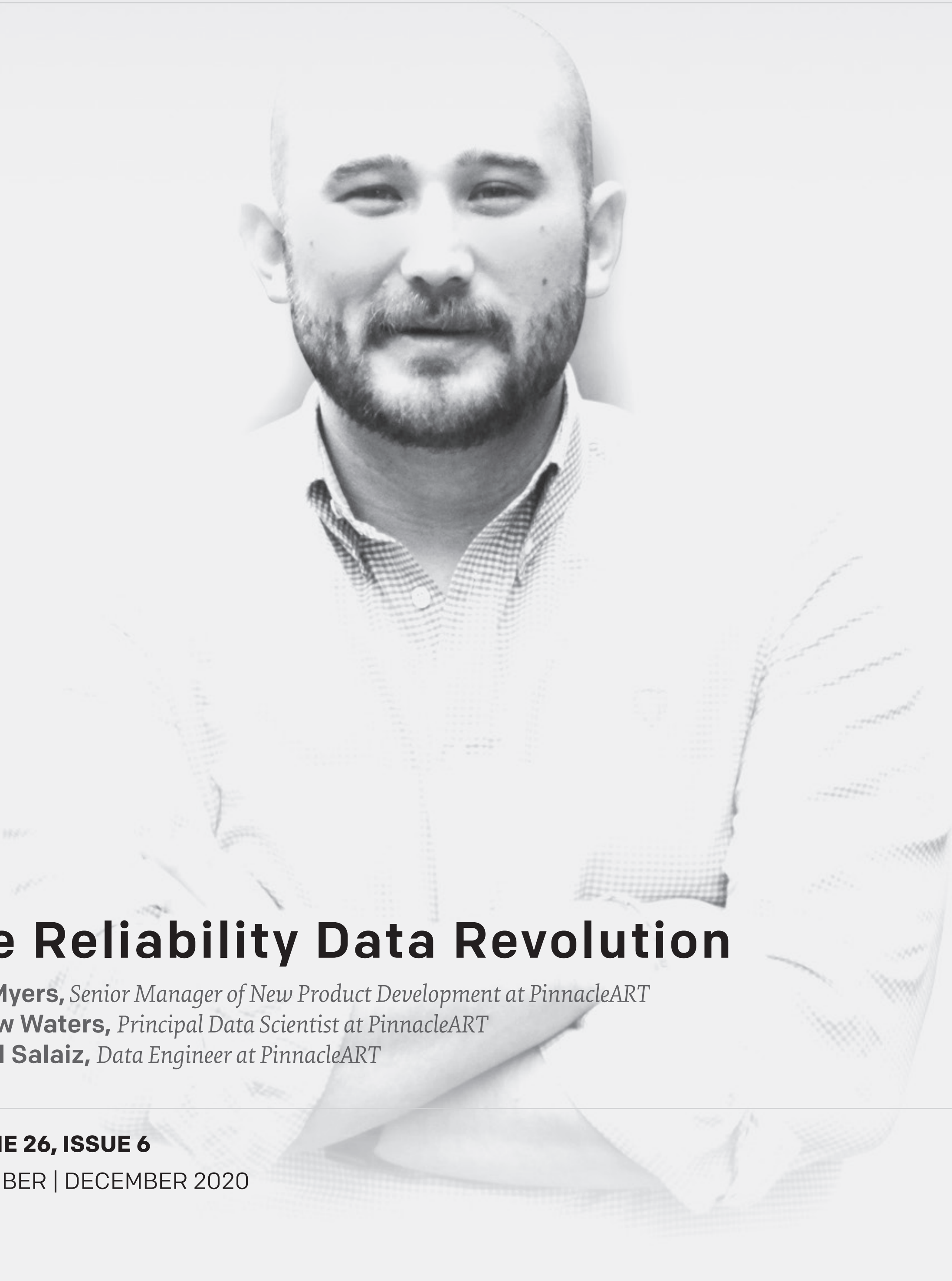
Inspectioneering Journal

ASSET INTEGRITY INTELLIGENCE



The Reliability Data Revolution

Ryan Myers, *Senior Manager of New Product Development at PinnacleART*
Andrew Waters, *Principal Data Scientist at PinnacleART*
Rachel Salaiz, *Data Engineer at PinnacleART*

VOLUME 26, ISSUE 6

NOVEMBER | DECEMBER 2020

The Reliability Data Revolution

Ryan Myers, *Senior Manager of New Product Development at PinnacleART*
Andrew Waters, *Principal Data Scientist at PinnacleART*
Rachel Salaiz, *Data Engineer at PinnacleART*

Clive Humby famously said: "Data is the new oil." As a society, we spend inordinate amounts of time and resources creating, collecting, and organizing data. When harnessed effectively, data can transform entire industries as evidenced by companies such as Amazon, Netflix, Spotify, and many others. These companies have integrated data analytics into their products and entire business models, which has enabled them to provide unparalleled levels of value to their customers. In the reliability industry, we also significantly invest in the creation and collection of data. But, are we really using our data to drive smarter reliability decisions?

We are in the midst of the largest digital transformation era the world has ever seen.

We are in the midst of the largest digital transformation era the world has ever seen. Recent innovations have led many companies to invest in a variety of initiatives such as machine learning and digital twin technologies. While these tools can provide a multitude of benefits, they can easily become expensive investments that fail to provide the desired business value. After all, when you look at reliability performance over the past several decades, the pace of improvement has stagnated despite these new innovations and significant capital investments.

Why is this the case? We typically see two schools of thought when it comes to reliability management and performance improvement. One approach focuses on the application of first principles accompanied by subject matter expert (SME) based analysis and decision-making. The other focuses on leveraging large volumes of data (dubbed "big data") and sophisticated algorithms that conduct much of the analysis and decision-making processes. In other words, segments of the reliability industry are pitted against each other in a battle of humans versus machines, or more specifically, a battle of intuition versus data. However, this is flawed thinking on both sides—while machines have time and time again throughout history proven their ability to outperform humans at certain tasks, we need to recognize that people are still better than machines in many areas, especially where we lack data or the right algorithms to process the data effectively.

Facilities that can successfully leverage both their data and expertise while effectively integrating this unified model into their core business processes can improve performance, eliminate non-value-added activities, and ultimately, gain a competitive edge over their peers. The next step-change in reliability performance will come from humans working effectively with machines, rather than at odds with one another.

The Evolution of Data in Reliability

The buzzword of the decade seems to be "big data." One major application of big data is machine learning, which is simply a set of advanced mathematical techniques that can sort through massive datasets to identify correlations and patterns that are difficult or otherwise impossible for a human to detect. The result of this pattern recognition is the ability to make surprisingly accurate predictions when, for example, classifying data, identifying anomalous behaviors, determining likely outcomes, or providing recommendations. Machine learning algorithms can automatically learn, hence the name, and continually improve their predictive capabilities without being told how to do so.

Common applications of machine learning technologies in the reliability industry include multivariate anomaly detection, early failure prediction, and failure mode classification, to name a few. In most of these applications, abnormalities and data signatures are identified and correlated to make asset failure predictions much earlier in the P-F curve than conventional means, enabling owner/operators to take proactive action to prevent an incident from occurring or provide ample time to prepare. While these capabilities are powerful, they do have significant requirements such as large amounts of good quality and organized data, significant monitoring and data management infrastructures, and the need to observe failures before they can predict them. As a result, these capabilities can limit potential applications and require a significant financial investment from facilities.

Despite this increase in available data and advanced analytics, facilities still use many segmented and qualitative approaches for managing asset reliability. This results in introducing subjectivity and variability in how reliability programs are managed because the decisions are not informed by data and therefore, are not consistently applied. Often, these decisions are also overly conservative and the data that is leveraged has safety factors and margins of error built in which further drive the divergence of our perceived models further from reality. This results in overspending and increased risk across our facilities due to allocation of resources on the wrong prioritized tasks.

Blending both intuition and data analytics enables owner/operators to improve performance while simultaneously reducing spend.

While there are strengths and weaknesses with both approaches, blending both intuition and data analytics enables owner/operators to improve performance while simultaneously reducing

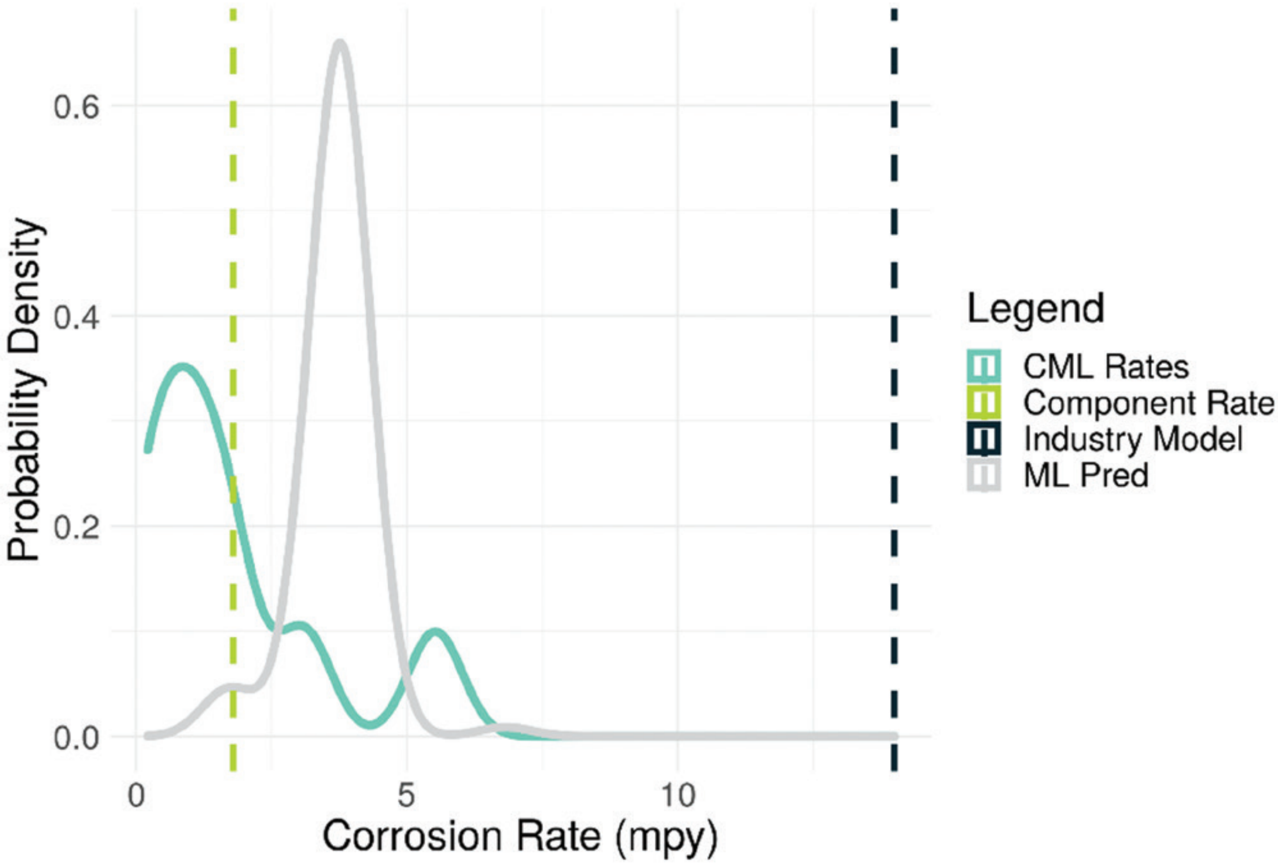


Figure 1. Degradation rate distributions of actual degradation, industry standard approach, and ML model.

spend by understanding performance drivers, quantifying and understanding how uncertainty in data drives risk, and dynamically focusing limited resources where they will add the most value.

Modeling Degradation Using Machine Learning

In an effort to leverage data to better predict asset health, identify reliability threats earlier, and more effectively allocate limited facility resources, Pinnacle's Research and Development team recently completed a study in which a data-driven approach was applied to modeling degradation compared to our own current industry practices. Improving both the accuracy and precision of asset degradation modeling can have a profound impact on the results and recommendations output from a Risk-Based Inspection (RBI) or other quantitative approaches to reliability management and performance optimization.

In this study, Pinnacle utilized a data-driven model that leveraged machine learning (ML) techniques to predict degradation rates, and associated variability, for a select set of assets from a dataset of catalytic reformer units. The model was fed large amounts of pertinent data including asset attributes, operating conditions, process stream data, inspection history, and other commonly available data to predict degradation. The results of the study showed that Pinnacle's data-driven model was able to predict degradation rates and show their variability with a higher level of accuracy compared to existing industry standard practices and subject matter expert estimation.

Figure 1 is an example output from the data-driven model, illustrating the predicted degradation rates for a drum in the reformer dataset. The actual observed degradation rates from inspection history are illustrated by the teal line. The machine learning, data-driven predicted rates are represented by the gray line. The data-driven distribution indicates that rates around 4 mils per year are likely, whereas rates greater than 7 mils per year and less than 1 mil per year are unlikely. The vertical green dashed line represents the average degradation rate (around 2 mils per year) and the black vertical line shows the modeled degradation rate (18 mils per year), which was calculated leveraging *API RP 581, 3rd Edition* and a materials & corrosion SME's judgement.[1]

Table 1. Mean absolute error for both the industry standard approach and ML model compared to actual degradation.

Metric	Industry Standard	ML Model
Mean Absolute Error	5 mils/year	3.1 mils/year

Results show the machine learning based model was much closer to the measured reality of the component than the industry standard approach (API RP 581), while also quantifying the uncertainty associated with the variability of degradation rates. This quantification of uncertainty is key to better understanding actual risk, thus improving the ability to predict potential failures and, more importantly, when to take action to effectively mitigate the risk.

Although **Figure 1** illustrates only a single example, this analysis was conducted on a population containing over 10,000 assets. **Table 1** contains the results comparing the mean absolute error of the industry standard approach and ML model as compared to actual degradation for this larger population.

Not only was the ML model able to predict degradation with approximately 38% less error, but the model can be run in near-real time on the entire population of assets. This enables timely evaluations of changes in degradation as a result of continually changing process conditions such as during an integrity operating window (IOW) excursion or if an owner/operator needed to quickly simulate the potential impacts of a change in feedstock as in each of these cases the prior measured corrosion rates are not necessarily a good indicator of future degradation.

While current industry practice and subject matter expertise is not always as overconservative as depicted in **Figure 1**, this study shows that there is an opportunity for the industry to leverage data analytics to identify potential threats sooner, reduce uncertainty to improve predictability, and focus limited resources to drive the largest performance impact.

Additionally, experts can also be focused where they can make the largest impact, leveraging their strengths to compensate for the potential gaps in data analytics including but not limited to: validation and continuous improvement of new algorithms, calibration of the data uncertainty used in these algorithms, providing experience-based assumptions in areas where we lack data, identifying the root causes of anomalies or issues, and developing solutions to improve the predictions of advanced data analytics.

The Future of Data in Reliability

The future of the reliability industry will depend heavily on how facilities utilize their expertise coupled with data analytics to make smarter reliability decisions. As an industry, we are creating more data every day, leveraging advanced technologies such as machine learning to continue to evolve our current capabilities and potential use applications. It is up to us as reliability professionals to challenge the status quo and work to integrate this unified model throughout our programs and into our decision-making processes so that we can realize the benefits and realize a step-change in performance.

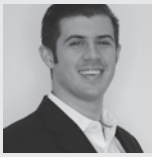
How are you using data to improve the reliability of your facility? ■

For more information on this subject or the author, please email us at inquiries@inspectioneering.com.

REFERENCES

1. API RP 581 3rd Edition, Part 2 - Probability of Failure Methodology, Annex 2.B - Determination of Corrosion Rates

CONTRIBUTING AUTHORS



Ryan Myers

Ryan Myers is the senior manager of New Product Development at Pinnacle, where he leads a cross-functional teams of engineers, data scientists, and software developers. His team focuses on creating new products and services that transform customers' reliability, integrity, and maintenance programs, enabling them to reach new levels of performance. Ryan specializes in product management, digital transformation, operational excellence, project management, mechanical integrity, and reliability engineering. He also has extensive experience leveraging predictive analytics to drive smarter decisions and improve business performance. Ryan obtained his Bachelor of Science in Mechanical Engineering with a Minor in Business from The University of Texas.

Rachel Salaiz

Rachel Salaiz is a data engineer at Pinnacle, where she leads efforts in data collection, standardization, and storage. Rachel specializes in quality control, automation, data cleansing, and data analysis. She also has experience implementing machine learning software solutions to help customers improve their reliability programs. She obtained her Master of Science in Data Analytics from the University of Houston - Downtown.

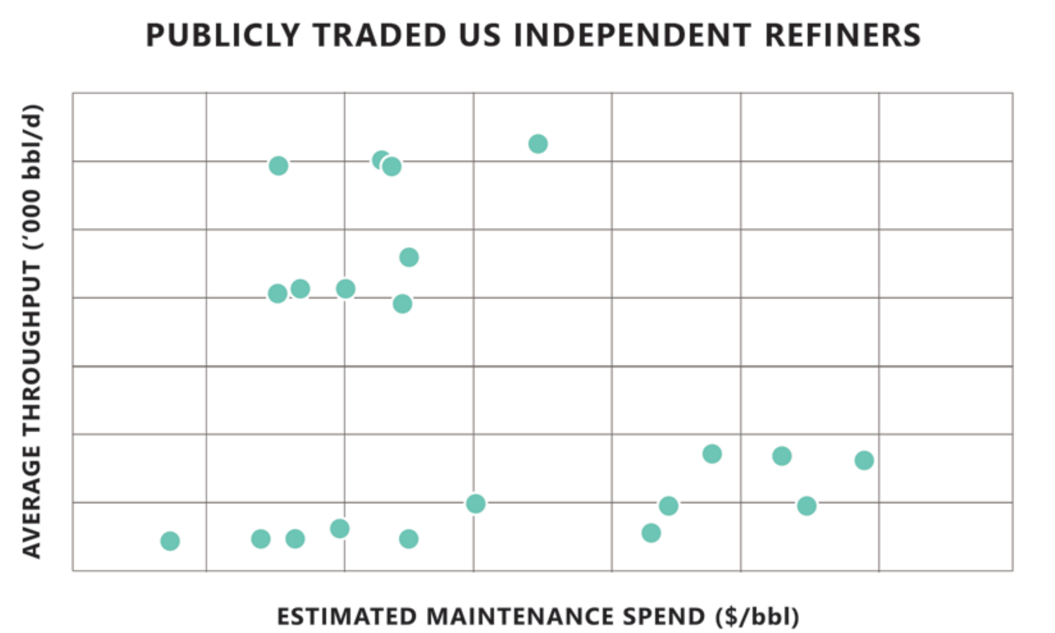
Dr. Andrew Waters

Dr. Andrew Waters is a principal data scientist at Pinnacle, focusing on developing data-driven algorithms to enhance a variety of reliability and maintenance applications. Dr. Waters also specializes in utilizing machine learning methods to improve and augment human decision making. He has utilized these skills across a diverse set of industries including finance, communication systems, engineering, signal processing, optimizing student learning outcomes, and hiring and recruitment programs. Dr. Waters holds a doctorate in Electrical and Computer Engineering from Rice University and is the author of over 20 publications in the areas of signal processing, machine learning, and Bayesian statistical methods. His research interests include sparse signal recovery, natural language processing, convex optimization, and non-parametric statistics.

PINNACLE

DATA-DRIVEN RELIABILITY

Reliability can mean the difference between a bankrupt business and a thriving one.



How much is reliability costing you, and what are you doing about it?

Pinnacle is exclusively focused on helping industrial facilities better leverage their data to improve reliability performance, resulting in increased production, optimized reliability and maintenance spend, and improvement in process safety and environmental impact.

+1 281 598 1330 | info@pinnaclereliability.com | pinnaclereliability.com