Data-Driven Reliability: Using Big Data to Model Degradation in Reformer Units

One Pinnacle Way,
Pasadena, TX 77504

281.598.1330

pinnaclereliability.com

Table of Contents

Purpose

The objective of this study is to compare the accuracy of Pinnacle's automated, data-driven model with the results of current industry practices regarding the prediction of degradation rates for assets and components.

Executive Summary

- Pinnacle has developed an automated, data-driven model that uses machine learning (ML) techniques to predict degradation rates for assets and components. When calculating degradation rates, the ML model considers information about the asset including its operating temperature, process stream data, system type, and other asset attribute data.
- Rather than employing rule-based models, such as API 581, the ML model learns how to predict degradation rates by being exposed to a large amount of data. It uses this data to naturally learn how different variables influence the overall degradation rate and will continue to improve over time as it is exposed to additional, higher-quality data.
- Pinnacle conducted an analysis that compared the accuracy of the degradation rates predicted by Pinnacle's ML model to the degradation rates predicted by a human subject matter expert using the current industry standards. Pinnacle's ML model performed significantly better than standard industry practice in overall accuracy.
- The results of the analysis exemplify the exciting possibilities for how "Big Data" can be used to solve real-world reliability challenges faster and more accurately than current industry practices.

Introduction

Complex processing facilities often struggle to make confident reliability decisions because of an overload of available data and ineffective application of data analysis. Facilities often encounter challenges gathering the right data at the right time from the correct sources and are not able to easily upload their data into a multitude of systems. As a result, facilities have trouble efficiently analyzing this data to make critical decisions on how to manage their assets and invest their capital to drive overall system performance.

Pinnacle's vision is to "Make the World Reliable, One Customer at a Time." Pinnacle strives to achieve this vision though data-driven reliability, leveraging data to help customers maximize availability, mitigate risk, and optimize cost. Pinnacle's unique perspective on reliability focuses on complex systems as a whole, allowing customers to better leverage their data and make more strategic reliability decisions.

To support Pinnacle's vision, the Pinnacle Research and Development team recently completed an analysis utilizing Pinnacle's reformer dataset and industry-leading ML degradation model. In summary, the analysis illustrated that Pinnacle's ML model was able to predict degradation rates and associated variability with a higher level of accuracy and half the margin of error compared to existing industry standard practice and subject-matter expert (SME) estimation.

Data Availability

To properly assess and predict degradation rates across a variety of reformers, the following data was collected from 37 reformer units:

- Inspection history: thickness measurements in thousandths of inches (mils) on each asset, component, and Condition Monitoring Location (CML). Inspection history comprised of measurements from a variety of inspection technologies including those appropriate for detecting and measuring both general and localized degradation.

- Asset attributes: data specific to the specific asset (e.g., operating temperature, operating pressure, metallurgy, etc.).
- Process stream data: chemical composition of the process stream associated with the asset or component (e.g., H_2S concentration, NH_3 concentration, etc.).
- Assigned damage mechanisms: information such as which damage mechanisms are potentially active for each asset or component.

Of the 37 reformer units analyzed, 13 were complete enough for full model creation and analysis of general degradation rates based on thickness data. An additional 17 units were used for degradation rate benchmarking analytics.

Preliminary Analytics

Before diving into the results of the ML model, it's useful to show some high-level analytics regarding Pinnacle's dataset on reformer degradation. First, the analytics validate Pinnacle's approach by showing that the dataset is sensible and the general trends that are observed match expectation. Second, the analytics illustrate the diversity observed in the collected data, which was ultimately leveraged by the ML model in order to make more accurate degradation predictions.

Degradation Rate by Operator/Site

Degradation rates naturally vary between different operators and their individual sites. In this study, the average degradation rate for the reformer dataset is calculated for each CML in the dataset. The variance of degradation rates across different companies and sites is illustrated below, color coded by operator.

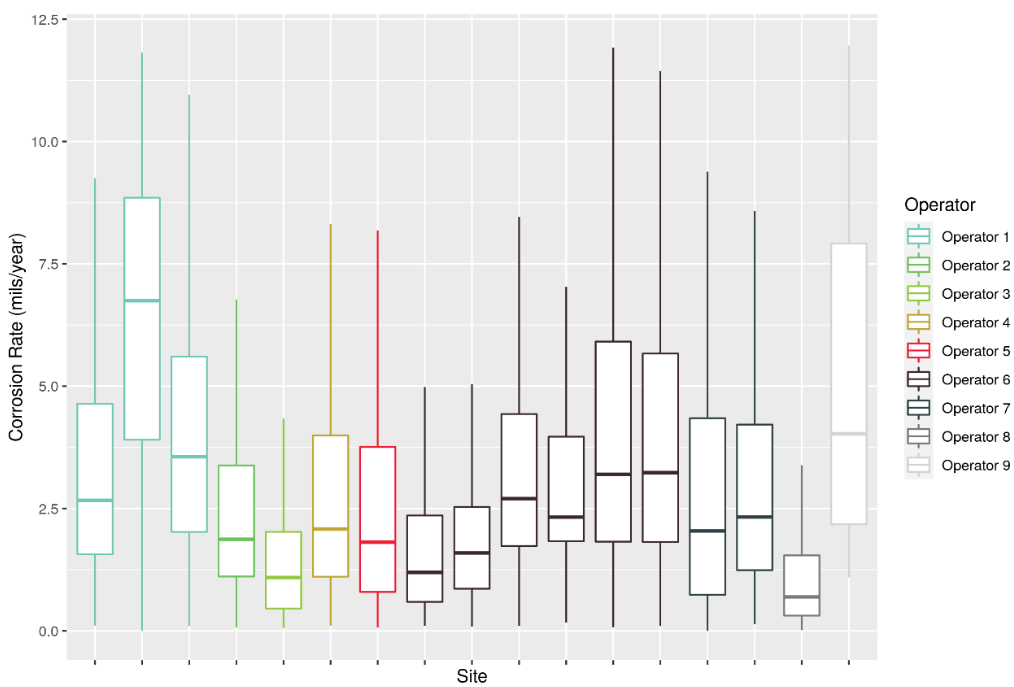

Figure 1: Boxplot of degradation rate by operator and site

Additional insights regarding the variability in degradation rate can be gained by casting the analysis as a density plot:

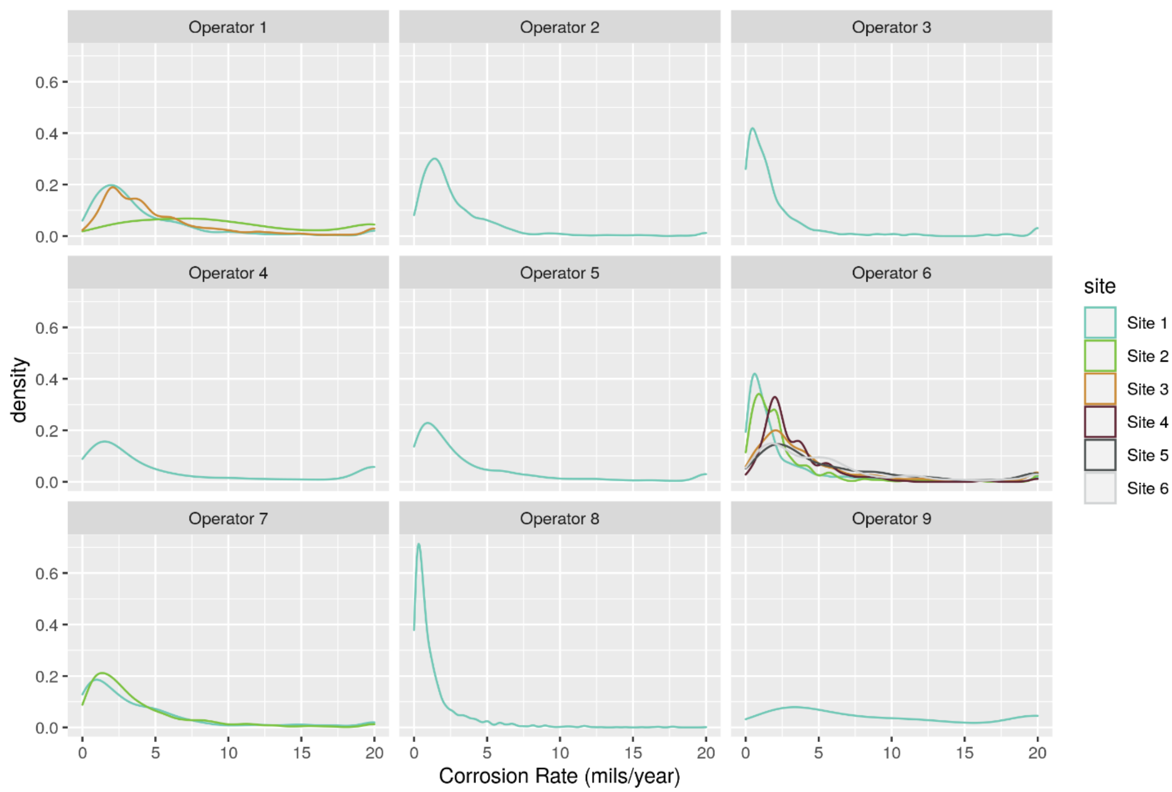

Figure 2: Density plot of degradation rate by operator and site

Most sites experience degradation rates less than 5 mils/year but also tend to have heavy tails that extend upwards to around 20 mils/year. Note that Operator 1, overall, has much higher degradation rates than the other operators, such as Operator 3 or Operator 8. Additionally, different sites may have heavier distribution tails indicating that there is higher variability in degradation. In this specific case, there appears to be areas of significantly higher degradation that is not widespread throughout the system. Also, Operator 1 Site 2 has a characteristically different shape – peaking at around 7.5 mils/year and being relatively flat indicating high variability in degradation rates throughout the system.

Degradation Rate by Reformer Type and System Function

Degradation rates can also be examined as a function of reformer type (e.g., fixed bed) as well as the system type (e.g., reaction). Shown below are the nine generic systems seen in both continuous catalytic reformers (CCR) and fixed bed reformers for reference. Higher degradation rates are expected in systems such as the combined feed or reaction system due to the expected conditions of those areas. Below is a boxplot illustrating the average degradation rates by system type in each type of reformer.

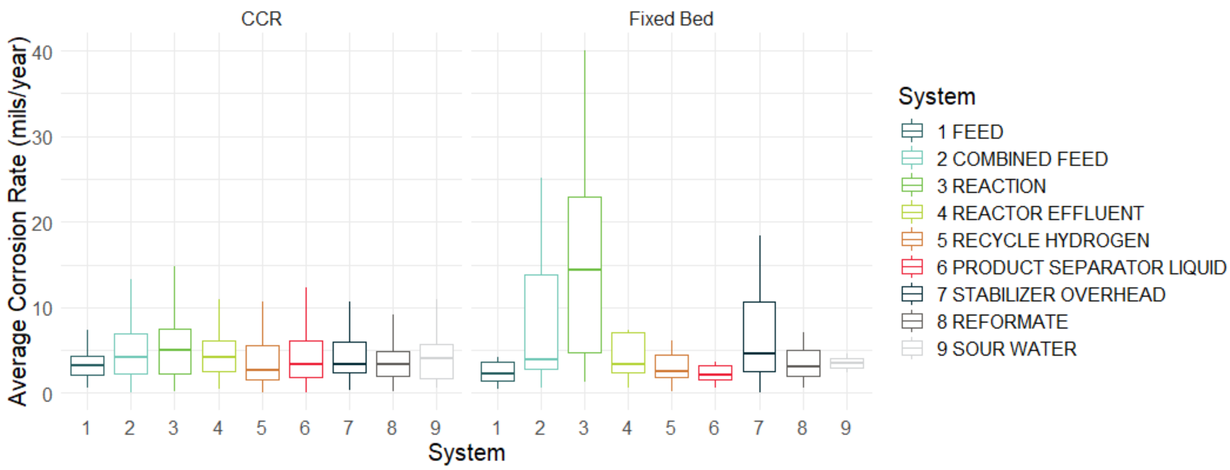

Figure 3: Boxplot of degradation rate by system type

Another view of each system's average degradation rate and average max degradation rates are illustrated below overlaid on Process Flow Diagrams (PFDs) for each reformer type:

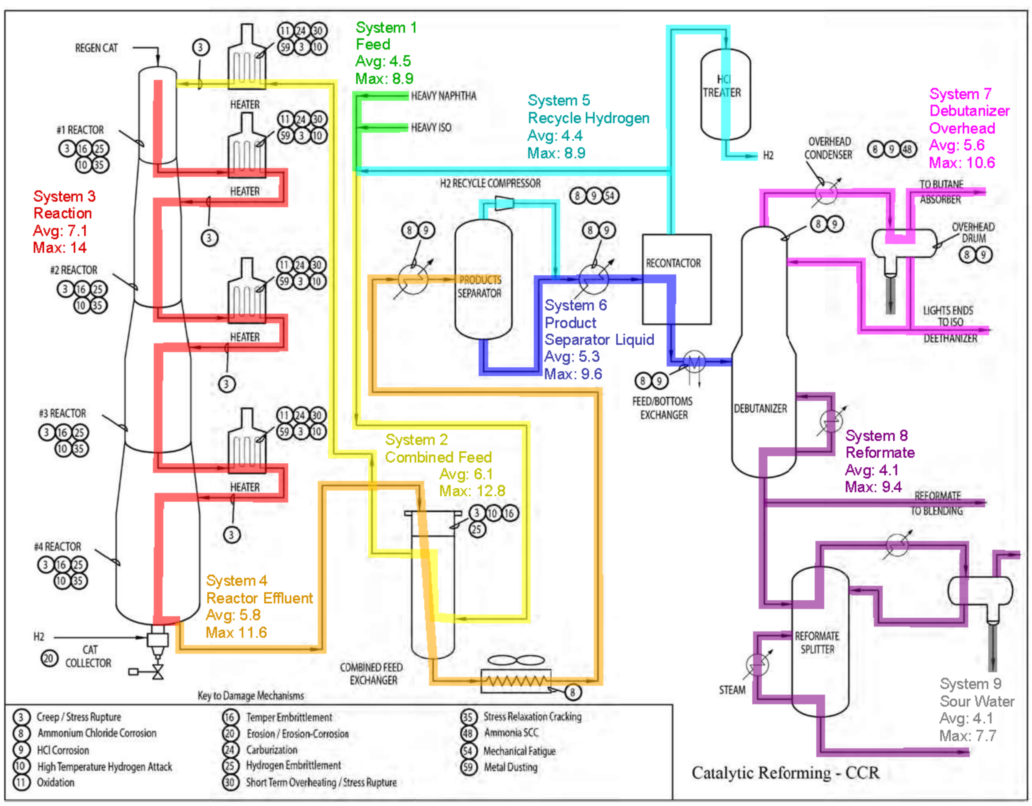

Figure 4: API 571 CCR process unit flow diagrams (From API 571 Damage Mechanisms Affecting Fixed Equipment in the Refining Industry 2nd ed. 2011, Section 5.2, Figure 5-69) overlaid with calculated average degradation rates per system type

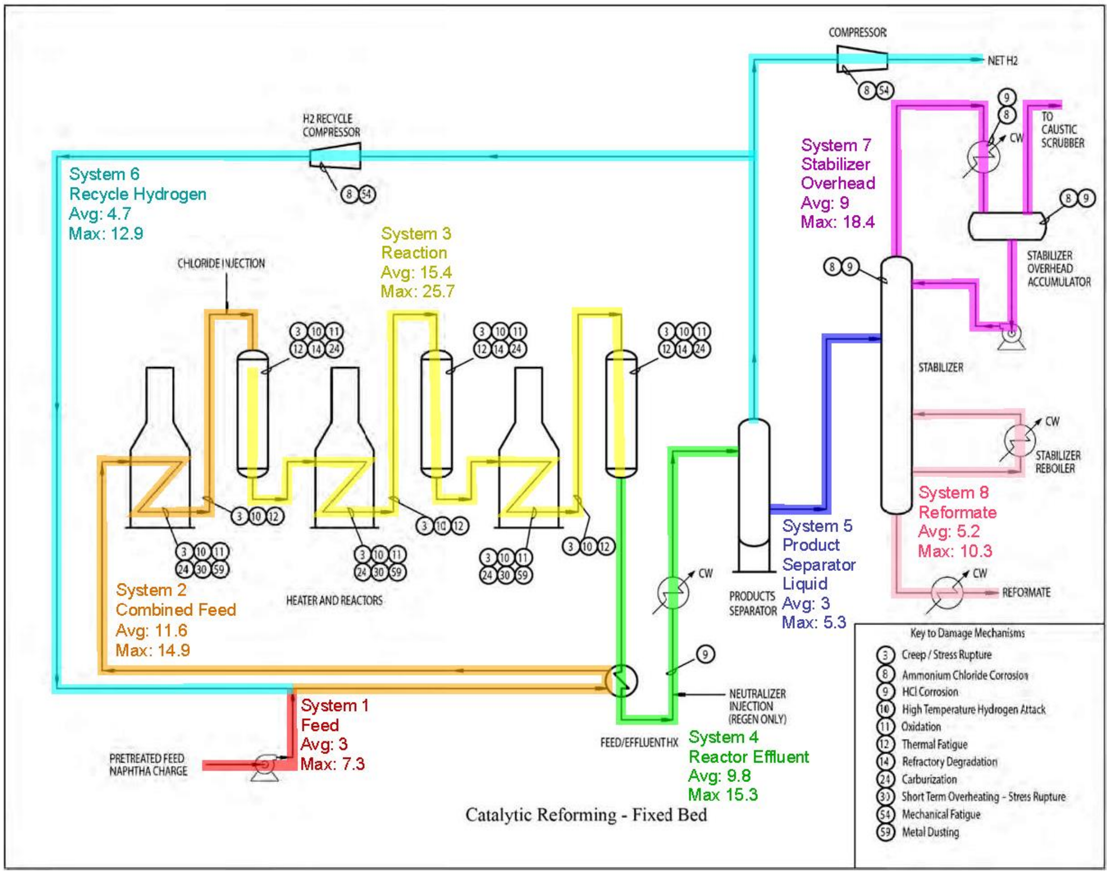

Figure 5: API 571 Catalytic Reforming – Fixed Bed process unit flow diagrams (From API 571 Damage Mechanisms Affecting Fixed Equipment in the Refining Industry 2nd ed. 2011, Section 5.2, Figure 5-70) overlaid with calculated average degradation rates per system type

The systems shown in each PFD include an average degradation rate and an average maximum degradation rate.

The Data-Driven Model

The ultimate goal of this study is to utilize Pinnacle's ML model to accurately predict degradation rates for reformer units and associated assets, and to compare Pinnacle's ML model to the current industry practice. While the exact implementation details of this model are beyond the scope of this writeup, it is important to differentiate the work of Pinnacle's ML model from the current industry practice. For example, consider degradation due to hydrochloric acid (HCl). This type of degradation can be modeled using the methodology found in API 581 or expertise from a materials and corrosion engineer. For HCl corrosion, API 581 specifies the expected degradation rate of an asset as a function of its metallurgy, temperature, pH, etc. Materials and corrosion engineers would typically use this information, in additional to a wealth of experience, to estimate degradation rates for a given asset. In this sense, the method prescribed by API 581 is a set of "rules" that describe how degradation is expected to proceed given certain variables.

While such information is incredibly valuable, it can also be somewhat misleading when carrying out analysis due to the potential for multiple active damage mechanisms and their molecular interactions for example. Additionally, this information can be limiting regarding what information may be useful in predicting degradation. The goal of Pinnacle's ML model is to learn how to deal with common situations using all available and pertinent data.

Pinnacle's ML model is not explicitly coded with rules similar to API 581. Instead, the model is fed data examples that describe a given asset or component (operating conditions, process stream data, etc.) along with the measured degradation rates. The model then uses this data to learn how different variables impact degradation rates and will make inferences such as higher temperatures generally coincide with higher degradation rates without being explicitly told that this is the case. By using data-driven, inference-based learning, the model is able to make better, more informed predictions.

In addition to predicting degradation rates, the model can assess the relative importance of each variable in the dataset as illustrated in the plot below.

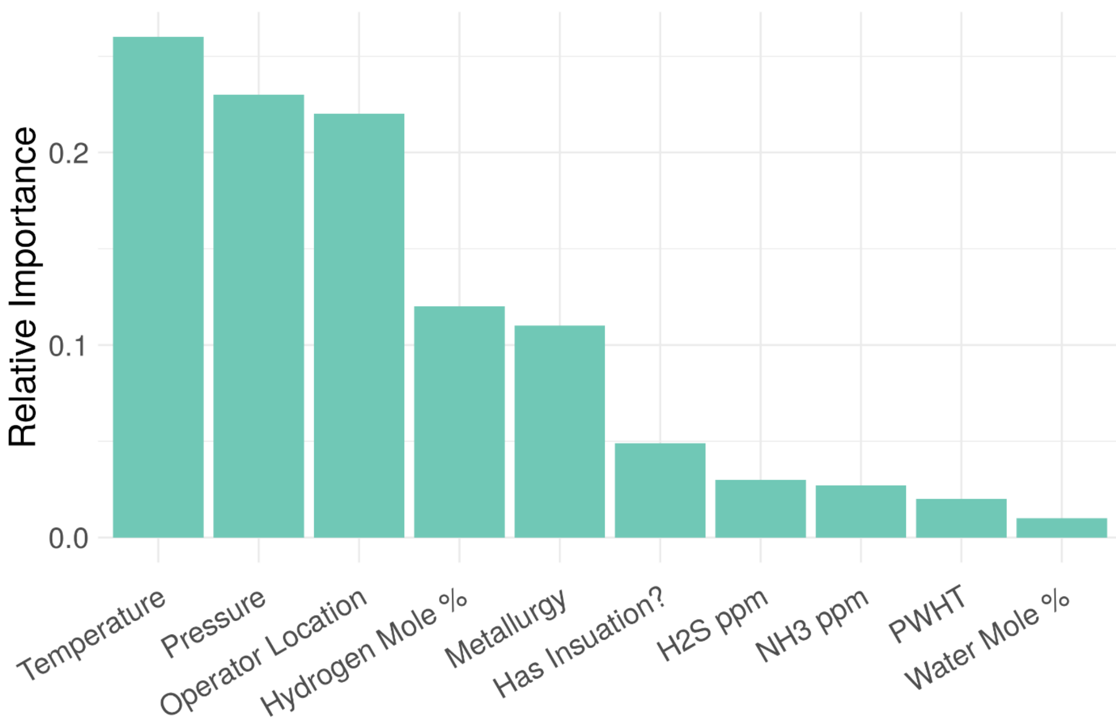

Figure 6: ML model variable relative importance

Certain variables are more important to the model relative to others. As an example, operating temperature was found to be the most relevant variable in predicting degradation rates, followed by operating pressure and location (operator site) of the reformer. Stream information, such as Hydrogen Mole % and H2S ppm are also considered highly informative. In contrast, water mole %, while providing some boost to the model's performance, is far less important to the model than some of the other variables as typically these are dry systems.

Below is example of what the model produces for a drum from one operator in the dataset. The observed degradation rates at the CML level is shown by the teal curve. The model estimates a degradation rate distribution shown by the grey curve which shows which degradation rates are likely for this component and those which are not. The degradation distribution indicates that rates around 4 mils/year are likely, whereas rates greater than 7 mils/year are highly unlikely. Vertical lines are also included to illustrate average degradation rate across the component (around 2 mils/year) as well as a degradation rate provided by the industry standard approach (18 mils/year). As seen below, the model is much closer to the measured reality of the component than the industry standard approach (581 tables), which is considerably more conservative in this case.

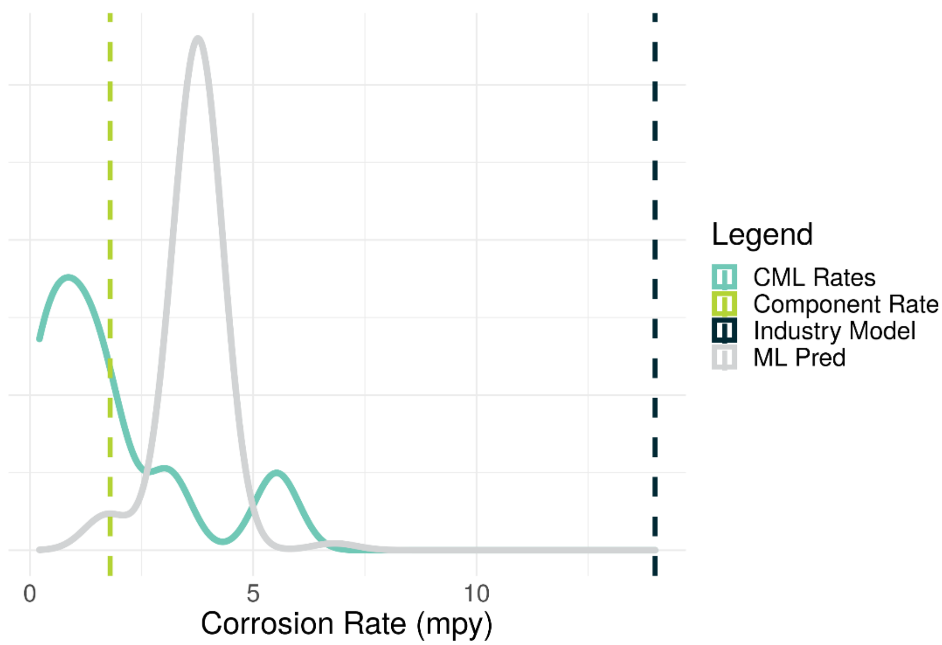

Figure 7: Example degradation rates for a drum

Comparison of the Data-Driven Model with a Human Subject Matter Expert

Pinnacle conducted an experiment to further demonstrate the power of the data-driven degradation model. In this experiment, 11 components were randomly selected from the dataset and removed from the training set of the model. The model was then trained on all remaining examples and was used to predict the degradation rates for the selected components. A materials and corrosion subject matter expert (SME) was also given the same dataset and task of predicting the degradation rates for each component using API 581 tables given available data including system diagrams, operating conditions, process stream data, etc. The data utilized by the ML model and the SME were identical to provide a fair and unbiased testing environment. The experimental results are summarized in the following table:

Component	Measured degradation rate (mils/year)	SME Estimated degradation rate (mils/year)	Model Estimated degradation rate (mils/year)
1	11.4	1	7.5
2	0.8	10	4.2
3	18.9	6	3.2
4	3.8	6	2.6
5	2.1	6	2.6
6	3.1	6	3.3
7	8.2	6	3.9
8	3.3	3	2.9
9	6.9	4	4.9
10	1.9	6	3.1
11	3.3	6	4.6

Table 1: Measured, ML model, and SME estimated degradation rates by component

The mean absolute error between the measured degradation rate and the estimates provided by the model and the industry standard are summarized in the following table:

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

Table 2: Mean absolute error for both the industry standard approach and ML model

The ML model achieved a sizable reduction in mean absolute error (38%). It is also noteworthy to point out that while the SME needed significant time to read through documentation and apply industry standards, the model could make predictions in near real-time. The model also effectively modeled not just the expected average degradation rate but also the potential range of expected degradation improving overall predictability and usability.

Constraints and Limitations

While there was a large volume of data for the model to use, significant gaps in the data limited the performance of the ML model. For example, operating temperature and pressure was present in only about 70% of the dataset and stream information (H_2S concentration, etc.) was even more scarce. Additionally, the dataset only contained single values for all variables even though these quantities typically fluctuate over time. Despite these limitations, the model still performed well and gave more accurate predictions than industry standards would produce. As both the volume and quality of the data improve, so will the ability of the model to accurately predict degradation rates.

Future Applications

This study has focused exclusively on predicting degradation rates in this document. However, the possibilities of using ML models within this industry are limitless. Some additional examples of future work include:

1. Automatically assigning applicable failure mechanisms for assets based on observed data
2. Recognizing poor choices in metallurgy (e.g., carbon steel being operated at excessively high temperatures) and *recommending* specific metallurgy replacements both to minimize cost and improve overall reliability
3. Applying this type of ML based model to predict the probability of failure for rotating assets, such as a centrifugal pump or compressor, leveraging available time series data such as vibration, temperature, discharge pressure, RPM, manufacturer, and other asset attribute data

Conclusion

Even with limited data quantity and quality, this study proves the ability of Pinnacle's ML model to more accurately predict degradation rates in reformer units as compared to industry standard practice.

Pinnacle is uniquely positioned to not only develop the most valuable reliability solutions in the world, but has the ambition, focus, and proven track record to make these innovative solutions work for Pinnacle customers, ultimately enabling them to reach new levels of performance and efficiency.

Pinnacle Asset Integrity Services, LLC dba
Pinnacle Reliability
1 Pinnacle Way
Pasadena, TX 77504
info@pinnaclereliability.com