on whether the tree was located subsea or at the surface. Macros needed to be written to determine which time column the statistical analysis should be run on. These macros were dependent on the user's selection on the data sorting page.

The user decides whether the data should be sorted by subsea tree only, surface tree only, or all wells regardless of tree location. Additionally, the user determines which times to include in the statistical analysis, results will change accordingly. Therefore, if the prospect well plan did not include a well test, the user would select the radio button of "no well test" in whichever column was most applicable.

Originally, the only statistical results were going to be 10th, 25th, 50th, 75th, and 90th percentiles from the sample dataset of the closest 30 wells. This is due to outliers have much less of an impact on the median as opposed to the mean.

For instance, the 100th percentile of the entire dataset is 210 days, which would significantly skew the average of the 30 wells if it were to be selected sample where the remainder of the wells averaged to be about 25 days. If the 100th percentile of the sample set was the only value to change, it would not skew the median at all. It would be irrelevant whether the 100th percentile was 100 days or 45 days. The same procedure was followed for the calculating NPT for the sample wells.

Though this method produced accurate results, a more detailed analysis needed to be performed. First, it was necessary to determine the distribution of the entire dataset. A program, Rose & Associates Toolbox, was utilized to determine the statistical distribution of the all the wells in the dataset. The program resulted in a log-normal distribution.

Second, for reassurance, the program was run on a few sample sets of thirty wells depending on the 15 different input parameters. Results from the sample sets also resulted in a log-normal distribution (Log-normal Solutions, 2012). For the program to remain user friendly, a log-normal distribution was used for the statistical analysis, regardless of the sample set.

For the closest thirty wells that resulted from the sort, the natural log needed to be taken for each of their completion times (Equation 1).

$\ln(t_i)$ = ln(completion time for each well)……….…….(1)

The average of the thirty log-normal distribution times were taken to get a new distribution mean (Equation 2).

$$\mu = \sum_{i=1}^{N} \frac{\ln(t_i)}{N} \quad \text{……………………………………….(2)}$$

The standard deviation of the log-normal distribution times were taken to get a new distribution standard deviation.

$$\sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\ln(t_i) - \ln(t))^2} \quad \text{……………………………….(3)}$$

Then the log-normal inverse could be taken to determine the probabilities (Equation 4).

$$F(X) = F(X, m, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^X \frac{e^{-\frac{(\ln(t)-m)^2}{2\sigma^2}}}{t} \quad \text{………………....(4)}$$

The natural log of the total time for every well in the database was taken, then the mean and standard deviation of these values were obtained to develop probabilities from the lognormal dataset. The lognormal distribution of the closest thirty wells and every well in the database resulted in the following probabilities: P10, P25, P50, P75, and P90 (Table 9).

Additionally, the Swanson Mean was available for those who choose to use it (Log-normal Solutions, 2012).

The Swanson mean is calculated using Equation 5:

Swanson Mean = 0.3*P10+0.4*P50+0.3*P90 (5)

Example results from running the program are displayed in Table 9.

To ensure there were no errors as to which parameters the data was sorted from, a sorting criteria box displays the inputs to the left of the results.

The process was repeated using 75 percent of the wells to train the data, and 25 percent of the wells to test the data. Entirely new relative impact values were obtained using the 126 wells. Maximum values for every parameter were taken after the Prediction Calculator was run against the eight different time combinations. The 42 test wells were then plugged into the developed program to test the reliability. While the average of the program results was still within 40 percent of the targets, approximately 33 percent, taking out 42 wells hindered the accuracy. Even without knowing the accuracy of the model, it is understood that more data used to train a model will produce better results. Thus, the original program will be used for the remainder of this study.

Input parameters for three Stone wells were used for quality control of the program: Dark Star, Liberty, and Terrapin. Prospect well names have been altered to protect Stone Energy confidentiality. Dark Star and Liberty were completed in early 2014, thus, program results could be compared to actual completion field days that were submitted to Dodson Datasystems. Unfortunately, Terrapin was planned to be completed in mid-2015, so field days were unavailable for a comparison. However, Terrapin's Approval for Expenditure (AFE) had already been completed and the normalized AFE days could be compared to the program results.

It is important to note that Dodson had a clear definition for the time breakdown, so that all time data was consistent throughout the wells in the database. AFE days needed to be normalized to subtract out BOP certification times, because many of the wells were pre-Macondo and did not include BOP certification times. It would be impractical to estimate completion time of a well if the data used in the analysis had different time definitions.

Results

A statistical time estimate is only as accurate as the data used in the analysis. Estimations in this study used data obtained from Dodson Datasystems. Thus, program developed in this study produced total completion time results, as Dodson Datasystems described it. Luckily, Dodson did an excellent job normalizing the data, so that there was consistency throughout the dataset. The clear definitions of the time segments ensured that there were no discrepancies of time data between the wells. BOP certification times were not included in the statistical time estimates, as they were not included in the Dodson data. Multiple wells were submitted to Dodson prior to the Macondo blowout, thus, they did not include BOP certification times. Wells submitted after the Macondo blowout were required to certify the BOPs, which took an additional two weeks on average.

The entire objective of this study was to estimate total completion time for deepwater Gulf of Mexico wells, based off of well attributes. A dataset where some wells included BOP certification times and some wells did not would severely skew the results of the time estimate. Dodson Datasystems mitigated this problem by normalizing the data and subtracting out the BOP certification times all together.

Operators typically included BOP certification time in their definition of total completion time, but these statistical estimates cannot be performed without enough data. For best results, approximately two weeks should be added to the total completion time after running the program.

Results of the analysis for the three wells were extremely consistent. Completion parameters for Dark Star were input into the program. The program was run using wells with subsea trees only and with well test times subtracted out. Percentile and probability results are displayed in Table 10. Completion parameters for Liberty were input into the program. The program was run using wells with subsea trees only and with well test times subtracted out. Percentile and probability results are displayed in Table 11. Completion parameters for Terrapin were input into the program. The program was run using wells with subsea trees only and inclusive of all times. Percentile and probability results are displayed in Table 12.

Conclusions

Discussion of Results

For this study, it was determined that the 50th percentile was the most accurate representation of well time, as it eliminates the effect of outliers.

Previously, the 50th percentile was used as the AFE days and the 25th percentile as the TF days. However, the program can calculate the TF days by subtracting the NPT from the corresponding total time. Additionally, the probabilities can be used for the time estimation, as the results were similar to the percentiles. The program allows for the user to determine which values to use for the time estimation. Results from this study indicated that the most accurate representation of the actual field days was the 50th percentile, which was what the program was originally designed to do. The program 50th percentile results, normalized AFE days, and actual field days for the three Stone wells is displayed in Table 13.

For Dark Star and Liberty, the program results were within five percent of the normalized field days. These were actually more accurate than the normalized planned AFE days, as shown in the "Days (% of Field)" row. For Terrapin, the program estimation was within one day of the planned AFE days.

Though the probabilities from the lognormal distribution did not give as accurate of results as the 50th percentile for Dark Star, Liberty, and Terrapin in this study, they will improve as the dataset increases. It is likely that the probability results will surpass the percentile results when it comes to accuracy when the dataset becomes more robust.

As stated earlier, the purpose of this study was to develop a program that could quickly estimate the amount of completion days within 40% accuracy. The results far exceeded the original intention of the program. However, this is an ongoing study and there is still much work to be performed.

Future Work

When a large enough amount of post-Macondo wells are added, the BOP certification times will be available to be included in the statistical time estimate. While the developed program has the capability of subtracting times from total times, it cannot add times in; these results would be erroneous. However, with more data, the sample datasets can be selected to only include specific times.

Additionally, the Prediction Calculator will give more consistent results without any data gaps that were discussed in the methods section, so there will be no need to take the maximum of all the parameters for the time segments (Microsoft, 2014).

In addition to annual updates of the deepwater statistical time estimate program, the deepwater drilling statistical time estimate needs to be updated. The entire process, with the exception of a few variations, will need to be repeated for Dodson Datasystems deepwater drilling database. The two programs will then be incorporated linked together in order to produce accurate well time estimates. With the addition of tangible equipment costs, rig rates, and spread rates, the program will be able to quickly provide total well costs.

Eventually, an entire new program will need to be developed around the data to apply additional sorting options. However, for the time being, this program is effective in quickly producing statistical time estimates for deepwater completions in the Gulf of Mexico within a 40% accuracy range.

Acknowledgments

Authors would like to thank Stone Energy Corporation for providing access to the data used in this paper.

Special thank goes to Dodson Datasystems for providing the data used in the analysis.

Nomenclature

CRI = *Completions risk index*
AFE = *Approval for expenditure*
BOP = *Blow out preventer*
BP = *British Petroleum*
DDR = *Daily drilling report*
ft =*Feet*
i = *Well number in dataset*
in = *Inches*
MD = *Measured depth (feet)*
μ = *Mean*
N = *Number of samples in dataset*
NPT = *Non-productive time*
Σ = *Standard deviation*
t = *Time (days)*
TA = *Temporarily abandon*
TF = *Trouble free*
TFT = *Trouble free time*
WOW = *Wait on weather*
X = *Well value for particular parameter in statistical analysis*

References

1. Adams, A., Gibson, C., and Smith, R.G.: "Probabilistic Well Time Revisited." SPE 119287, SPE/IADC Drilling Conference and Exhibition, Amsterdam, March 17-19, 2009.
2. Dawson, R., Floetra, A., Noerager, J.A., Norge, E., and White, J.P. "Drilling Time Predictions from Statistical Analysis." SPE

16164-MS, SPE/IADC Drilling Conference, New Orleans, March 15-18, 1987.
3. Dodson Datasystems V6.6. "Deepwater Gulf of Mexico Completion Data. 2014. Retrieved from: http://www.dodsondatasystems.com/Private/OffshoreCompletions/Query/Default.aspx
4. Heaton, J., "A Non-Mathematical Introduction to Using Neural Networks." Heaton Research, Inc., Missouri, 2010.
5. Heaton, J., "Introduction to the Math of Neural Networks." Heaton Research, Inc., 2012.
6. Hubele, N.F., Montgomery, D.C., and Runger, G.C., "Engineering Statistics." John Wiley and Sons, New Jersey.
7. IHS. "Offshore Completions Performance Benchmarking: Data Definitions." Version 6.6.0., 2014. Retrieved from: http://www.dodsondatasystems.com/Private/OffshoreCompletions/Resources/Default.aspx
5. IHS. "Offshore Completions Performance Benchmarking: Data Definitions." Version 6.6.0., 2014, Retrieved from: http://www.dodsondatasystems.com/Private/OffshoreCompletions/Resources/Default.aspx
6. Lognormal Solutions, Inc. Rose & Associates Toolbox V4-3-109, 2012, Retrieved from: http://www.roseassoc.com/Toolbox.html
7. Microsoft. Content Types (Data Mining). 2014. Retrieved from: http://msdn.microsoft.com/en-us/library/ms174572.aspx
8. Microsoft. Feature Selection (Data Mining). 2014. Retrieved from: http://msdn.microsoft.com/en-us/library/ms175382.aspx
9. Microsoft. Microsoft Logistic Regression Algorithm. 2014. Retrieved from: http://msdn.microsoft.com/en-us/library/ms174828.aspx
10. Microsoft. Microsoft Logistic Regression Algorithm Technical Reference. 2014. Retrieved from: http://msdn.microsoft.com/en-us/library/cc645904.aspx
11. Microsoft. Prediction Calculator (Table Analysis Tools for Excel). 2014. Retrieved from: http://msdn.microsoft.com/en-us/library/dn282357.aspx
12. Microsoft. SQL Server Data Mining Add-Ins for Office. 2014. Retrieved from: http://msdn.microsoft.com/en-us/library/dn282373.aspx

Tables

Table 1: Parameters used in Dodson Datasystems CRI formula.

Rig Type	Production Casing Size
Tubing Metallurgy	Sand Control Type
Production Casing Type	Intelligent Completion?
Completion Type	Bottom Hole Temperature > 300°F
Tree Type	Production Casing Squeezed?
Mechanical Type	Re-Entry to Temporarily Abandoned Well?
Interval Bottom Depth	Hole Angle at Perforation
Interval Length	Completion Fluid Weight

Table 2: Dodson Datasystems completion time segments.

Casing Run and Cement Time	Install Tree Time
Pick Up Completion String, Displacement, and Filter Time	Rig Down and Move Off Time
Perforation Time	Well Test Time
Sand Control Time	Rig Up Time on Re-Entry Well (TA Re-Entry Time)
Run Tubing Time	

Table 3: Example of interval data as it is exported from Dodson Datasystems

AREA	WATER DEPTH	INTERVAL	INTERVAL BOTTOM DEPTH	INTERVAL LENGTH	TOTAL COMPLETION TIME
East Breaks	3673	1	12044	18	12.1
East Breaks	3673	2	11836	61	2.59
East Breaks	3673		12044	79	14.69

Table 4: Example of Prediction Calculator results of impact values for Interval Number, Interval Bottom Depth, Rig Type, and Tree Type.

ATTRIBUTE	VALUE	RELATIVE IMPACT
Interval	1	0
Interval	2	37
Interval	3	91
Interval Bottom Depth	<10999	0
Interval Bottom Depth	10999-14469	24
Interval Bottom Depth	14469-17800	19
Interval Bottom Depth	17800-24659	68
Interval Bottom Depth	>=24659	62
Rig Type	DS	24
Rig Type	PF	24
Rig Type	SS	0
Tree Type	PF	32
Tree Type	SP	95
Tree Type	SS	77
Tree Type	TL	0
Tree Type	UK	66

Table 5: Screenshot of well time data with potential combinations of time segments subtracted out.

TIME COMBINATIONS	EXAMPLE WELL TIMES (DAYS)
TOTAL COMP TIME	210
NO RE-ENTRY NO WELL TEST NO CASING RUN	147.2
NO RE-ENTRY NO WELL TEST	166
NO RE-ENTRY NO CASING RUN	150.2
NO RE-ENTRY	169
NO WELL TEST NO CASING RUN	31.8
NO WELL TEST	188.2
NO CASING RUN	191.2

Table 6: Prediction calculator impact value results for eight possible time combinations.

ATTRIBUTE	VALUE	ALL TIMES	NO CASING RUN	NO WELL TEST	NO TA RE-ENTRY	NO WELL TEST, NO CASING RUN	NO TA RE ENTRY, NO CASING RUN	NO TA RE ENTRY, NO WELL TEST	NO CASING RUN, NO WELL TEST, NO TA RE ENTRY
INTERVAL	1	0	1	0	0	0	6	0	0
INTERVAL	2	24	18	24	24	24	30	24	26
INTERVAL	3	60	0	52	54	13	0	53	4
INTERVAL BOTTOM DEPTH	<10999	0	0	0	0	0	0	0	0
INTERVAL BOTTOM DEPTH	<14469	16	38	23	12	37	35	35	39
INTERVAL BOTTOM DEPTH	<17800	13	33	16	10	42	24	22	38
INTERVAL BOTTOM DEPTH	<24659	45	59	41	46	60	61	52	67
INTERVAL BOTTOM DEPTH	>= 24659	41	70	43	38	73	66	59	73

Table 7: Maximum relative impact values for every potential parameter.

SORT VALUES	ALL TIMES	NO CASING RUN	NO WELL TEST	NO TA RE-ENTRY	NO WELL TEST, NO CASING RUN	NO TA RE ENTRY, NO CASING RUN	NO TA RE ENTRY, NO WELL TEST	NO CASING RUN, NO WELL TEST, NO TA RE ENTRY
6	0	1	0	0	0	6	0	0
30	24	18	24	24	24	30	24	26
60	60	0	52	54	13	0	53	4
0	0	0	0	0	0	0	0	0
39	16	38	23	12	37	35	35	39
42	13	33	16	10	42	24	22	38
67	45	59	41	46	60	61	52	67
73	41	70	43	38	73	66	59	73

COMPLETION FLUID WEIGHT	TEMPERATURE >300?	SQUEEZED?	SAND CONTROL TYPE	**TOTAL**
51	3	15	57	**593**
51	3	15	57	**597**
23	3	15	57	**587**
51	3	15	57	**583**
39	3	15	57	**603**
51	3	15	57	**603**
39	3	15	57	**577**
51	3	15	57	**607**
23	3	15	57	**574**
39	3	15	77	**572**
51	3	15	57	**568**
39	3	15	57	**566**
39	3	15	57	**564**
39	3	15	77	**563**
39	3	15	74	**561**
39	3	15	57	**556**
51	3	15	57	**551**
51	3	15	57	**637**
51	3	15	57	**545**
51	3	15	74	**639**
39	3	15	57	**543**
51	3	15	74	**530**
51	3	15	57	**526**
23	3	15	74	**525**
23	3	15	74	**523**
23	3	15	74	**520**
51	3	15	57	**520**
51	3	15	57	**516**
5	3	15	74	**511**
51	3	15	74	**504**
5	3	15	57	**501**
51	3	15	74	**499**
21	3	43	74	**473**
51	3	15	74	**466**

Table 8: Image showing four of the 15 well parameters for database wells being summed to give a total well score with respect to completion time.

Table 9: Example results from running program.

Sorting Criteria
All Subsea Tree Wells
No Well Test
No Casing Run

All Subsea Tree Wells	Total Days	NPT Days
10th Percentile	15.54	0
25th Percentile	18.7725	1.2775
50th Percentile	25.1	3.605
75th Percentile	30.8325	9.49
90th Percentile	51.033	32.71

All Subsea Tree Wells	TF Days	NPT%
10th Percentile	15.54	0.00
25th Percentile	17.495	6.81
50th Percentile	21.495	14.36
75th Percentile	21.3425	30.78
90th Percentile	18.323	64.10

Top 30 Subsea Tree Wells	Total Days	NPT Days
10th Percentile	20.803	1.495
25th Percentile	24.025	3.2025
50th Percentile	28.985	6.37
75th Percentile	36.475	16.65
90th Percentile	59.9	34.948

Top 30 Subsea Tree Wells	TF Days	NPT%
10th Percentile	19.808	7.19
25th Percentile	20.8225	13.33
50th Percentile	22.615	21.98
75th Percentile	19.825	45.65
90th Percentile	24.952	58.34

Probabilities	Total Days
P10	16.86
P25	23.14
P50	32.87
P75	46.71
P90	64.08

Swanson Mean	37.4309

Table 10: Dark Star time results after running the program.

Sorting Criteria
All Subsea Tree Wells
No Well Test

All Subsea Tree Wells	Total Days	NPT Days
10th Percentile	20.09	0
25th Percentile	24.89	1.2775
50th Percentile	34.2	3.605
75th Percentile	47.93	9.49
90th Percentile	69.86	32.71

All Subsea Tree Wells	TF Days	NPT%
10th Percentile	20.09	0.00
25th Percentile	23.6125	5.13
50th Percentile	30.595	10.54
75th Percentile	38.44	19.80
90th Percentile	37.15	46.82

Top 30 Subsea Tree Wells	Total Days	NPT Days
10th Percentile	20.61	0
25th Percentile	24.89	1.08
50th Percentile	32	3.23
75th Percentile	38.66	4.07
90th Percentile	42.89	6.75

Top 30 Subsea Tree Wells	TF Days	NPT%
10th Percentile	20.61	0.00
25th Percentile	23.81	4.34
50th Percentile	28.77	10.09
75th Percentile	34.59	10.53
90th Percentile	36.14	15.74

Probabilities	Total Days
P10	20.14
P25	25.04
P50	31.90
P75	40.63
P90	50.53

Swanson Mean	33.961

Table 11: Liberty time results after running the program.

Sorting Criteria
All Subsea Tree Wells
No Well Test

All Subsea Tree Wells	Total Days	NPT Days
10th Percentile	20.09	0
25th Percentile	24.89	1.2775
50th Percentile	34.2	3.605
75th Percentile	47.93	9.49
90th Percentile	69.86	32.71

All Subsea Tree Wells	TF Days	NPT%
10th Percentile	20.09	0.00
25th Percentile	23.6125	5.13
50th Percentile	30.595	10.54
75th Percentile	38.44	19.80
90th Percentile	37.15	46.82

Top 30 Subsea Tree Wells	Total Days	NPT Days
10th Percentile	30.79	2.59
25th Percentile	35.15	3.48
50th Percentile	42.99	7.65
75th Percentile	64.25	17.74
90th Percentile	92.27	39.63

Top 30 Subsea Tree Wells	TF Days	NPT%
10th Percentile	28.2	8.41
25th Percentile	31.67	9.90
50th Percentile	35.34	17.79
75th Percentile	46.51	27.61
90th Percentile	52.64	42.95

Probabilities	Total Days
P10	26.54
P25	36.11
P50	50.84
P75	71.58
P90	97.40

Swanson Mean	57.518

Table 12: Terrapin time results after running the program.

Sorting Criteria
All Subsea Tree Wells
All Times

All Subsea Tree Wells	Total Days	NPT Days
10th Percentile	21.48	0
25th Percentile	27.53	1.2775
50th Percentile	37.4	3.605
75th Percentile	55.92	9.49
90th Percentile	76.94	32.71

All Subsea Tree Wells	TF Days	NPT%
10th Percentile	21.48	0.00
25th Percentile	26.2525	4.64
50th Percentile	33.795	9.64
75th Percentile	46.43	16.97
90th Percentile	44.23	42.51

Top 30 Subsea Tree Wells	Total Days	NPT Days
10th Percentile	33.37	2.59
25th Percentile	39.58	3.48
50th Percentile	45.01	7.65
75th Percentile	66.42	17.74
90th Percentile	115.08	39.63

Top 30 Subsea Tree Wells	TF Days	NPT%
10th Percentile	30.78	7.76
25th Percentile	36.1	8.79
50th Percentile	37.36	17.00
75th Percentile	48.68	26.71
90th Percentile	75.45	34.44

Probabilities	Total Days
P10	28.11
P25	38.38
P50	54.23
P75	76.64
P90	104.63

Swanson Mean	61.514

Table 13: Dark Star, Liberty, and Terrapin results compared to normalized AFE days and normalized field days.

			Dark Star
	New Database Estimate	Normalized Plan Days (AFE)	Normalized Field Days
Days	32	33.41	32.26
Days (% of Field)	99.19%	96.56%	

			Liberty
	New Database Estimate	Normalized Plan Days (AFE)	Normalized Field Days
Days	42.99	39.73	44.92
Days (% of Field)	95.70%	88.45%	

			Terrapin
	New Database Estimate	Normalized Plan Days (AFE)	Normalized Field Days
Days	45.01	46	
Days (% of Field)			

Figures

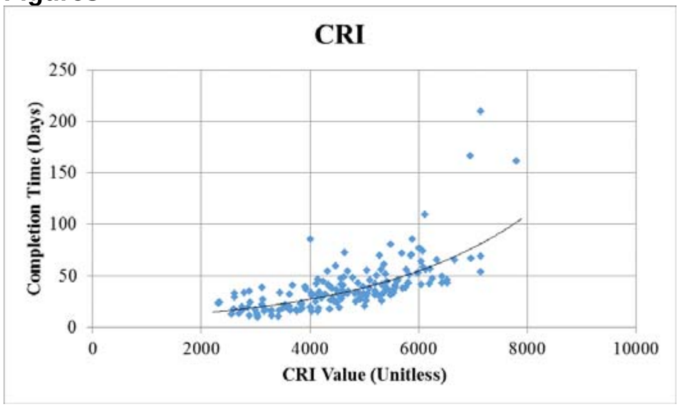


Figure 1: Dodson Datasystems CRI plotted against time

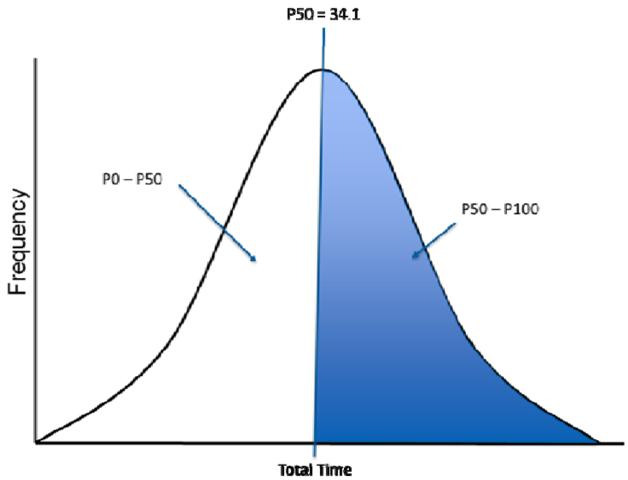
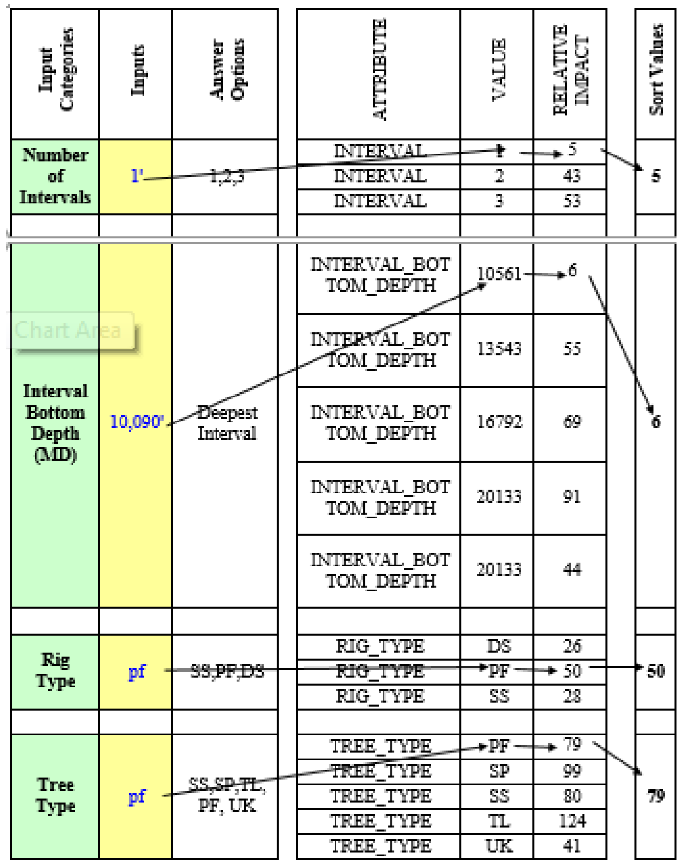


Figure 2: Bell curve with median dividing the pair of outcomes to show wells with completion properties that increase time.



Input Categories	Inputs	Answer Options	ATTRIBUTE	VALUE	RELATIVE IMPACT	Sort Values
Number of Intervals	1'	1,2,3	INTERVAL	1	5	5
			INTERVAL	2	43	
			INTERVAL	3	53	
Interval Bottom Depth (MD)	10,090'	Deepest Interval	INTERVAL_BOTTOM_DEPTH	10561	6	6
			INTERVAL_BOTTOM_DEPTH	13543	55	
			INTERVAL_BOTTOM_DEPTH	16792	69	
			INTERVAL_BOTTOM_DEPTH	20133	91	
			INTERVAL_BOTTOM_DEPTH	20133	44	
Rig Type	pf	SS,PF,DS	RIG_TYPE	DS	26	
			RIG_TYPE	PF	50	50
			RIG_TYPE	SS	28	
Tree Type	pf	SS,SP,TL, PF, UK	TREE_TYPE	PF	79	79
			TREE_TYPE	SP	99	
			TREE_TYPE	SS	80	
			TREE_TYPE	TL	124	
			TREE_TYPE	UK	41	

Figure 3: Image of input page automatically populating prospect well score with respect to time from well characteristics.