Data-Driven Fault Detection for Neutron Generator Subsystem in Multifunction Logging-While-Drilling Service

A. Mosallam[1], L. Laval[2], F. B. Youssef [3], J. Fulton[4], and D. Viassolo[5]

[1,2,4] *Schlumberger, 1 rue Henri Becquerel, 92140 Clamart, France*
AMosallam@slb.com
LLaval@slb.com
JFulton@slb.com

[3] *Schlumberger, 42920 KLANG, Selangor, Malaysia*
FYoussef@slb.com

[5] *Schlumberger, 1430 Enclave Pkwy, Houston TX 77077, USA*
DViassolo@slb.com

Abstract

This paper presents a method for constructing a health indicator to detect neutron-generator faults in a multifunction Logging-While-Drilling (LWD) service and predict maintenance requirements due to wear. The method is based on extracting features from selected channels that hold information about the subsystem degradation with time. These features are used to build a decision-tree model which estimates the tool condition from the recorded data. The model demonstrates excellent value for both maintenance and field engineers due to the fact that in just a few minutes the physical condition of the neutron generator can be determined with high confidence. This work is part of a long-term project with the aim to construct a digital fleet management for drilling tools.

1. Introduction

EcoScope (Figure 1) is a multifunction Logging-While-Drilling (LWD) tool developed for oilwell drilling applications (Hansen & White, 1991). It is typically used in conjunction with other LWD equipment during the drilling phase of oil and gas extraction. It comprises an integrated suite of formation-evaluation, well-placement and drilling-optimization measurements into a single housing. The key differentiator that elevates this LWD service above all other competitors in the industry is the inclusion of a Pulsed Neutron Generator (PNG) which is a self-contained particle accelerator that produces neutrons using a fusion reaction. The PNG eliminates the need for the traditionally required Americium Beryllium (AmBe) chemical source. The PNG usage substantially reduces the transportation and wellsite health and safety risks and provides additional, advanced measurement (Figure 2). The PNG is fundamental to the LWD service provided to clients to deliver porosity, density and spectroscopy of the drilled formation. Yet it is an extremely complex piece of sensitive equipment, which is required to operate in particularly harsh environmental conditions.

Figure 1. Multifunction LWD service

A typical usage cycle is around 100 hours operating at temperatures up to 150degC and pressures of 20,000psi, with significant levels of shock and vibration from the aggressive cutting of rock at depths of several kilometers beneath the sea floor. Onboard sensors, electronics and memory acquire and store roughly 1000 channels of critical tool information which not only includes drilling and formation information for the client, but also tool diagnostic information for the Original Equipment Manufacturer (OEM). After each job, the tool is sent to a maintenance base where technicians use the acquired tool data to decide what level of maintenance is required before the tool is ready to perform the next job. A critical part of this decision-making process is analyzing the diagnostic information of the PNG and its subsystem to determine, whether the downhole conditions have had any detrimental effects on the operability.

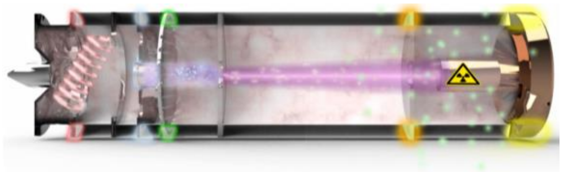

Figure 2. Pulsed Neutron Generator

Due to the complexity of the system, analysis of this vast amount of data is very time consuming and prone to error if performed manually. Furthermore, the electrical and physical complexity leads to a large number of potential failure modes, many of which are intermittent or only evident under the extreme stress of downhole conditions and are nearly impossible to reproduce in a maintenance base. Main reported failures are mechanical damage of PNG inside components, leakage, shortcut or open circuit and electronic boards soldering issues or component delamination. The PNG functionality is critical to the core tool measurements, so that any failure can have catastrophic consequences. Therefore, fault detection of the PNG is of utmost importance for operation (Zhan et al., 2010).

An automated diagnostic tool, which can determine the health of the PNG system with minimal user input, removes variability, eliminates human error and provides an efficient decision on the required maintenance in a fraction of the time. There are situations at the rigsite, where a LWD tool must be run two or more times without being sent back to the base for maintenance. In these cases, the field engineer must make the re-run decision, often under critical time pressure. Here, the fault detection application provides a useful aid to determine the tool's reliability for the next run (Isermann, 2006). The reliability benefits are clear and provide significant cost savings both for the client in terms of reduced Non-Productive Time (NPT) at the rigsite and for the OEM in terms of reduced Materials and Supplies (M&S) during maintenance and troubleshooting.

We present a data-driven fault detection method for the PNG subsystem in this LWD tool. The idea is to construct a Health Indicator (HI) from sensor data acquired from the PNG subsystem which can be used as a fault detection model. This method generates a statistics summary of selected channels to reduce computation time. The resulting features are used to extract the first principal component (eigenvalue). Then empirical mode decomposition algorithm is used to decompose the generated first principal component into successive intrinsic mode functions and a residual signal (Huang et al., 1998). The residual signal is retained because it shows rate of change in variance with time and therefore the developing degradation of the subsystem. A decision-tree model is trained on HIs of different runs labeled as healthy or faulty by a domain expert.

This paper is structured as follows. Section 2 presents a description of the PNG subsystem. The method and the results for the fault detection model are presented in section 3 and 4, respectively. Finally, section 5 concludes the paper.

2. PULSED NEUTRON GENERATOR (PNG)

Generators of high energy neutrons have been used for a long time for neutron-gamma ray or neutron-neutron logging (Tittle, 1961). A neutron generator has multiple advantages over traditional chemical sources. The ability to turn off the PNG means there is a negligible radiation risk when the generator is not downhole in a well. A higher yield of neutrons is available, enabling better measurement statistics, and the ability exists to control the yield of neutrons. The yield of neutrons from a generator can be either in bursts or continuously. Commonly in logging operations, the neutron generator has a controllable yield of neutrons in burst. PNGs are used in wellbore formation evaluation tools to evaluate how neutrons interact with the drilled formation. Porosity, salinity, formation density, hydrogen content, formation elemental fractions, etc., can all be determined by measurements of interaction products of high energy neutrons from a PNG with such formations.

The PNG is a self-contained particle accelerator that produces neutrons using a fusion reaction (Figure 3). A high-voltage potential accelerates ionized deuterium and tritium isotopes of hydrogen toward a target doped with tritium.

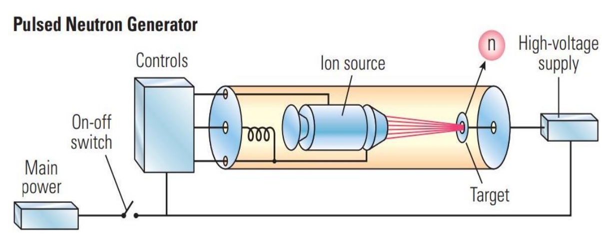

Figure 3. Architecture of PNG

The fusion reaction produces a 4He nucleus and a neutron as shown in Figure 4. The reaction energy is transferred into the kinetic energy of the two particles and is dissipated as heat when the particles are stopped in matter. The neutrons leave the reaction with very high speed, having kinetic energy of approximately 14.4 MeV of the total 17.6 MeV released. When the main power is disconnected, the PNG produces no neutrons.

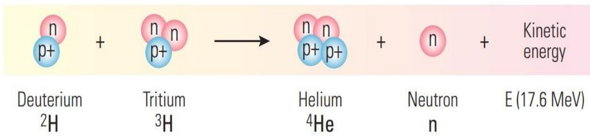

Figure 4. Fusion reaction inside the PNG

Most nuclear logging measurements are carried out by emitting pulses of neutrons which irradiate the earth formations, and by detecting the radiation (neutrons or gamma rays) resulting from the interaction of earth formation atoms and the emitted neutrons. Thus, it is critical to have a good knowledge of the characteristics of the neutron pulse, such as the neutron output (number of neutrons emitted) and the pulse timing. Such knowledge means having control over these characteristics. It is highly desirable to generate neutron pulses having a substantially square shape with a short rise time (to reach the plateau value) and a short fall time (once the voltages are turned off). These features require the following elements:

- Neutron detector to control the output neutron flux
- Electronic timing board for pulse output neutron flux shaping
- Three electronic power supply boards to permit nuclear fusion reaction
- Electronic acquisition board to control the neutron output flux

The PNG subsystem consists of the generator, a monitor detector, a timing board, acquisition board and three power supply boards as presented in Figure 5.

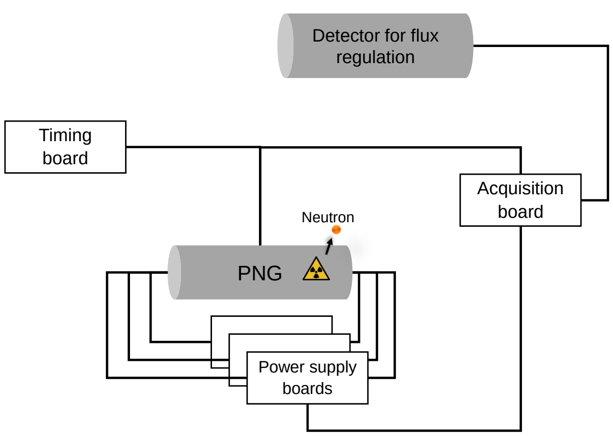

Figure 5. Pulse neutron generation subsystem

Known failure modes for the PNG subsystem are:

- High Voltage (HV) leakage due to improper insulation
- Broken inside components as filament or cathode
- Harnesses discontinuity
- Damage or broken component on acquisition boards or power supply boards
- Inefficient or broken detector for neutron flux regulation
- Wear of the target

Such failures modes affect the overall function of the PNG and lead eventually to failure. In the next subsection we present a method to extract HIs from selected channels that hold information about degradation of the PNG. Such HIs are then used to build a machine learning model to estimate the physical condition of the PNG subsystem.

3. PROPOSED METHOD

The idea is to construct a health indicator from recorded channels that hold information about the PNG subsystem degradation. This health indicator is then modeled using a decision tree model, which can be used to discriminate between healthy and failed tools. The method is divided into four main steps presented in Figure 6.

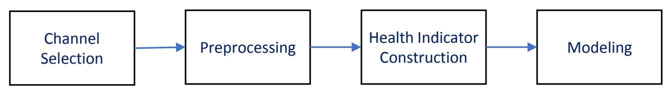

Figure 6. The method's general scheme

3.1. Channel selection

After each run acquired tool data contain an immense number of data channels generated at a record rate, which results in millions of data points from a single run. Not all of those channels hold information about the degradation of the PNG subsystem over time. Removing channels that hold no information about the degradation should result in a better HI while increasing the efficiency of the algorithm. Domain knowledge of nuclear physics and nuclear instrumentation determines the channels that contain relevant information about the degradation. Therefore, Subject Matter Experts (SME) select the most important channels for the health status. For the PNG subsystem the following channels are selected:

- Detector output
- Input and output currents
- Input and output voltages

These channels shall be used to construct the HI; the remaining channels can be ignored.

3.2. Preprocessing

The LWD data-acquisition system starts recording data once the engineers initialize it for the upcoming job. The LWD tool goes through the following steps:

- Tool initialization - The engineer configures the acquisition parameters for the upcoming job, formats the tool memory, and begins the tool recording.
- Shallow hole test - A confirmation that the tool is functioning as expected inside the well before deploying to full well depth.
- Casing logging for caliper calibration - Using the known internal diameter of the metal casing connecting the rig to the wellbore and the known drilling fluid properties, the tool calibrates its ultrasonic measurement before descending into the drilled well.
- Drilling operations - The physical drilling of a well with the measurement acquisition equipment situation immediately behind the drill bit.

Once the tools is initialized and in the well the onboard system records measurements every 2 seconds. This data is not available until the job is done and the tool is back to the surface. Data collected during the steps before drilling operations contain millions of data points a lot of them can be irrelevant information about the degradation of the PNG and are removed. After removing irrelevant data, a moving window is computed to summarize the data. The window size is determined based on the job length. The resulting signal contains 300 samples for the selected channels. The reason behind reducing the number of samples is to reduce the processing time required to extract HI as will be explained in the next section.

3.3. Health indicator construction

To construct the HI from preprocessed signals we used a method proposed by (Mosallam et al., 2013). Three steps are applied to the preprocessed channels. The health indicator is calculated incrementally: it starts by the first two samples and then adds the next on and so on. The result of this step is a HI which consists of 299 samples. The details of the algorithm are explained below.

Channel compression: this step has two goals: 1) extract representative features and 2) reduce the dimensionality of the selected channels into a single channel. Standard Principal Component Analysis (PCA) method is applied on all channels within every window:

$$Cv_i = \lambda_i v_i \tag{1}$$

where C is the covariance matrix, λ_i are the eigenvalues, and v_i are the eigenvectors of the processed channels. The first principal component retains the maximal variance while reducing the dimensionality to one dimension (Jolliffe, 1986). Therefore, only the first principal component is used to represent the health-status evolution with respect to time.

Trend extraction: in this step the algorithm extracts a trend from the compressed signal to represent the degradation in a simple monotonic signal using Empirical Mode Decomposition (EMD) algorithm. EMD is a method employed to decompose a signal into successive Intrinsic Mode Functions (IMF) and a residual signal (Huang et al., 1998). The EMD algorithm performs the following steps:

- find all local maxima and minima of the input signal and compute the corresponding upper and lower envelopes using cubic spline, respectively.
- subtract the mean value of the upper and lower envelopes from the original signal.
- repeat until the signal remains nearly unchanged and obtain IMF_i.
- remove IMF_i from the signal and repeat the previous steps.
- stop when the generated residual $r_n(t)$ is a constant or a trend.

The residual $r_n(t)$ of this process should be a constant or monotonic signal that can be represented as:

$$r_n(t) = v_1(t) - \sum_{i=1}^{n} imf_i(t) \tag{2}$$

where $v_1(t)$ is the input first principal component, imf_i is the IMF and n is the maximal number of IMFs. The residual can be used as a way to characterize the degradation of the PNG. A constant residual could be due to no change happening at this particular point. However, a residual following a trend could be due to degradation change happening as described by (Mosallam et al., 2016).

Feature extraction: the last step of the algorithm extracts statistical features from the generated residuals (Mosallam et al., 2012; Witten & Frank, 2005). Such features represent each residual in one dimension as variance σ^2 of each residual signal:

$$\sigma^2 = \frac{\sum_{i=1}^{n}(r_n - \mu)^2}{n} \tag{3}$$

where μ is the mean of the residual signal r_n and n is the number of samples in the residual signal.

3.4. Modeling

Each job used to build this model is labeled first as either healthy (1) or faulty (-1) by the SME. A decision tree classification model is used in this work (Bishop, 2006). The model is trained to map the relation between the input HI and the corresponding label (x, Y) where:

$$x = [x_1, x_2 ....x_{299}] \tag{4}$$

and

$$Y = \begin{cases} 1 & where\, x\, is\, healthy \\ -1 & where\, x\, is\, failed \end{cases} \tag{5}$$

4. RESULTS

The first step in this work collects LWD run data to build the model. We collected around 200 run files from different locations with different environmental conditions to build the fault detection model. For each run, the SME reviews the corresponding maintenance reports and checks the raw data to label it as healthy or faulty. The SME also selects which channels contain information about the degradation and should be used to build the model. The algorithm then starts by preprocessing the raw data. The idea is to remove irrelevant data and to summarize the raw data. Figure 7 shows the result of preprocessing one raw channel.

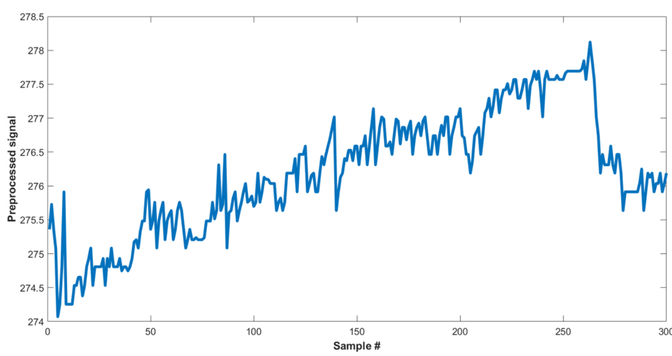

Figure 7. Results of preprocessing one raw channel

Figure 7 shows that the number of generated samples is 300 regardless of the run length. In this way the computation needed to construct HI is dramatically reduced. Next, the method constructs the HI from preprocessed data. The HI construction consists of three main steps. The first step computes the first principal component of the first two samples of the preprocessed signal. Then the EMD residual is extracted from the first principal component. Finally, the variance is calculated from the residual. The third value from the preprocessed signal is added to the first two samples and the previous three steps are repeated. The result of this recursion is a HI consisting of 299 samples. Figure 8 shows the HIs for four different healthy runs. As can be seen HIs for healthy runs did not exceed 4000.

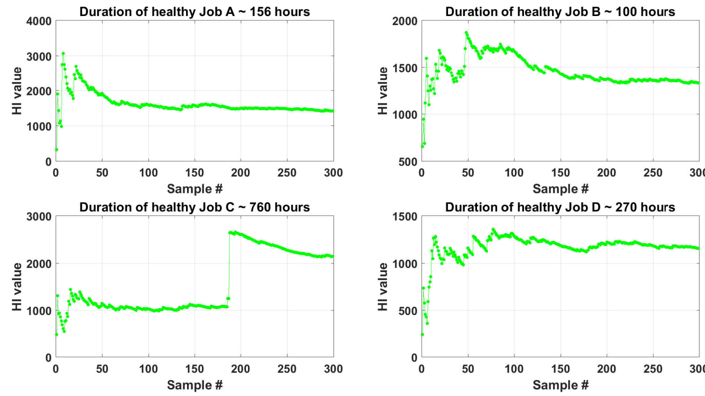

Figure 8. HI results of four different healthy jobs

Figure 9 shows failed runs. HIs for failed runs started on very low values and then it had sharp increase. HIs for failed runs can reach up to 8×10^6. Also, in many cases we noticed that the point when the HI start increasing can reflect the begging of the failure on the raw data.

Figure 9. HI results of four different failed jobs

HIs are generated from the data set to build a decision tree classification model. Healthy runs are labeled as 1 and faulty runs are labeled as -1. Decision trees are then used to train a model using the input HIs and their corresponding labels. Figure 10 shows distribution of mean values calculated for healthy and faulty groups separately. The plot shows that mean value of healthy HIs are much lower than failed HIs. Also, the majority of HIs mean values for faulty runs are centered around one value whereas the healthy mean is spread between $10^{22.65}$ to $10^{3.6183}$.

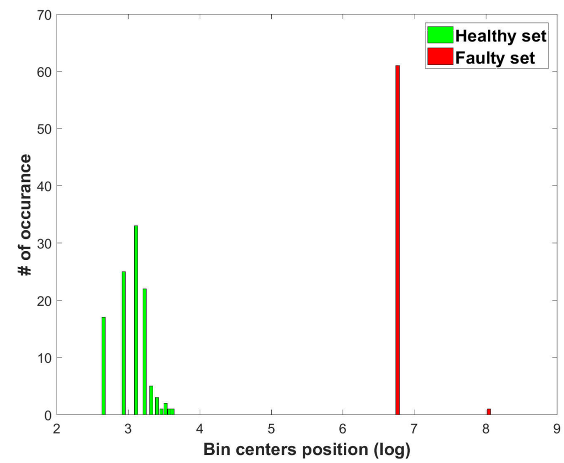

Figure 10. Histogram of mean values of healthy and faulty HI in the training set

Figure 11 shows distribution of standard deviation values calculated for all generated HIs. Standard deviation values of healthy HIs are much lower than faulty HIs. The reason is that healthy HI values are not spread out from the mean. HIs for failed run start rapidly increase after failure begines during the job.

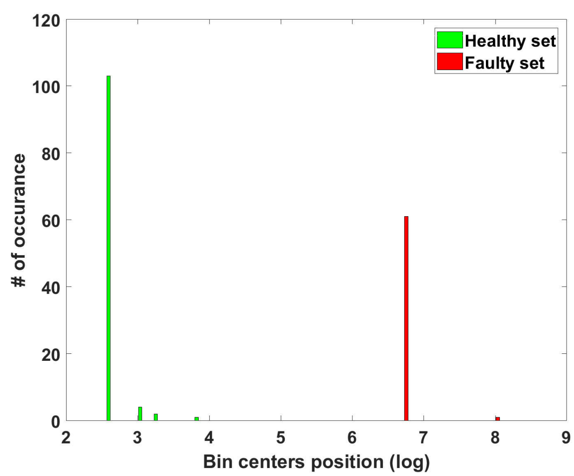

Figure 11. Histogram of standard deviation of healthy and faulty HI in the training set

We validated the model on a test data set of 60 new run files, which we did not use for training the model. The test data set consists of 25 failed and 35 healthy files. The model shows high confidence with only one misclassification and an accuracy of 98.33% (see Table 1).

		Correct label		
		Healthy	Faulty	Total
Predicted label	Healthy	34	1	35
	Faulty	0	25	25
	Total	34	26	60

Table 1. Confusion matrix for model testing

5. CONCLUSION

This paper presents a data driven method for fault detection of a nuclear generator subsystem. The method builds on constructing representative features that can be used as health indicators. Such health indicators are modeled using a decision tree classifier. The model was validated using operation data that was not used to train the model. The results show that the model can discriminate between failed and healthy runs with high accuracy. Also, the model demonstrated excellent value for maintenance and field engineers, because the physical condition of the neutron generator can be determined with high precision and in just a few minutes using run data.

REFERENCES

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Secaucus, NJ, USA: Springer-Verlag New York, Inc.

Hansen, R., & White, J. (1991). Features of logging-while-drilling (lwd) in horizontal wells. In *Spe/iadc drilling conferencespe/iadc drilling conference.* doi: 10.2118/21989-MS

Huang, N., Shen, Z., Long, S., Wu, M., Shih, H., Zheng, Q., ... Liu, H. (1998). The empirical mode decomposition and the hilbert spectrum for nonlinear and non-stationary time series analysis. In *Proceedings of the royal society of london series a. mathematical, physical and engineering sciences* (p. 903-995).

Isermann, R. (2006). *Fault-diagnosis systems: An introduction from fault detection to fault tolerance)*. Heidelberg: Springer-Verlag, Heidelberg.

Jolliffe, I. (1986). Principal component analysis and factor analysis.. doi: 10.1007/978-1-4757-1904-8-7

Mosallam, A., Medjaher, K., & Zerhouni, N. (2012). Unsupervised trend extraction for prognostics and condition assessment. *IFAC Proceedings Volumes*, *45*(31), 97 - 102. (2nd IFAC Workshop on Advanced Maintenance Engineering, Services and Technology) doi: https://doi.org/10.3182/20121122-2-ES-4026.00014

Mosallam, A., Medjaher, K., & Zerhouni, N. (2013). Nonparametric time series modelling for industrial prognostics and health management. In (Vol. 69, p. 16851699). doi: 10.1007/s00170-013-5065-z

Mosallam, A., Medjaher, K., & Zerhouni, N. (2016). Data-driven prognostic method based on bayesian approaches for direct remaining useful life prediction. In (Vol. 27, p. 10371048). doi: 10.1007/s10845-014-0933-4

Tittle, C. W. (1961). Theory of neutron logging i. *GEOPHYSICS*, *26*(1), 27-39. doi: 10.1190/1.1438839

Witten, I., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Zhan, S., Ahmad, I., Heuermann-Kuehn, L., & Baumann, J. (2010). Integrated pof and cbm strategies for improving electronics reliability performance of downhole mwd and lwd tools..

BIOGRAPHIES

Ahmed Mosallam is a data scientist at Schlumberger technology center in Clamart, France. He has his PhD. degree in automatic control in the field of Prognostics and Health Management (PHM) from University of Franche-Comté. His main research interests are signal processing, data mining, machine learning and PHM.

Laurent Laval is a physicist at Schlumberger technology center in Clamart, France. He has an engineering degree in physics from Ecole National de Physique de Grenoble. His interests include software development, failure analysis, condition based maintenance, prognostics and health management.

Fares Ben Youssef is an electrical engineer at Schlumberger Asia Center for Reliability and Efficiency in Malaysia. He has a Master degree in electrical engineering from the engineering school of Paris XI University. His interests include electrical board and component failure analysis, equipment efficiency, condition based maintenance, prognostic and health management.

James Fulton is a reliability engineer at Schlumberger technology centre in Clamart, France. He has a first class Masters degree in aeronautical engineering from the University of Glasgow. His interests include design for reliability, condition based maintenance and failure analysis.

Daniel Edgardo Viassolo is a Principal and global PHM Analytics manager with Schlumberger's Enterprise Solutions organization. Daniel has made impactful contributions to the field of Industrial Asset Health Management and Controls in diverse applications domains - oil & gas services, renewables, jet engines. He co-authored 19 US patents, 30+ publications and 1 book. Daniel obtained his PhD from Purdue University.