Paper SAS6421-2016

From Sensors to Solutions: An Example of Analytics in Motion

Moray L. Laing, Gilbert Hernandez, Philip Easterling, and Milad Falahi, SAS Institute Inc.

ABSTRACT

Oil and gas wells encounter many types of adverse events, including unexpected shut-ins, scaling, rod pump failures, breakthrough, and fluid influx, to name but a few common ones. This paper presents a real-time event detection system that presents visual reports and alerts petroleum engineers when an event is detected in a well. This system uses advanced time series analysis on both high- and low-frequency surface and downhole measurements. This system gives the petroleum engineer the ability to switch from passive surveillance of events, which then require remediation, to an active surveillance solution, which enables the engineer to optimize well intervention strategies. Events can occur simultaneously or rapidly, or can be masked over longer periods of time. We tested the proposed method on data received from both simulated and publicly available data to demonstrate how a multitude of well events can be detected by using multiple models deployed into the data stream. In our demonstration, we show how our real-time system can detect a mud motor pressure failure resulting from a latent fluid overpressure event. Giving the driller advanced warning of the motor state and the potential to fail can save operators millions of dollars a year by reducing nonproductive time caused by these failures. Reducing nonproductive time is critical to reducing the operating costs of constructing a well and to improving cash flow by shortening the time to first oil.

INTRODUCTION

This paper deals with the art of converting sensor measurements taken at the edge of oil and gas processes and converting them into solutions that improve efficiency and drive out unnecessary costs. But first let us consider the ecosystem of an oil and gas operation, in particular the various types of solutions that are needed to turn sensor measurements into solutions. The examination begins with the sensor data itself. Many people talk about “real-time data” and “real-time analytics,” but what does that really mean?

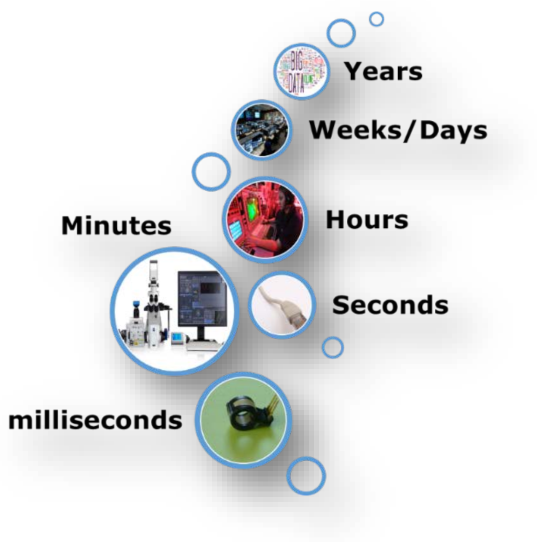


Figure 1. The Ecosystem of Real-Time Data

Is it data in motion at the asset? Or, do we also include data that is captured in real time, but that is stored offline for exploration at a later date? In fact, by looking at Figure 1 above, we see several levels of real-time data that are used in building-out solutions. These range from data in motion at the point of measurement and control all the way back to the enterprise, where data scientists and engineers might be combining recently acquired data with years of historical and comparative real-time information.

Now that we understand the ecosystem of the data, we must examine the solutions. Luckily for us, the International Society of Automation has already published a standard that helps to define this. ISA95 defines the levels of architecture involved in an automated solution in five distinct levels, the functional model of which is shown in Figure 2.

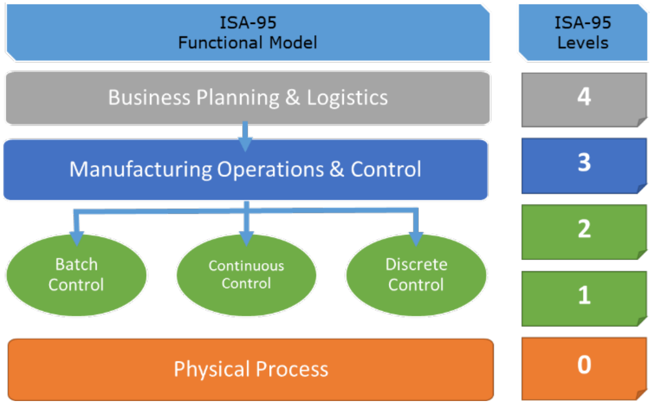


Figure 2. The ISA95 Functional Model (ISA95, Enterprise-Control System Integration, 2015)

What is important to note on these levels are the three key elements of functional operation shown at Level 1.

- batch control
- continuous control
- discrete control

Continuous control with analytics necessitates the ability for the analysis to effectively work in stream with the data but comes at the expense of not having a lot of the unstructured external content (historical data and reports). Similarly, enterprise centers of analytics at Level 4 have access to years' worth of real-time data that can be more fully explored for insight and enriched with unstructured content from a plethora of sources. The ISA95 model shows both discrete and batch levels of control. The conclusion is that we also need an analytical solution that can operate between levels 1 and 4. In this solution, we blend sensor data with unstructured content in an advisory role, which integrates with the human at the edge. This solution can then use the greater depth of knowledge available from the enterprise in a semi real-time mode. See Figure 3.

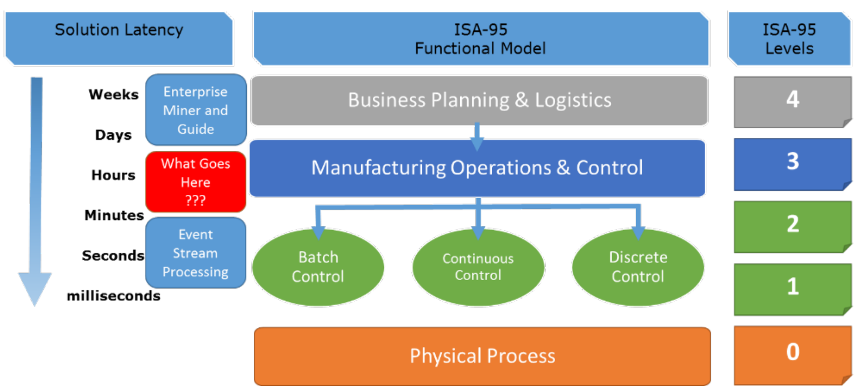


Figure 3. Solution Latency Needs versus ISA95 Levels

An example of this mid-latency real-time solution is in oil and gas downhole equipment. Consider the example of mud motor reliability. A mud motor is a device that provides increased rotational power to the bit as it cuts rock, as well as steering capabilities for wellbores that have horizontal sections. To optimize the performance of a mud motor, we need all of the current in-stream real-time data that is being acquired from the operation: the torque being applied, the weight on the drill bit, and more. But, we also need the operational specifications of this motor, what flow ranges can be used, how much dog leg it can create, and at what temperature the stators will begin to fail. Taking it further, what have been the optimal operational settings for this type of motor been in the past, and what has caused failures before?

The input data is varied and stretches across the ISA95 levels described above. However, the solution is not one that can, or indeed should, be run in continuous mode. First, humans are not especially great at thinking strategically at high speed. If we provide near real-time advice on the status of the motor, the human becomes reactive to its condition, rather than thinking strategically about how best to drill the next stand of drill pipe. Second, we need the extended information that is not in-stream in order to fully understand the performance capabilities of the motor. In this case, it is better for the advisory output to be batched at a time when the human is able to consider the advice and act strategically on it as the operation proceeds (as with a connection). Mud motors do not have instantaneous responses. Therefore, the solutions that advise on their performance need to consider not just the here and now but also the future state of both the motor and the wellbore being drilled. In Figure 4 we add a third element to our architecture where we consider cognitive levels of consumption (Hobbs & Shively, 2014).

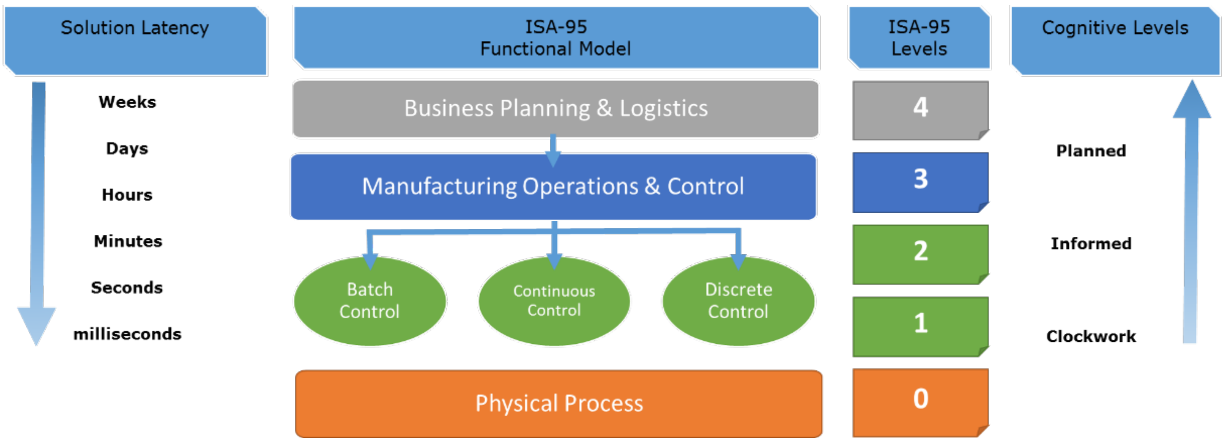


Figure 4. Solution Latency Needs versus Cognitive Levels

BUILDING AN EFFECTIVE ARCHITECTURE

So, what does a successful mid-tier architecture that can operate on real-time data at this mid-latency of solution look like? Consider the five elements of architecture shown in Figure 5.

Figure 5 Architecture for Mid-tier Analytics

DATA QUALITY AND INTEGRATION (1)

High-quality, decision-based analytics require a number of data management tools. Real-time data and, in particular, drilling data is messy. There are a multitude of papers on this subject. It is worth noting the key capabilities that a mid-tier analytics solution needs in order to work effectively with real-time data.

Missing values are a major issue because many of the communication methods used in a real-time operation do not guarantee a level of fidelity in measurements from the sensors. For example, the downhole sensors have to communicate to the surface via a method known as mud pulse telemetry (MPT). Owing to the medium and methodologies of transmitting data, and then acquiring it through MPT, the likelihood of missing values in the data stream can be relatively higher than we would like when using this data. To compound this problem, the process of drilling is mechanically noisy. Vibrations that can be in excess of 1G in force mean that the sensor measurements often contain bad values that we need to be able to manage in our analytical solution. The advantage of placing the analytics at a mid-tier level is that we can better manage the quality of this data across a window of time where more mature analytical techniques can be used to account for missing data through various forms of inference.

DATA ENRICHMENT (2)

Using our mud motor case as an example, there are several pieces of information required to optimize the performance of the operation. Firstly, the driller might be performing directional control of the wellbore using the mud motor. To do so, he needs a planned trajectory for the wellbore. As the wellbore is drilled, this planned wellbore might need to be changed due to operational conditions that arise. The analytical model has to be able to incorporate any changes to this plan. Further, what is the driller trying to achieve? Is he trying to reduce induced vibrations to extend the life of equipment? Or is he trying to finish the current hole section before the rig has to halt operations due to inclement weather?

Contextual information enrichment of the sensor data is critical for the analytics to be of value. Luckily there are products out there that can help us. For example, SAS® Decision Manager “delivers more precise, accurate, and traceable decisions into operational processes using analytics and business rules. SAS Decision Manager closes the loop for analytics, pushing advanced predictive analytic models into operations” (Taylor, 2014).

DATA MINING AND MODEL BUILDING (3)

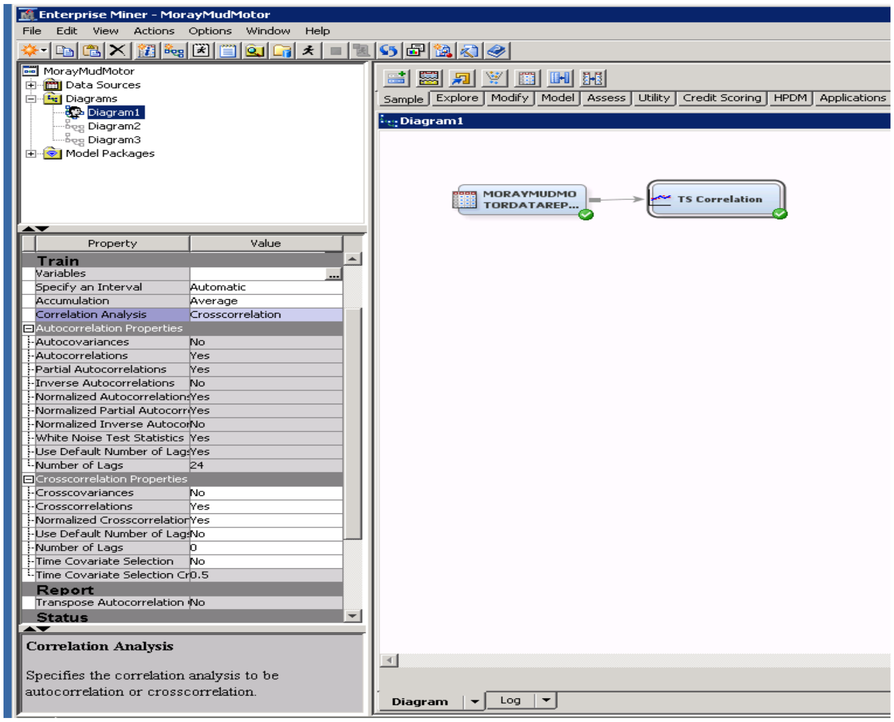
The heart of the solution requires a scalable analytic workbench that can incorporate sensor measurements and enterprise-level contextual information. Let us examine the key modeling techniques that enable us to build effective solutions that operate at the fidelity and scale that is appropriate to our operational need.

Building the Analytics

The three main purposes of analytics are descriptive, predictive, and prescriptive. Within each of these, the models can be time dependent or independent of time. While all models are appropriate for this paper’s topic, we present only a few of them to illustrate the types of analysis that are relevant for the data and business problem. Given that the drilling data is recorded over time, we present two time-dependent descriptive analyses and one time-independent predictive model.

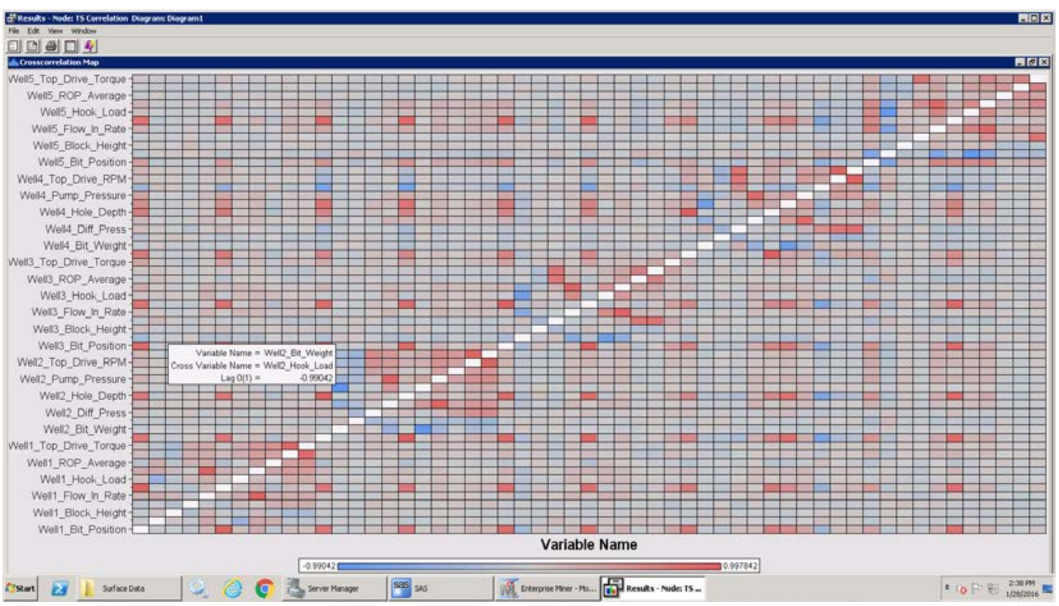
Time-Dependent Descriptive Analysis

Time-dependent descriptive analytic techniques include time series cross correlation, time series similarity analysis, and time series clustering. Each of the different drilling data elements is a time series, and we would like to measure the strength of correlation (positive or negative) among them. In Display 1, we see the SAS® Enterprise Miner process flow for a time series cross correlation analysis of data from five wells with zero lags. The Properties panel to the left shows the property settings we used for the analysis.



Display 1. SAS Enterprise Miner Cross Correlation Process Flow

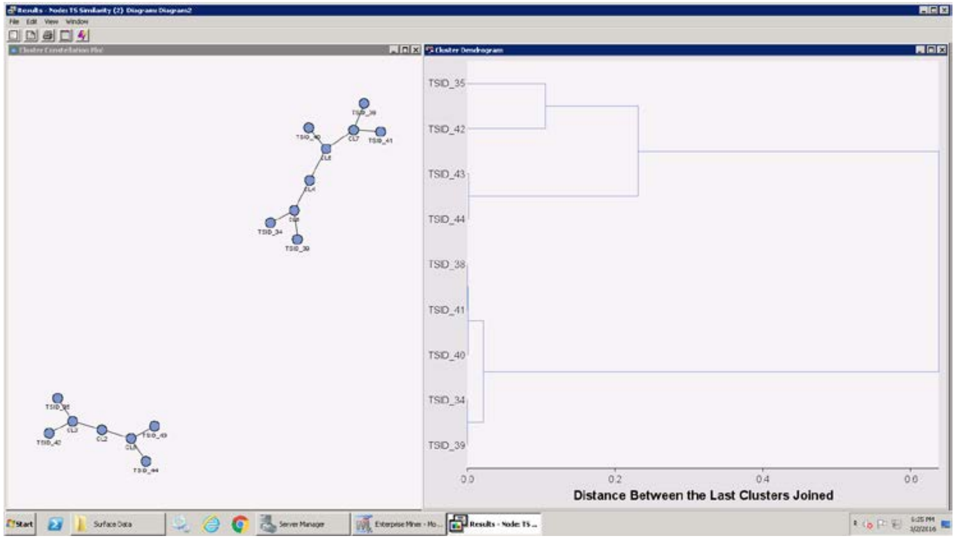
A time series cross correlation matrix is a common graphic method for presenting the results, as shown in Display 2. Note that the matrix is symmetric, and the color scale represents the strength of correlation. We have highlighted a data point in the matrix and its associated values.



Display 2. SAS Enterprise Miner Cross Correlation Results Matrix

From these results, we can see the brightly colored cells where certain pairs of time series are strongly correlated (either positively or negatively). This knowledge can help us to interpret highly similar or dissimilar behavior between any pair of drilling metrics. Looking across a row or column of the results can help us to determine which drilling metrics are highly correlated (positively or negatively) with all other recorded metrics for any given well or across wells.

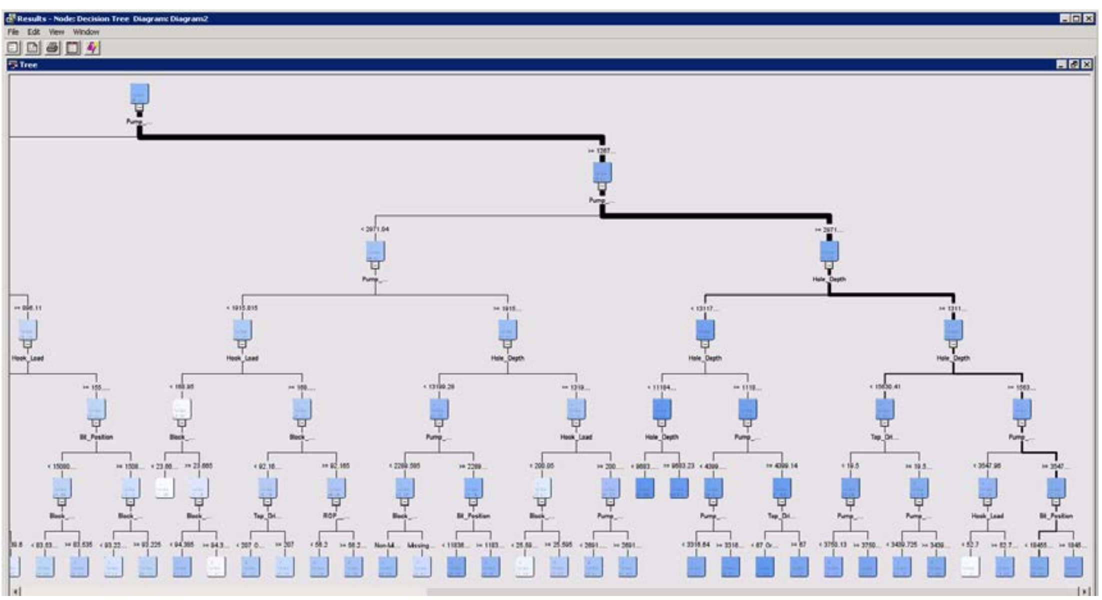
After determining that there appears to be correlation among some of the drilling metrics, it might be helpful to group “similar” drilling metrics into clusters using time series similarity analysis. This is a quantitative method that mimics how a human might perceive relative similarity among the patterns or shapes of time series data. Display 3 shows two results graphs from similarity analysis, a cluster constellation plot and a dendrogram plot.



Display 3. SAS Enterprise Miner Time Series Similarity Results Graphs

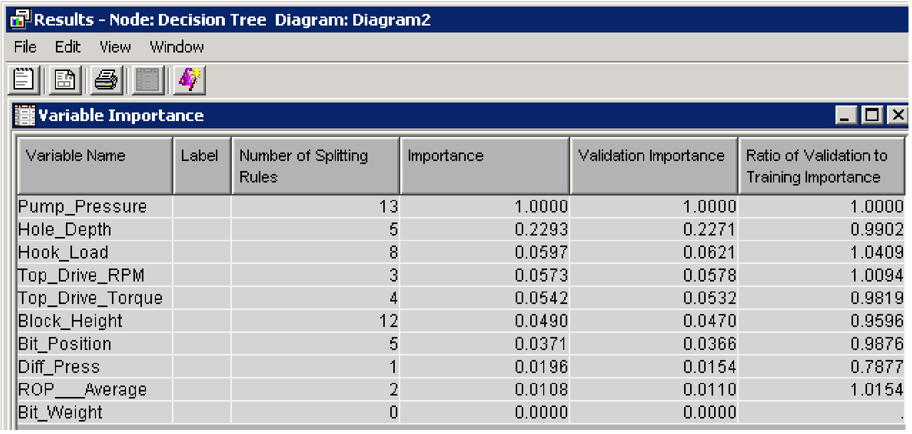
Time-Independent Predictive Analysis

Popular time-independent predictive analytic techniques include regression models, tree models, and neural networks, among others. In order to use these techniques, we need an objective. In the case of mud motor performance, we conduct a predictive modeling “tournament” for mud motor wear based on data sets that have been historically enriched with this information post-well. These models can then be used in our mid-tier advisory analytics to provide levels of prediction on motor wear as we drill the well. In this particular scenario, a decision tree was the model with the best predictive accuracy for motor failure. Display 4 shows part of the tree model.



Display 4. SAS Enterprise Miner Decision Tree Model Partial Results

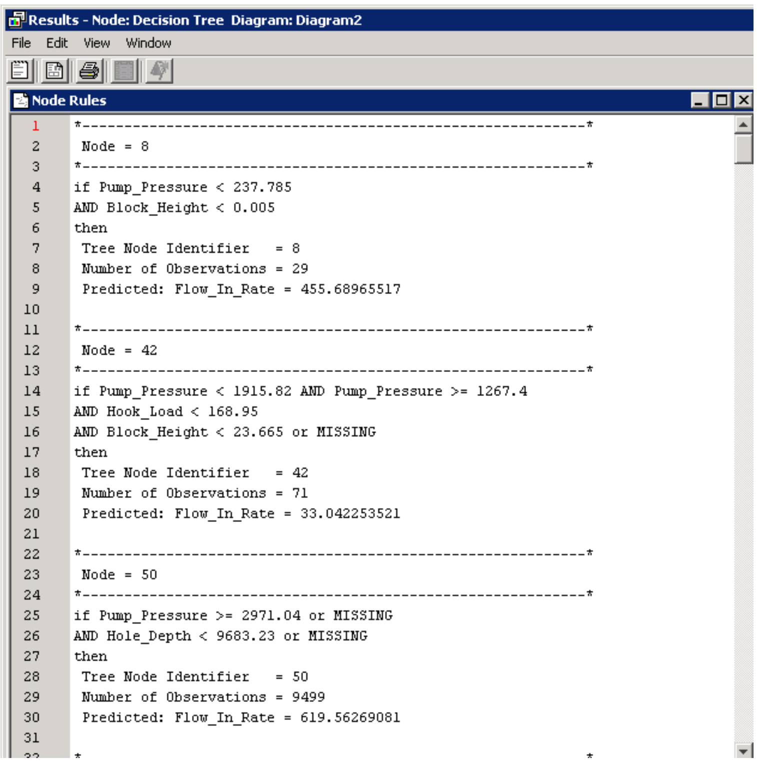
Also, in the decision tree model results in Display 5, we receive a summary of the relative importance of the independent variables in the model, sorted in descending order of importance.



Variable Name	Label	Number of Splitting Rules	Importance	Validation Importance	Ratio of Validation to Training Importance
Pump_Pressure		13	1.0000	1.0000	1.0000
Hole_Depth		5	0.2293	0.2271	0.9902
Hook_Load		8	0.0597	0.0621	1.0409
Top_Drive_RPM		3	0.0573	0.0578	1.0094
Top_Drive_Torque		4	0.0542	0.0532	0.9819
Block_Height		12	0.0490	0.0470	0.9596
Bit_Position		5	0.0371	0.0366	0.9876
Diff_Press		1	0.0196	0.0154	0.7877
ROP___Average		2	0.0108	0.0110	1.0154
Bit_Weight		0	0.0000	0.0000	.

Display 5. SAS Enterprise Miner Decision Tree Model Results Variable Importance Table

The decision tree model automatically produces scoring code in the form of if/then/else statements, do loops, mathematical expressions, and assignment statements. The resultant code as shown in Display 6 can then be used in a scoring application, and this scoring code can be deployed into our operational application for use by both field- and office-based personnel.




```
*------------------------------------------------------------*
 Node = 8
*------------------------------------------------------------*
if Pump_Pressure < 237.785
AND Block_Height < 0.005
then
 Tree Node Identifier   = 8
 Number of Observations = 29
 Predicted: Flow_In_Rate = 455.68965517

*------------------------------------------------------------*
 Node = 42
*------------------------------------------------------------*
if Pump_Pressure < 1915.82 AND Pump_Pressure >= 1267.4
AND Hook_Load < 168.95
AND Block_Height < 23.665 or MISSING
then
 Tree Node Identifier   = 42
 Number of Observations = 71
 Predicted: Flow_In_Rate = 33.042253521

*------------------------------------------------------------*
 Node = 50
*------------------------------------------------------------*
if Pump_Pressure >= 2971.04 or MISSING
AND Hole_Depth < 9683.23 or MISSING
then
 Tree Node Identifier   = 50
 Number of Observations = 9499
 Predicted: Flow_In_Rate = 619.56269081

```

Display 6. SAS Enterprise Miner Decision Tree Model Scoring Code Snippet

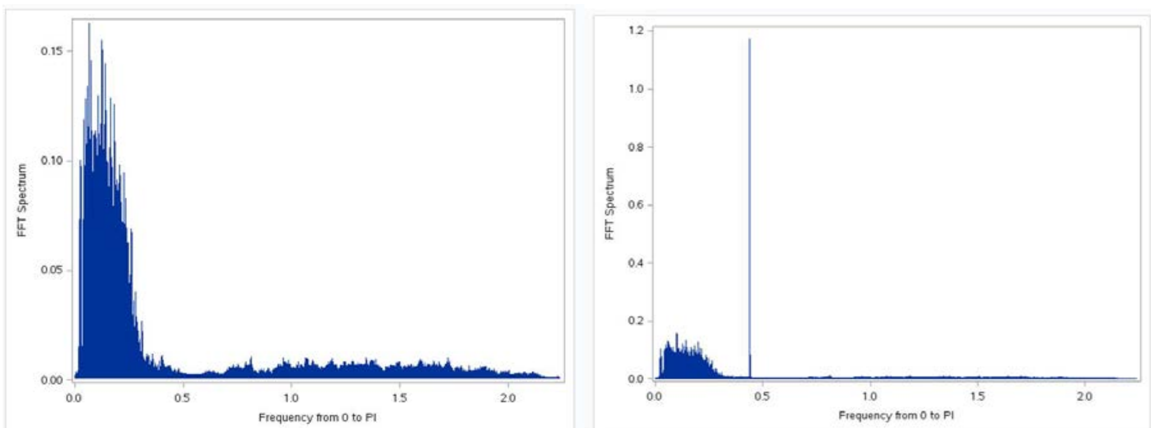
Time Series Decomposition

The authors would also note that the time series decomposition process is particularly useful when analyzing data from processes such as mud motor drilling in which there are regular cyclical events. Time series decomposition decomposes all sensor data to achieve trend, seasonal, and cycle components. The trend component shows the long-term increase or decrease in the values. The cycle component shows the cyclic pattern in the data. A cyclic pattern exists when values rise and fall but not over a fixed time period. A seasonal component shows the seasonal cycles (such as monthly, daily, or hourly) in the data.

Components of the time series decomposition can help us get more information about motor operation. For example, the trend component shows major trends in motor operation. Therefore, we can detect major misalignment or drift in the drilling operation and backtrack to the root cause, which is sometimes equipment calibration error. The cycle component gives insight into irregular vibrations or other short-term events in the motor, which can be helpful in detecting events during drilling. Online monitoring and scoring the cycle component using a predictive model can help us detect and prevent unwanted mud motor events in drilling.

Fast Fourier Transform (FFT)

Many motor applications use FFT spectral analysis as a handy tool to get information about motor operation. FFT transform refers to extraction of series of sine and cosine functions that, when superimposed, transform the signal from the time domain into the frequency domain. In doing so, we can discover hidden patterns in motor behavior that relate to frequency, such as vibration. Figure 6 shows an FFT transform of a normal mud motor on the left versus a mud motor with a high parasitic vibration on the right.



a **b**

Figure 6. FFT Transform of Mud Motor Measurements

OPERATIONALIZATION OF THE ANALYSIS (4)

Operationalizing the solution requires a careful evaluation of the operating environment, in particular the Human Machine Interface (HMI). In an article in the Journal of Petroleum Technology, it was noted that drillers can typically focus on only five different control elements at any one time (Rassenfoss, 2012). With this in mind, it's important that the operationalization does not add to the confusing stream of information being presented on the drill floor.

However, there is still a need to communicate the results of any analytical solution to the enterprise. In our example, the simplest way to achieve this is to embed the analytics into the existing solutions on the drill floor by exporting the model in native code such as C or Java. For the enterprise, the solution can be operationalized through SAS Visual Analytics. In both cases, there is a need to scale the analytical insight so that is more readily consumed.

In our example, a simple indicator that shows that there is a predicted problem with the mud motor can be displayed while drilling, with more detailed analytical information available to driller during the connection. From the enterprise, the same scenario applies. The office-based engineering team might want to see all operations within a geographic region and an indication of which ones have potential problems, as shown in Figure7.

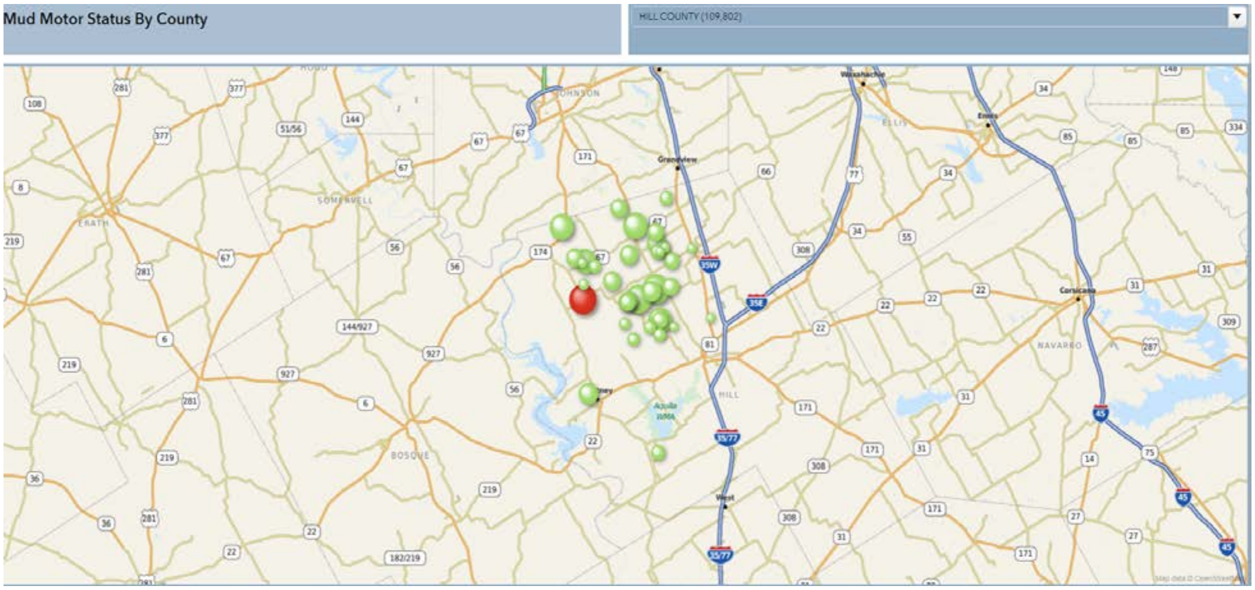
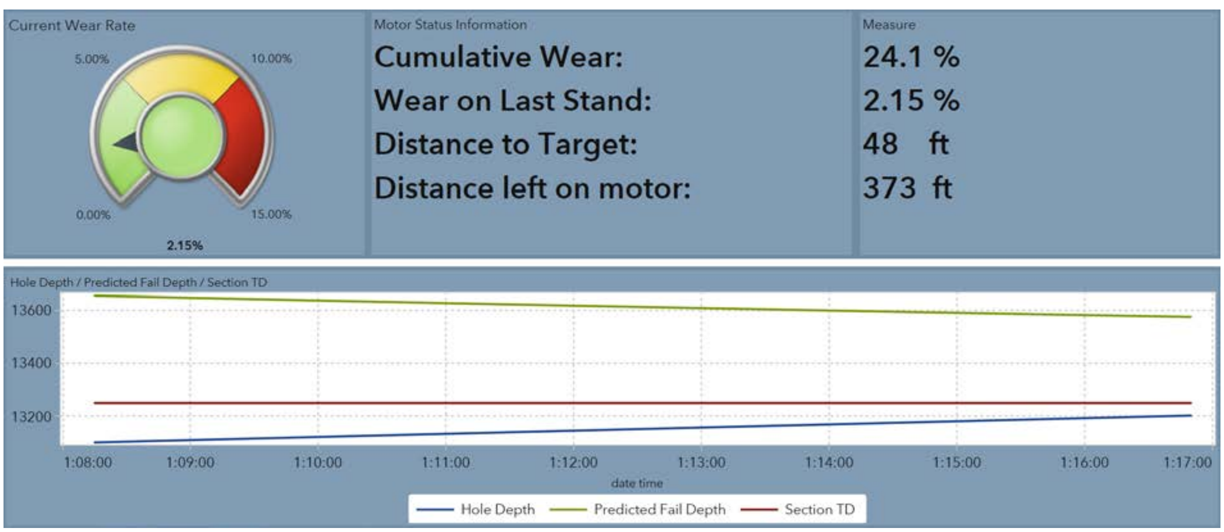


Figure 7. Office-Based View of Mud Motor Status on Several Drilling Operations in a County

However, a more detailed analysis needs to be made available on demand to the driller on the rig itself. The display needs to be informative and simple, and it needs to enable the driller to consider what his strategy for drilling the well going forward needs to be, based on the analysis. In displays 7 and 8, we show both a well-performing motor in green and a poorly performing motor in yellow. In the displays, we keep the information simple enough that the driller can quickly consume it during a logical break in operations, such as a connection. In this example, a gauge of current wear level, four key parameters of motor wear and time-to-failure relevant to the target, and a plot showing the target depth in red, with the wear curve and hole depth indicating whether the motor will fail before the driller reaches the target.



Display 7. Advisory Rig Display of Good Mud Motor Performance, Displayed during Connection

Display 8. Advisory Rig Display of Poor Mud Motor Performance, Displayed during Connection

MODEL MANAGEMENT (5)

There are two remaining issues that also need to be addressed. Firstly, as new data is received, the in situ models degrade. Secondly, the model will be deployed to multiple locations. In order for the architecture to remain valid and useful over the life of the operation, a model management solution also has to be implemented to ingest new data as it arrives and automate the process of keeping the deployed models up to date. To achieve this, we employ SAS® Model Manager. Display 9 shows an example of the comprehensive publishing environment for model delivery provided by SAS Model Manager, which supports the lifecycle and performance of the model (SAS, Time Is Precious, So Are Your Models, 2013).

Display 9. SAS Model Manager Dashboard

BRINGING ALL THE PIECES TOGETHER

Fortunately for us, all of the elements required in this architecture are available in SAS® Real-Time Decision Manager. Although this product is intended to support customer intelligence and real-time marketing campaigns, it features all of the core elements that we list above to build out our mid-tier analytics solution, but also includes a business workflow capability. As Figure 8 shows, we can now create a closed-loop analytical solution that seamlessly integrates into the drillers existing workflow. For more details, see the SAS Institute white paper, “SAS Decision Manager: A Technical Supplement” (Taylor, 2014b).

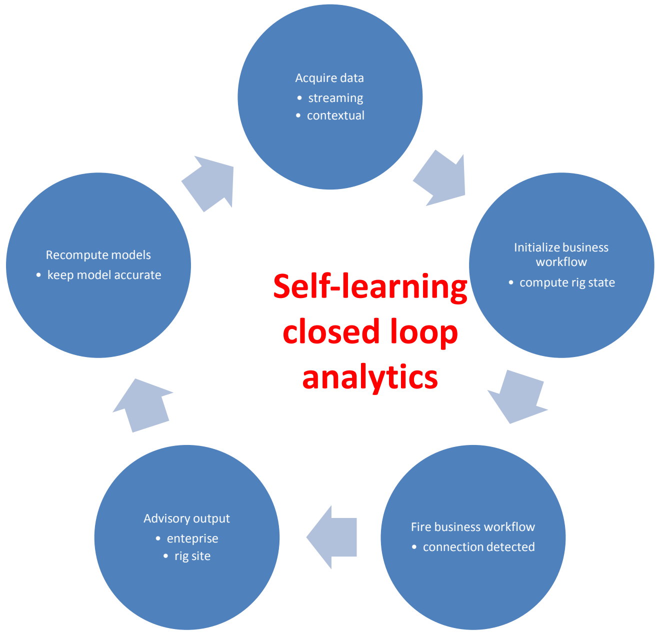


Figure 8. Closed-Loop Analytical Workflow

FINAL THOUGHTS

In this paper, we have bound our story of mid-tier analytics to a particular example from the drilling industry. Hopefully we’ve explained the advantages that a pseudo real-time analytics architecture brings, in particular to a chaotic or busy operation where the end user doesn’t always have the time to look at in-stream advice. In fact, as you start to look at the potential use cases, it becomes evident that you should always consider the solution fidelity as much as you consider the data fidelity. We leave you with some examples of this from the energy industry and encourage you to consider this paper the next time you build a real-time analytical solution.

ADDITIONAL MID-TIER EXAMPLES

- Monitoring for fluid breakthrough on partially instrumented production wells
- Optimal operating parameters to maximize artificial lift
- Batch analysis of pipeline pressure waves to predict future integrity issues.
- Operational controls for Amine treatment to prevent foaming and flooding.

REFERENCES

“Directional deviation tools.” PetroWiki. Available at http://petrowiki.org/Directional_deviation_tools.

Hobbs, A., and R. Shively. "Human Factor Challenges of Remotely Piloted Aircraft." NASA. Available at http://human-factors.arc.nasa.gov/publications/Hobbs_EAAP.pdf.

"ISA95, Enterprise-Control System Integration." International Society of Automation. Available at https://www.isa.org/isa95/.

Rassenfoss, S. 2015. "Drillers Find Themselves in a Tricky Spot at the Human/Machine Interface." Journal of Petroleum Technology 64:48-54.

SAS Institute Inc. SAS Institute white paper. "Managing the Analytical Life Cycle for Decisions at Scale." Available at http://www.sas.com/en_us/whitepapers/manage-analytical-life-cycle-continuous-innovation-106179.html.

Taylor, J. 2014. SAS Institute white paper. "SAS Decision Manager: A Product Review." Available at http://www.sas.com/en_us/whitepapers/sas-decision-manager-product-review-107272.html.

Taylor, J. 2014. SAS Institute white paper. "SAS Decision Manager: A Technical Supplement." Available at http://www.sas.com/content/dam/SAS/en_us/doc/whitepaper2/sas-decision-manager-technical-107273.pdf.

Wexler, Jonathan, Wayne Thompson, and Kristen Aponte. 2013. "Time is Precious, So Are Your Models: SAS® Provides Solutions to Streamline Deployment." *Proceedings of the SAS Global Forum 2013 Conference*. Cary, NC: SAS Institute Inc. Available at https://support.sas.com/resources/papers/proceedings13/086-2013.pdf.