Inspectioneering Journal

ASSET INTEGRITY INTELLIGENCE

Data-Driven Reliability—Reliability Done Right

Sean Rosier, *Principal, Asset Management Solutions at Pinnacle*

VOLUME 27, ISSUE 2

MARCH | APRIL 2021

Data-Driven Reliability—Reliability Done Right

Sean Rosier, *Principal, Asset Management Solutions at Pinnacle*

If you have ever had the pleasure of hearing Mr. Paul Barringer speak about using probability statistics to perform reliability, then you know what kind of visionary he was on the subject. The key phrase that he continually emphasized was "if you are not using statistical probability, then you are not doing reliability." He was, and still is, right. All facilities produce data. Many of the programs produce their own data beyond what is needed to operate the plant. However, despite the amount of available data that exists, numerous facilities continue to struggle, even after implementing a reliability program. Why do failures continue to occur to even the best reliability programs? Many times, we immediately focus our thoughts on people, process, or technological issues. In reality, the real answer is data. Data is the lifeblood of any maintenance or reliability program and is the basis of decisions, the steward of objectivity, and the driver for sustainability.

> Data is the lifeblood of any maintenance or reliability program and is the basis of decisions, the steward of objectivity, and the driver for sustainability.

The manner by which reliability is achieved is experiencing a much-needed paradigm shift in the way it uses data. In this article, we'll explain how you can leverage this industry shift in the approach to data to implement data-driven reliability at your facility.

What's wrong with reliability today?

Before diving into the benefits of data-driven reliability, it's important to examine its current challenges. A few common reliability challenges that many facilities face today include:

- The inability to quantify the value of maintenance, resulting in qualitative decision making
- A lack of focus on what drives availability and performance
- Overconservative analyses based on intuition rather than data, resulting in performing activities that increase costs and do not add value to the program
- Siloed execution, primarily stemming from the various types of equipment, resulting in most organizations conducting maintenance and reliability work in functional groups separate from each other

With all of the information and tools that are available today, why do facilities still face these challenges? For starters, from the first day we are exposed to reliability programs, we are compelled to focus on the process and function of the program itself rather than how to use data to drive the analysis. All too often, facilities build their programs based on industry best practices such as a standard risk analysis matrix or criticality analysis and expect them to produce analytics that will help them increase their safety and availability, all while only being fed qualitative information. While this qualitative, program-based approach will produce some initial results at first and will often be considered a success, it is subjected to sustainability issues. The underlying commonality is that industry, in general, is bound by the constraints of best practices that promote intuition over data.

At the end of the day, your program should not be the driver of data; the data should drive your program. Instead of implementing and running a program that is designed to produce a qualitative result, the maintenance and reliability of an asset, and facility for that matter, should be rooted in what the data is telling you to do. An example of this is a standard oil change in a pump. In most cases, the condition of the oil in a pump is not sampled or tested to determine when it needs to be changed. As a result, most changes are performed on a time-based manner, whether or not the oil actually needs to be changed. Understandably, it could be argued that the economics of this type of activity does not merit analysis and that it's cheaper to just change the oil. However, when you consider studies that indicate the number of time-based oil changes on a macro scale, the economics do come into play.

In some ways, the industry already uses data to drive reliability, but there are many more ways that it does not. Many people who perform mechanical integrity inspections will say that they already use data to drive their inspections. While the inspection of fixed equipment is modeled somewhat after this data-driven approach, if these inspections are truly driven by data, then why is there so much conservatism built into the methodologies?

On the other hand, a typical reliability centered maintenance program utilizes more of a static criticality analysis ranking where the consequence of failure (CoF) is based on some parameters initially set forth in the design of the implementation and a high, medium, or low likelihood of failure that is usually someone's opinion. Now, couple these with fact that both fixed and non-fixed programs have built-in levels of interpretations as to how they should work (which greatly varies) and it's easy to see why costs and availabilities fluctuate and programs stagnate over time.

What does data-driven reliability look like?

In its simplest form, the core of a data-driven reliability program is having the ability to understand the uncertainty surrounding the data you collect from an asset. Specifically, understanding the uncertainty through characterizing the variability of the data to estimate the probability of failure (PoF) of an asset. When the variability is not characterized and the uncertainty is high, there is less confidence in the failure estimation. Conversely, if the variability is well defined, then the range of uncertainty is understood thus giving a clear estimation of the asset's failure profile, or PoF.

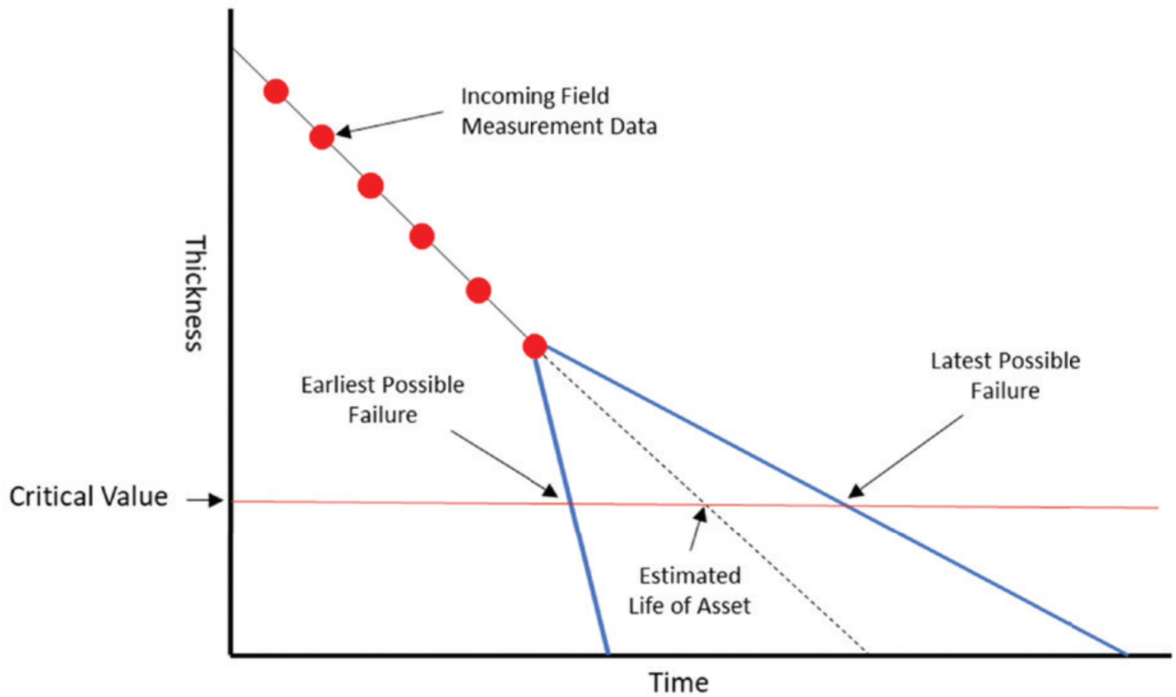

Figure 1. Example LVC for trending wall thickness to determine an asset's PoF.

There are a few philosophies that facilities need to embrace to successfully implement data-driven reliability:

- All data tells a story but not all data improves reliability. A typical facility has millions of data points that are being captured, stored, and trended. However, when you really start to consider how to use data, estimates show that only about 30-40% of it impacts the reliability of a facility.
- Big data is a term that is widely used today to describe using data to make decisions. However, running a data-driven reliability program with large amounts of data is fairly easy. The difficulty arises when you only have a couple of data points. This is a very common situation, especially regarding fixed equipment.
- Data-driven reliability impacts both fixed and non-fixed assets. Whether it's a pressure vessel, pump, or valve, all of these assets must operate safely and reliably and have data that can be analyzed.
- Regardless of whether a maintenance task is preventive or corrective, data-driven reliability alerts facilities whenever a maintenance activity needs to be performed to ensure the operation of the asset.

How does data-driven reliability work?

To calculate an asset's PoF, a data-driven model utilizes a variety of statistical probability modeling curves in conjunction with incoming predictive and inspection measurements, process data, and other inputs to translate the measurements collected in the field into a calculated PoF.

This sounds straightforward. Probability statistics have been used for years in a number of applications; however, the question remains: how can you use your asset's specific data to determine its probability of failure?

Oftentimes, PoF statistics are used on a population of assets rather than an individual asset. For instance, a reliability engineer trying to determine the PoF of a pump would gather the failure data from a group of pumps that have a similar make/model or are the same type of pump. The engineer would then determine the Mean Time Between Failure (MTBF) for each time a failure occurred and would then use a linearized form of the Weibull distribution to calculate the PoF for the combined population. The result would be something like "we predict that we will experience 1.5 failures in this population of pumps this year." The problem with this approach is that you do not know what pump is going to fail or when a failure in an individual pump would occur. So, for all of the effort that was put into the PoF analysis, the only tangible data that you got was being able to get a better insight into where to focus your predictive efforts to keep an eye on where you think a failure might occur.

How do you make the transition from the data you have collected about your asset to a calculated PoF? It's done through what's called a Lifetime Variability Curve (LVC). The LVC is the means by which **quantitative** data is used to project future performance and degradation while **quantifying** the uncertainty in how said data is behaving. Ultimately, this will predict when a failure will likely occur and therefore when an action should be taken.

Figure 1 is an example of an LVC for a piece of pipe where wall thickness is the measure being used to determine the PoF.

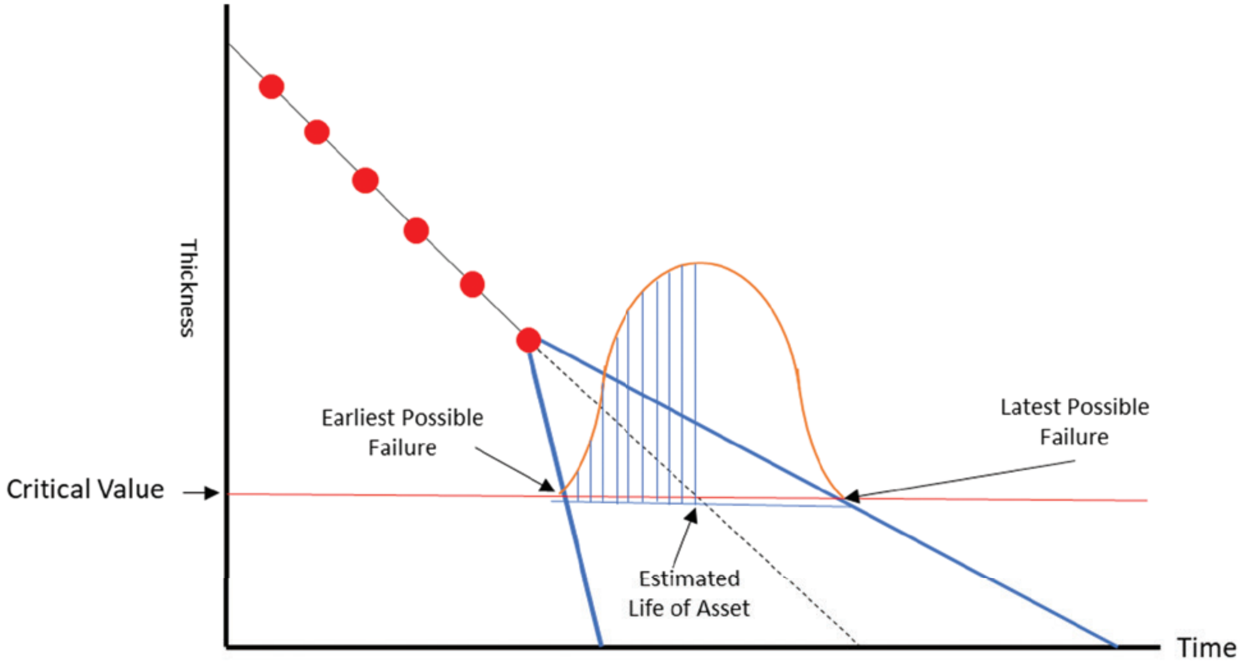

Figure 2. LVC for trending vibration to determine an asset's PoF.

In the LVC, the Y-axis corresponds to the value you want to trend, and the X-axis is time. The scale of each axis depends on the type of data that is being trended such as thickness, velocity, speed, years, months, hours, or seconds.

As you can see, the incoming field measurements data is represented by the descending red dots. This is the data that is being collected on a particular asset. In this case, these are measurements taken at a particular CML. As a new data point is measured, the measured thickness of the wall is decreasing.

There are four other values that make up the LVC.

The first is the Critical Value. The Critical Value (CV) is the threshold line that corresponds to the value that is considered to be the point where an action must be taken based on the data that is being trended. To clarify, this does not mean that the asset itself has failed but that the specific data being collected has passed an acceptable level indicating that an action should be performed.

The value of the CV depends on the type of data that is being trended. For fixed equipment, the CV will primarily be the minimum integrity requirement for safe operation (i.e., Critical Limit for thickness; remaining creep life; or delta T). For non-fixed equipment, the CV will come from a variety of sources depending on what data is being trended.

It is important to know that although most CVs will be in one direction (up or down) with respect to the data that is trended, there are some LVC data trends that will require an upper and lower CV. For instance, when trending process data, there will often be an upper and lower limit by which the asset must operate within instead of a direct correlation that increases or decreases. The upper and lower limits help determine whether the process results in too little or too much flow. The same can be said for certain parameters in the lubrication analysis and other data types.

The second is the Estimated Life of a component or asset. The Estimated Life of a component or asset is the point on the LVC where the estimated lifeline of the asset crosses the Critical Value. This is the point where, based on the measured data, a mitigating action will need to be performed.

The Estimated Life of an asset is calculated by taking the measured data that is collected through a regression analysis and forecasting its interpolated trajectory line in order to estimate where it will cross the X-axis. The point where it crosses the CV is the date where it is deemed unacceptable to run and action should be taken to mitigate the failure.

The third and fourth points are the earliest an asset can possibly fail and the latest an asset can possibly fail. These points are the statistical estimations of the earliest and latest probability that a component or asset can fail based on the measured data.

Each time a new data point is entered and the LVC is recalculated, the Estimated Lifeline date can either extend out or occur sooner than the previous date. It all depends on the data. For our pipe wall example, if a measured thickness reading was to be less than expected based on the previous measurements, then the new failure date will extend past the previous failure date. If the measurement indicates a greater wall loss, then the failure date will be sooner.

Also, each time a new data point is entered, the time range between the earliest and latest possible failure points will decrease. This is because each time a new data point is measured, the accuracy of the LVC estimations increases. In turn, the estimations of the earliest and latest failure points become more accurate as well. In essence, all three of the lines (Estimated Life, earliest failure, and latest failure) all converge at the point of failure.

It is true that as the three lines converge, the estimation of the failure dates becomes more accurate. However, the key statistical concept to understand here is that with every new data point, the convergence of the lines is decreasing the variability associated with the data. In addition, the convergence also decreases the uncertainty, bringing the models closer to the true failure points.

From LVC to Probability of Failure

In order to translate the LVC into a useable form for the PoF calculations, a statistical representation of what the data is telling us must be determined. This is done by calculating a population distribution in regard to the three lines in the LVC. For the below example, the Weibull distribution will be used to calculate the PoF curve.

There are three primary steps that use the observed data to calculate the PoF and represent this as a Weibull function.

The first is what's called the Probability Density Function (or PDF). The PDF is a continuous representation of a histogram that shows how the number of failures is distributed in a period of time. If you recall from the LVC, the earliest and latest possible failure points are calculated based upon the failure date population distribution. That population distribution is the PDF. **Figure 2** illustrates this on the pipe wall thickness example. You'll notice the bell-shaped line extending from the earliest failure point to the latest failure point. This is where the population distribution is developed.

The next step is to determine the parameters for the Weibull curve. The Weibull distribution curve uses three parameters in its calculations: Alpha–the Life parameter, Beta–the Shape parameter, and Gamma–the time offset parameter. In order to obtain these parameters, a regression model is applied to the data in the PDF from step one to calculate the coefficients for the resulting regression line. These coefficients are then converted into the parameters Eta, Beta, and Lambda for use in the Weibull curve.

The final step is to plug in the values and calculate the PoF for the asset over a period of time. This is done by calculating the PoF of the asset over a period of time.

Conclusion

While probability statistics have been around for decades, it has never fully found its way into the core of reliability. With the advent of the Lifetime Variability Curve (LVC) and the data-driven reliability model, it is now possible to capture the specific data and apply probability calculations on an asset-by-asset basis in order to see how actions, or inactions, will affect the overall availability and cost of a facility. In order for you to elevate your reliability program to its full potential, the data-driven reliability model must become the guiding first principle of your program.

How can you shift to a data-driven reliability model at your facility? The first step is to embrace the next evolution of reliability modeling–Quantitative Reliability Optimization (QRO).

QRO is an approach to reliability that streamlines current reliability models into a single, comprehensive analysis that will enable reliability and operations leaders to make smarter reliability decisions for their facilities. QRO leverages the best elements of current models and methodologies while introducing novel analysis concepts to drive process safety, improved availability, and overall facility performance. QRO takes large volumes of relevant reliability and economic data as inputs, analyzes system reliability performance using cutting-edge data science, and delivers optimized reliability plans.

With QRO, facility leaders will finally be able to manage reliability through a single data-driven lens. ■

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

CONTRIBUTING AUTHOR

Sean Rosier

Sean Rosier is a post graduate qualified engineer with 25 years of experience in reliability engineering, maintenance management, operations management, project engineering, project management, and market research. He is experienced in a range of industries including oil & gas, chemicals, steel production, mining, manufacturing, and LNG, in roles, including consultancy, service contracting, field execution and leadership in both technical and operations management. He is currently the Principal, Asset Management Solutions for Pinnacle.