Putting Your Data to Work: recent experiences in driving marine operational excellence & asset management

Subrat Nanda
American Bureau of Shipping, Houston. TX.
snanda@eagle.org

Abstract

Advancements in the field of data science are presenting new opportunities for marine fleet operators to adopt a more effective asset management strategy, that combines advanced data analytics with maintenance and operational experience, to achieve a reduction in unplanned downtime and help realize fleetwide efficiencies. These methods are used to quantify the reliability of maritime assets and drive improved decision making for fleet operations, emergent risk identification, and ultimately improved operational availability and flexibility. This also enables operators to move beyond traditional calendar-based philosophy into condition-based paradigm, where maintenance efforts and classification are driven by condition of equipment, making them more targeted and optimized for scope and timing. Key in this is anomaly detection to distinguish early onset of failure conditions. This further ensures operating costs reduction, maximum equipment life utilization and total life-cycle optimization. We will describe essentials of condition-based class, equipment health assessment methods using anomaly detection and additional considerations for deployment. Finally, we will describe recent case studies in these areas as well as challenges in applying them to real world operations.

Keywords: *Condition-Based, Anomaly Detection, Classification*

Section 1: Introduction

Marine fleet operators are challenged to reduce unplanned failure expenditures. Recent data science applications have enabled marine fleet operators to adopt an effective asset management strategy – combining advanced data analytics with marine-specific maintenance and operational experience to reduce unplanned downtime. For this paper, the terms asset and equipment are used interchangeably, as the concepts discussed apply equally from a high-level asset, vessel or offshore platform, to equipment or components contained therein.

This paper is organized as follows. We discuss the concept of condition-based paradigm especially relating to the condition-based class model in Section 2. We describe the relation between condition-based maintenance and condition-based class, how it relates to what classification societies are evolving towards, and the data needed for such a transition. Section 3 discusses in detail the traditional maintenance practice and explores factors precipitating the move from a calendar-based to a more condition-based thinking. We discuss some historical reasons for continuing using a calendar-based approach and the advantages of moving to condition-based. We then provide an anomaly detection overview and the important role it plays in modern maintenance practice. We describe the major elements, challenges and lessons learned, and include a brief survey of major data-based methods being developed and used. We also discuss the overall process for implementing an anomaly detection process. Finally, we conclude with a summary of current progress, and future road map to tie data-driven condition-based maintenance practices to condition-based class in Section 5.

Section 2: Condition Based Class

Classification societies play an important role in helping to ensure the overall risk profile for operating assets is kept at acceptable limits, such that the safety of life, property, data, etc., is never jeopardized. This necessitates operators maintain a classification requirement and periodically obtain renewals post detailed surveys and inspections. Traditionally, this activity follows a calendar-based model with requisite ‘open and inspect’ centered process for certain equipment.

Condition-Based Maintenance (CBM) is a methodology for implementing a maintenance strategy, wherein maintenance intervention is performed when needed. This is determined by evaluating various indicators of asset performance or life [1]. The need for CBM arises due to a variety of factors, such optimizing costs for maintenance, and reducing major overhaul downtime (leading to high costs over the asset life-cycle) and sub-optimal management and use of parts and spares. The evolution of maintenance strategy has followed the path from corrective maintenance to preventive maintenance to condition-based maintenance [11]. CBM offers the ability to detect incipient failures, enabling overall asset life extension and optimal replacement strategy development. Over the long run, this results in decreased maintenance cost and reduces unplanned downtime. CBM however is not a strategy to prevent machines from degrading or stopping failures completely [10], [12].

With the development of advanced methods to understand the condition of equipment using multiple types of data collected, a condition-based class model (CBCM) is emerging whereby asset surveys are more targeted and efficient, less intrusive, and improve the overall experience for operators while ensuring high safety standards for classed assets. This requires aligning the classification requirements to evolve from calendar-based to condition-based, as depicted in Figure 1. Moving to a condition-based class enables customized survey planning, making it tailored to a specific asset, resulting in increased efficiency in the classification process while maintaining overall risk and safety metrics.

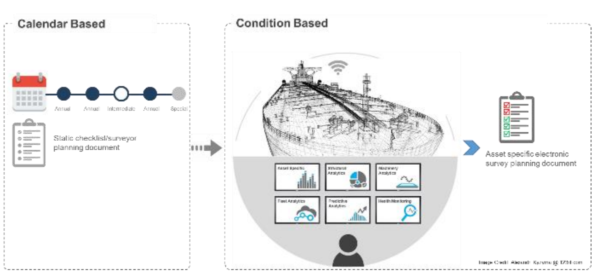

Figure 1: Condition-Based Class Model

To build such capability, a data model is required to capture, aggregate (data management) and integrate diverse data types from an asset over its design, operations and service history. Some data types used to develop the CBM process are as follows:

a) sensor data: time series, e.g., temperature
b) maintenance logs: transactional
c) inspection reports: digitized reports
d) design changes: and impact
e) survey reports
f) data generated from wearables and drones

Section 3: Traditional Maintenance Practices and the case for Condition-Based

Current maintenance practices for a major part of the industry are based on preventive recommendations driven by Original Equipment Manufacturers (OEM), which are typically determined based on testing during manufacturing resulting in time-based preventive maintenance schedules. This means equipment would be operated until preventive maintenance intervals, measured in calendar time, cycles or other such units of interval measurement, and then replaced. There are various advantages to this, primarily lower expected departures from set processes (scope, etc.), higher certainty about time and scope of maintenance activities, lesser requirement for real time sensor-based equipment understanding, and more certainty of associated costs, like installation or connectivity.

However, this approach poses other challenges such as high fixed costs of planned outages, lack of warning about emerging and insipient failures while operating, and sub-optimal parts replacement. The fixed scope of such maintenance also causes high base cost of operation driven by parts and spares.

Until recently, some reasons for continuing the use of traditional maintenance practices have been the lack of real-time connectivity, data acquisition, computing, digital technologies and intended design departures. However, today's equipment has hundreds of sensors on-board that sense or derive an even larger number of parameters such as temperatures, pressures, etc. This combined with high speed connectivity due to low-cost high-speed internet, or in extreme cases satellite data transmission, has provided an opportunity to use the large quantity of data generated continuously. Today, high resolution data storage, transmission and computing make it possible to perform large scale calculations – not possible until a few years ago. The ecosystem of the Internet-Of-Things (IoT) and digital technologies play a vital role in making the 'data flow' work seamlessly. The final aspect mentioned above is intended design departures. Equipment usage has mainly stayed within the original design intent. However, with economic changes, many equipment manufacturers and operators see a departure in usage, operating conditions and operator skill. This leads to sub-optimal maintenance schedules when using traditional methods, and further defines the need to develop condition-based maintenance practices.

Section 4: Overview of Anomaly Detection

Detection of an abnormal condition triggers further analysis that can confirm potential equipment damage, or characterize it as either a degraded sensor, or other data quality related issue. An anomaly can potentially arise due to several factors such as changes in operating conditions, modes, sensor failures, or actual equipment damage. The aim of anomaly detection as a process is therefore to pinpoint anomalous behavior and subsequently isolate if change detected is indeed an initiation of equipment damage, indicating a presence of damage, versus a data artifact due to other reasons unrelated to asset health or life [14].

Figure 2 illustrates an overview of the basic anomaly detection flow. Data from the equipment of interest is fed to an anomaly detection engine, which includes the definition of a 'normal' pattern. This sense of what constitutes a 'normal' is derived or learned from the data, by analyzing the correlations and relations between multiple variables or single parameters at a time, and their various states under multiple operating conditions. According to Hawkins, anomaly is defined as something that *arouses suspicions of a different mechanism* [13].

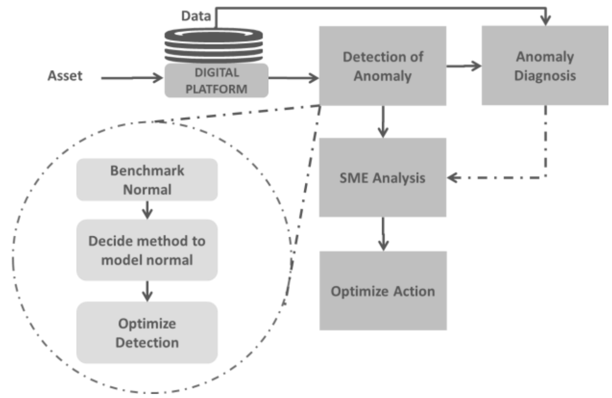

Figure 2: Anomaly Detection Process Overview

The next decision in the design process is choosing a technique to perform anomaly detection. Most methods available in open literature are one of the following broad categories: namely, supervised or unsupervised methods. Unsupervised methods find patterns in data by identifying commonalities among sub-groups of the data, which is unlabeled. As implied, supervised methods usually require a labeled historical data where past anomalies of interest are identified and categorized into root causes under various operating conditions [6], [7]. For either of the above two broad methodologies, a decision boundary to identify an anomaly. This is done using various optimization methods to minimize the risk of Type I and Type II errors (risk of making a false positive or false negative decision errors) and validated using historical data. Further details are mentioned in [2], [3], [4], [6], [7].

To identify anomalies in operational data, both univariate and multi-variate approaches are used. For complex equipment such as engines, pumps, etc., using multivariate methods using more than one measured or sensed parameter are found to be more robust as they account for different operating modes and interaction between different parameters.

To construct these models, a model of being 'normal' must be constructed and subsequently a measure of the 'distance' to normal must also be constructed – to identify a potential anomaly. Therefore, most methods calculate a real-valued outlier score to obtain a quantified estimation, from which a normal determination of a data point or observation is made. The following presents a summary of the methods used for anomaly detection [2], [14]:

a) Model-based Methods: if a data point does not fit an assumed model of known data, it can be announced as abnormal. Models that summarize data such as regression models, probability distribution models, or cluster models are employed to detect anomalies in data. For example, if the hypothesis is testing if two sets of data came from the same underlying assumed probability model, a test for anomaly can be constructed, such as using likelihood ratio test. Even if the data do not come from the assumed distribution, such as the multi-variate normal, these tests are still effective in pointing to regions of interest.
b) Density-based methods: methods finding natural 'clusters' of related data also detect data points, which are not part of known clusters. Regions in the data space, with sparse density surrounding them can point to potential anomalies.
c) Distance-based methods: various techniques to determine the distance between two data points or sets of data have been used to develop methods for detecting anomalies. For example, to test if a test data point occurs at the extreme of a probability distribution, a measure of the distance from a known distribution such as Mahanolobis or Hotelling-T squared (multi-variate) distance have been used.

In Figure 3, data are taken from three sensors, measuring different parameters of interest, onboard an asset propulsion system. These parameters of interest are identified as affecting the performance (shown here as Parameter 1, Parameter 2 and Parameter 3) over time, based on knowledge of equipment operation – both normal operations and several failure events. A combination of different anomaly detection methods, probability model and distance-based methods are used to detect anomalies in multiple variables simultaneously. The outputs are then combined using a weighted scheme to announce the presence of an anomaly with confidence.

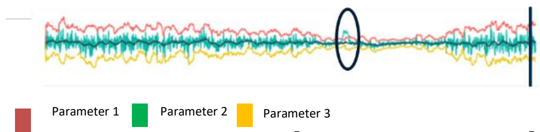

Figure 3: Anomaly Detection Example

As shown in the highlighted elliptical area, the aim of the anomaly detection process is detecting these 'early' warning movements in data that could point to a potential event, shown here by the vertical line. The time on the x-axis between the detection of the event (shown by the ellipse) and the event (vertical line) is the 'lead time' before a mitigative action is required to prevent a downtime incident.

There are several important lessons learned while developing the anomaly detection process, described below. These were broadly related to the availability of data coming from sensors; design of algorithms for anomaly detection, and consumption of the output from anomaly detection process.

(i) *Sensor Variation:* the units of measurement and location of installation of sensors on multiple equipment typically vary across the deployed fleet. Appropriate corrections accounting for these must be deployed
(ii) *Acceptable Type I and II errors:* this bears directly on the assumed risk due to either missed anomaly alerts (false negatives) or effort required to characterize and respond to all alerts generated (false alarms). The methods must be optimized based on the acceptable levels of these risks.
(iii) *Parameters to use:* in a typical deployed marine asset, there could several thousand measured parameters. Deciding which parameters to include for the

anomaly detection processing, for a given set of equipment, poses a data dimensionality challenge. This can be addressed using equipment design and operations knowledge.

(iv) *Edge vs Central deployment:* deploying anomaly detection algorithms at a central location provides an advantage to gain insights from across the fleet. However, deploying at edge can provide initial alerts (thresholds based) to on-board personnel

(v) *Anomaly Consumption Process:* a deliberate process to consume the outputs of the anomaly detection algorithms must be developed. These processes include: characterizing actual alerts, versus sensor issues; the feedback cycle from on-board personnel; and the operating procedures to respond to specific alerts for effective anomaly detection.

Section 5: Summary and future

We discussed the role of data-driven methods to improve and scale existing maintenance practices for marine assets. We discuss the relevance and evolution of operators, class societies and fleet managers to a more condition-based paradigm, which enables optimal decision making which serves to preserves life and property, while also reducing costs of operation and uncertainty associated with unplanned downtime. This directly supports the preservation of availability. We discussed the relation and factors guising the transition from traditional maintenance practices to condition-based and their relation to the condition-based class model. Finally, this paper describes one of the major elements for the above transition – anomaly detection and it role. We described the main elements and lessons learnt in implementing an anomaly detection process. This requires design of both the data-driven methods and subsequent alarm consumption process design.

For future work, we continue to explore the relation between data-driven methods and 'soft' factors such as the human element, and their impact on overall success of the condition-based process. In addition, the authors work continues to expand on the inter-connectivity of data-driven advanced methods, data acquisition and connectivity with business operations, to achieve higher "on the ground" availability. This will be reported in our future work.

References

[1]Wikipedia page for condition based maintenance:https://en.wikipedia.org/wiki/Condition-based_maintenance

[2] Chandola, V, et. al Anomaly Detection: A Survey. ACM Computing Surveys, Vol. 41(3), Article 15, July 2009.

[3] Mitchell, T Machine Learning

[4] Nanda, S and Krueger, M: Enhancing equipment monitoring capabilities using anomaly detection. Presentation at Machine Learning, Deep Learning and AI in Oil & Gas Conference, 2016 Houston.

[5] Casella, George and Berger, Roger L. Statistical Inference

[6] Markou, M and Singh, S Novelty Detection: a review – part 1: statistical approaches. ACM Digital Library

[7] Markou, M and Singh, S Novelty Detection: a review – part 2: neural network based approaches. ACM Digital Library

[8] Si, Xiao-Sheng, et al. Remaining useful life estimation – A review on the statistical data

driven approaches, European Journal of Operations Research, 2011
[9] Wikipedia page on Bayesian Inference: https://en.wikipedia.org/wiki/Bayesian_inference
[10] Banks, Jeff Overview of Condition Based Maintenance, The Pennsylvania State University
[11] Moubray, John Reliability Centered Maintenance-II
[12] ABS, Guide for Surveys Based on Machinery Reliability and Maintenance Techniques, 2016
[13] Hawkins, D Identification of Outliers, Chapman and Hall
[14] Hans-Peter Kriegel, et al Outlier Detection Techniques, 2010 SIAM International Conference on Data Mining

Subrat Nanda is the Chief Data Scientist at the American Bureau of Shipping in Houston and leads data science and advanced analytics efforts. Subrat has over 15 years of experience in applying data science to industrial problems in the areas of asset condition-based maintenance (CBM), prognostics & health management (PHM), reliability engineering, statistical risk assessment and services enhancement. He has successfully developed and deployed applications using machine learning, artificial intelligence & data science technologies in diverse domains such as marine & offshore, gas turbines, wind turbines, financial risk, inspection and marketing. Subrat believes that fusing domain knowledge, data and data science are the key elements for successful industrial analytics. Subrat holds a Master's degree from the University of Exeter, England in Autonomous Systems and a Bachelor's degree in Engineering from University of Nagpur, India. He has filed 14 US patents, 6 trade secrets and has authored more than 25 publications and/or conference presentations in the general area of applied data science for asset health management.