Inspectioneering Journal

ASSET INTEGRITY INTELLIGENCE

Case Study: Successful Use of Machine Learning to Model Degradation in Hydrocrackers

Dr. Andrew Waters, *Director of Data Science at Pinnacle*
Ryan Myers, *Product Manager at Pinnacle*

VOLUME 28, ISSUE 2
MARCH | APRIL 2022

Case Study: Successful Use of Machine Learning to Model Degradation in Hydrocrackers

Dr. Andrew Waters, *Director of Data Science at Pinnacle*
Ryan Myers, *Product Manager at Pinnacle*

Introduction

One fundamental challenge many facilities encounter is accurately estimating the rate of degradation throughout their facility. Degradation rates are used to inform facility technical leadership about the present risk of individual assets and are used to schedule a variety of inspection and maintenance tasks. Incorrectly estimating degradation rates can lead to an inadequate understanding of risk. If estimated rates are overconservative, facilities may waste resources on unnecessary inspections. Alternatively, if estimated rates do not correctly capture all potential risk, facilities can experience large economic or health, safety, and environmental (HSE) consequences.

Currently, degradation estimates are made by subject matter experts (SMEs) who leverage industry standard tools such as API RP 581, various industry-recognized damage/corrosion models, prior inspection data, and the wealth of knowledge and experience they have gained throughout their careers. While these methodologies have historically been a credible method of predicting degradation rates, actual rates can differ significantly from the estimated rates for a variety of reasons. First, while there may be a tremendous amount of data available to an SME, sifting through and analyzing a large quantity of data can be daunting for a human. Further, data quality issues such as incomplete, missing, or poor-quality data can also drastically alter SME perceptions. Finally, even with a complete and clean set of data, actual degradation rates can differ significantly from theoretical values due to a variety of factors such as environmental considerations and changes in facility process conditions.

Machine learning can strengthen the natural limitations of human SMEs, resulting in methods that predict degradation rates more quickly and accurately. When used properly, machine learning models can quickly sort through, organize, and clean massive amounts of input data such as temperature, pressure, metallurgy, and stream information, and leverage this data to make more accurate degradation rate predictions. Further, models based on data science can continually evolve and learn based on newly acquired data, preventing results from becoming stagnant.

Machine learning models can be leveraged—even by facilities with limited data—to strengthen areas of natural human limitations when predicting degradation rates. The following study found that a machine learning model was able to predict degradation rates for a hydrocracker unit more accurately and with a smaller margin of error compared to current industry practice. This article will discuss the details of this study as well as future applications of machine learning models for the industry.

The Hydrocracker Study

The purpose of this study was to compare the accuracy of degradation rates estimated by a machine learning model to the degradation rates calculated by human SMEs leveraging API RP 581. This study specifically focused on degradation rates for a variety of piping circuits in a hydrocracker. The study focused on piping circuits specifically because they are responsible for the majority of loss of primary containment failures in refining facilities [1].

The hydrocracker analyzed in this study had 25 piping systems which included 70 piping circuits and 1,662 condition monitoring locations (CMLs). The available data consisted of:

- Historical thickness inspection data, including the measurement dates and measured thickness readings.
- Asset level information including asset metallurgy, operating temperature, and operating pressure.
- Process information associated with each circuit in the unit, including parameters such as H2S content, pH, and Total Acid Number (TAN).

Given raw inspection data taken over time, the corrosion rate at each CML in the population was calculated. **Figure 1** shows these observed degradation rates as a box plot broken out by piping system. Note that, on average, the degradation rate on any given CML for this hydrocracker is quite low (between 1 and 2 mpy), but that there is a large amount of variation in degradation rates both in the hydrocracker unit itself as well as across each system within the hydrocracker. Observed corrosion rates for many CMLs in the hydrocracker approached 10 mpy.

The machine learning model developed for this study relies directly on observed data, rather than industry standards such as API RP 581, to predict degradation rates. By leveraging asset data, the machine learning model learns how different variables affect the overall degradation rate and will continue to improve its predictions as it is exposed to new data over time. The machine learning model does this using supervised or example-based learning, where each input example corresponds to a single CML with its various elements of asset information as well as the measured degradation rate observed for that CML. By consuming these training examples, the machine learning model builds a representation of how strongly each variable influences the final measured degradation rate. For example, the model will

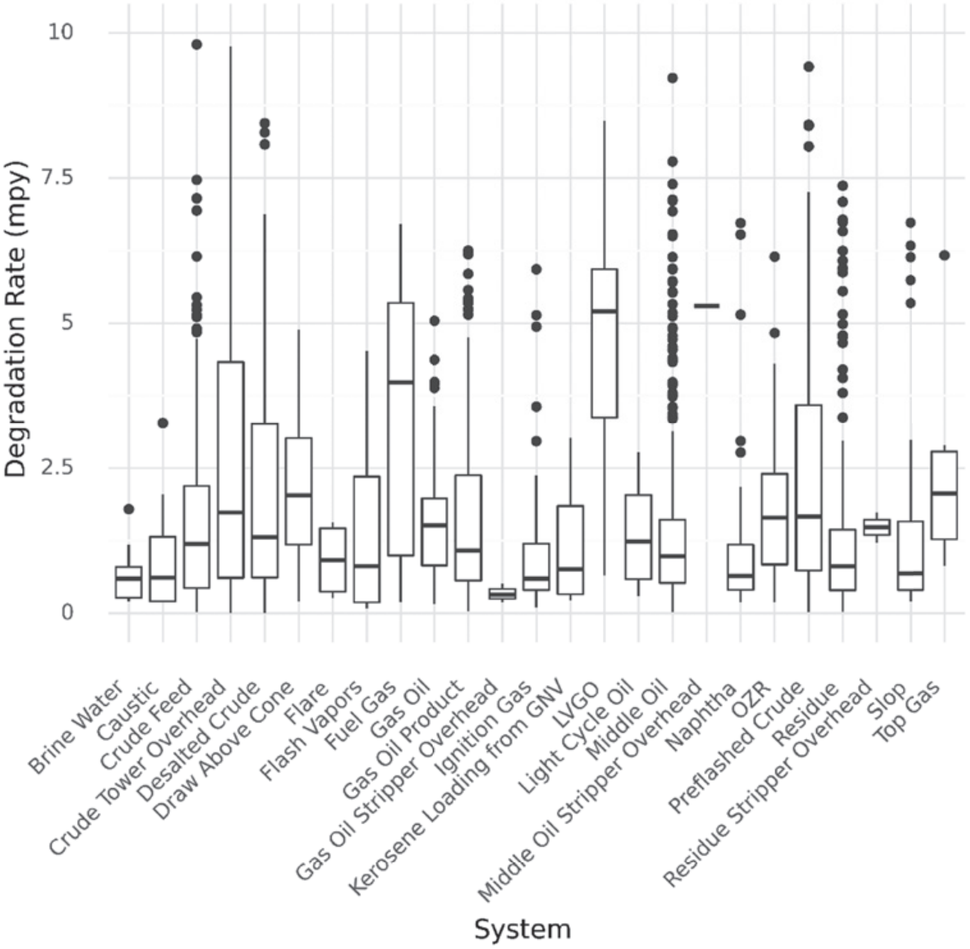


Figure 1. Box-whisker plot of the observed degradation rate ranges across the various systems in the hydrocracker. The horizontal lines in each box correspond to the median degradation rate for that system and the dots correspond to statistical outliers outside of the standard interquartile range. Note that while most data suggests low corrosion rates, there is considerable variation both across individual systems as well as the entire hydrocracker itself.

learn that, with all other variables being equal, higher temperature generally leads to higher degradation rates. Note that the machine learning method was not pre-programmed with these rules and assumptions. Rather, the model learns these relationships naturally by observing data. After training, the machine learning model can make predictions for new CMLs that it has using the rules that it acquired during its training, enabling it to make reasonable and accurate inferences for new degradation scenarios that may be different than the ones it has previously encountered.

Figure 2 highlights the degree to which the top ten most important variables contributed towards the prediction of degradation rates within the piping circuit. For this specific hydrocracker, operating pressure was the most impactful variable in predicting degradation rates, followed by temperature, O_2 levels, and pH. While still providing a meaningful contribution to the model, the hydrocracker's Total Acid Number (TAN) was a far less significant variable than many others.

The accuracy of the model was evaluated by training the model on subsets of the overall data. The resulting model was then used to make predictions on the remainder of the data, which corresponds to new cases that the model had not previously encountered. This procedure was repeated over multiple partitions of the data and then compared to the SME-estimated degradation rates that were provided with the initial dataset.

We evaluated accuracy by computing the absolute error between the predicted degradation rate (either given by the SME or by the machine learning model) and the actual corrosion rate observed in the data. This is done by computing the error (difference) between the two rates, taking the absolute value of that error, and averaging across all observations. The average absolute prediction error for the API RP 581 model was 7.2 mpy, whereas the average absolute error for the machine learning model was significantly smaller at 1.5 mpy. In 85% of the cases that were examined in the study, the API RP 581 rates were more conservative than the observed degradation rates. However, it is noteworthy that in 15% of the cases, the degradation rates predicted by API RP 581 were significantly under-conservative, meaning that API RP 581 underestimated the corrosion rates at these CMLs. The machine learning model was much more accurate at predicting corrosion rates for these CMLs. In one case, API

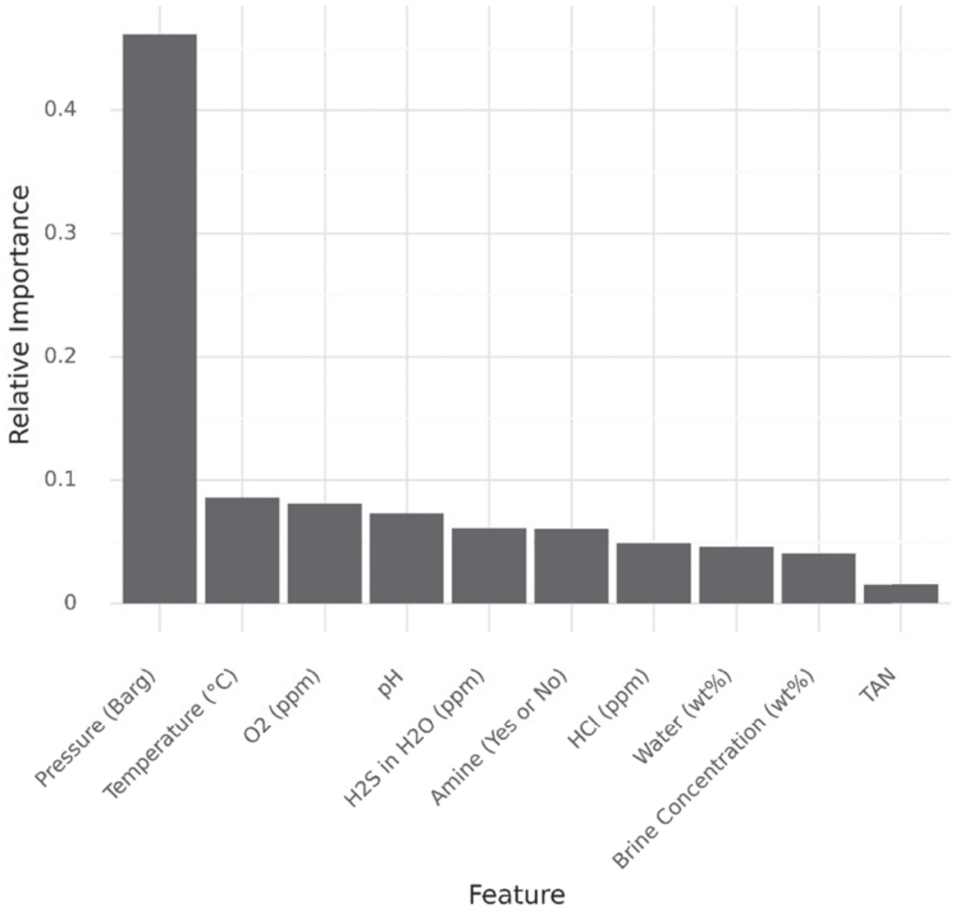


Figure 2. The relative importance of the top ten variables used by the machine learning model. Pressure was the most informative variable for the model.

581 underestimated the true rate of corrosion by nearly 10 mpy, whereas the machine learning model only underestimated by 2 mpy. While the machine learning model still underestimated the degradation rate for this specific CML, it did so by a significantly smaller margin and provided a better estimate of degradation.

Key Takeaways

While this study provides examples of how data science and "big data" can be applied to the industry, there are a few important points to highlight.

First, API RP 581 corrosion estimates tend to be conservative (by intent), which can make it appear easy for a machine learning model to outperform API RP 581 rates in terms of overall accuracy. However, a previous study that compared degradation rates in a reformer unit also found that machine learning-based methods outperformed an actual human SME tasked with providing accurate degradation rates [2]. These findings highlight how machine learning models can predict degradation rates more accurately than current standards used in the industry.

Second, while the machine learning methods explored in the study were shown to be more accurate than the SME-assigned rate, these models should not be considered as a replacement for the SME. The expertise that a qualified SME brings to the process of estimating degradation rates is irreplaceable and machine learning serves as a potential aid to the SME in their work. For example, the machine learning model can help SMEs quickly identify areas of particular concern within a facility or calculate changes in degradation rate as process conditions in the facility change.

Finally, each facility is unique, and the machine learning model will produce results that are specific to the facility that was used to train it. However, the general rules learned by the model regarding how individual variables affect degradation rates will transfer to any facility. Additionally, the machine learning model is capable of quickly calibrating to a new facility with only a small amount of data and naturally improves over time as it observes more data to refine its rules and predictive capabilities.

Future Applications of Machine Learning

The results of this analysis show how data science can be used to solve reliability problems faster and more accurately than current industry practices. The focus of this article was on piping circuits for a hydrocracker, but the methods employed can be easily applied to a variety of units such as reformers and crude units. These models can also be applied to assets other than piping such as pressure vessels, tanks, or exchangers.

Additionally, facilities can currently use data science and machine learning techniques to aid in a variety of tasks beyond degradation rate estimation. For example, data cleansing is a vitally important, yet time-consuming process for humans. Machine learning methods, on the other hand, can quickly identify potentially anomalous inspection readings and suggest reasonable values in the case of missing data. In addition to data cleansing, machine learning techniques can further be used to automatically assign damage mechanisms, grade inspection reports, and flag assets that may need retro-positive material identification (PMI).

Data science and machine learning have the potential to revolutionize degradation rate estimation throughout a facility. By implementing these machine learning methods, facility leaders will have a better understanding of the risk profile of their facility, enabling them to make more strategic decisions. ■

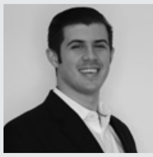
For more information on this subject or the author, please email us at inquiries@inspectioneering.com.

REFERENCES

1. Gysbers, A.C., 2012, "A Discussion on the Piping Thickness Management Process," Inspectioneering Journal, 18(5), pp. 10-13.

2. Pinnacle, 2022, "Data-Driven Reliability - Using Big Data to Model Degradation in Reformer Units," https://pinnaclereliability.com/learn/white-papers/data-driven-reliability-using-big-data-to-model-degradation-in-reformer-units/.

CONTRIBUTING AUTHORS



Ryan Myers

Ryan Myers, Product Manager at Pinnacle, oversees all new product development activities for quantitative reliability methods and the application of advanced analytics technologies. He leads multi-disciplinary technical teams across engineering, data science, and software development fields to drive the creation of new products and services focused on increasing customer value through transforming their reliability, integrity, and maintenance programs. Ryan specializes in mechanical integrity and reliability engineering, operational excellence, probabilistic modeling, decision analytics, digital transformation, and product management. Ryan obtained his Bachelor of Science in Mechanical Engineering with a minor in business from The University of Texas and is also a certified Lean Six Sigma Black Belt.

Dr. Andrew Waters

Dr. Andrew Waters is Chief 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.