Analytics Case Study Interview

September 3, 2020

Before we get started…

What is the goal of this case study?

You are going to be looking at 4 different, specific examples of issues that arise on pump & drive systems. Across these examples, you will be asked to:

1) Design features that would differentiate these conditions from their progression in peak acceleration and/or peak velocity over time, and/or from their individual burst time waveforms and FFTs.

   Utilize a diagram to show how your features would be calculated, and then create a python script that is able to read the associated files for each failure from csv and generate the feature sets.

   Note: You will begin this exercise by assuming that there is no variation in the given issues. E.g. the provided example of imbalance is identical to all other examples of imbalance that would occur.

2) Develop an algorithm that would be able to separate these specific four examples based entirely on rules or logic on the suggested feature sets. (Note: do not use a machine learning approach as you only have one sample of each fault.)

   This algorithm should be documented using a flow diagram as well. Be critical and consider how this output should be presented and provide your rationale. Should it be a single point in time? Should it be a stream of probabilities over time?

   Hint: There's no "right" answer. We just want to see how you think.

3) Consider why your algorithm would or would not generalize well to other examples of these types of failures. Suggest several, rules-based changes to your algorithm that could be incorporated to generalize but do NOT implement them.

4) Describe how you would take your rules-based approach and implement it within a machine learning or statistical framework to further improve results and generalizability. Consider how that approach would work well with low volume datasets.

   Hint: Deep learning has not worked well historically for our application.

How should I interpret the images and data structures?

Each example consists of several key pieces of information, which directly relate to the way in which we sample data from our sensors. By understanding how our sensors collect data, you'll be able to understand what information has been provided.

Our sensors:

- Our accelerometers take a "burst" sample at a frequency of ~1x / min. This burst is tied to the vibration at that instant in time.
- As mentioned, these burst samples do not correspond to an entire minute of continuous readings. Rather, they corresponds to a much small "instantaneous" interval of time (~0.2 sec).
- These individual bursts are sampled at a much higher frequency (typically 8192 Hz) than the 1x / min mentioned, providing discrete FFT and time waveforms.

Our graphics:

- The time series graphs that span over multiple hours (see slide 7) are either the max acceleration or the max velocity of individual bursts. This would be calculated as the peak of the burst sampled at 8192 Hz.
- From the max acceleration / velocity graphs, an individual reading can be examined with much more granularity. These graphs are shown in either the frequency domain graphs as an FFT or as the time waveform.

Our data:

- The data provided for each failure spans across an ~hour window around the event issue. These datapoints will come from sensor "monitoring points" in the x and y axis orientations.
- Each data point will consist of a timestamp and the raw acceleration values for the burst in g's. If an individual row of data is plotted, it would correspond to a time waveform sample from an individual burst.
- Therefore, the max of each row item plotted over its corresponding timestamp would match up to peak accel over time.

Other notes…

1. We understand that this is a difficult case study and requires a significant amount of time to formulate a plan of attack and execute against.

2. We also understand just how ambiguous the problem statement is. You might be confused on exactly what you are trying to accomplish, and we get it. Fight with yourself to clarify the problem – that is half of the battle.

   We fight to find this clarity everyday. We've provided you with information and want to see what you believe would drive actionable insight and generate value for our customers.

3. We do not expect perfection or complete generalization in your algorithm, but we do expect significant progress, we expect a program that is able to separate these specific examples. The more insight you can generate and the more generalized your approach, the better.

   Often, as scientists, we want to understand a problem in its' totality and identify every nuance, but that's not always practical. We must generate incremental value in our research at every step. Consider the pareto approach, where 80% of results are driven by 20% of casual factors. Apply it to your approach. How can I be pragmatic yet thorough?

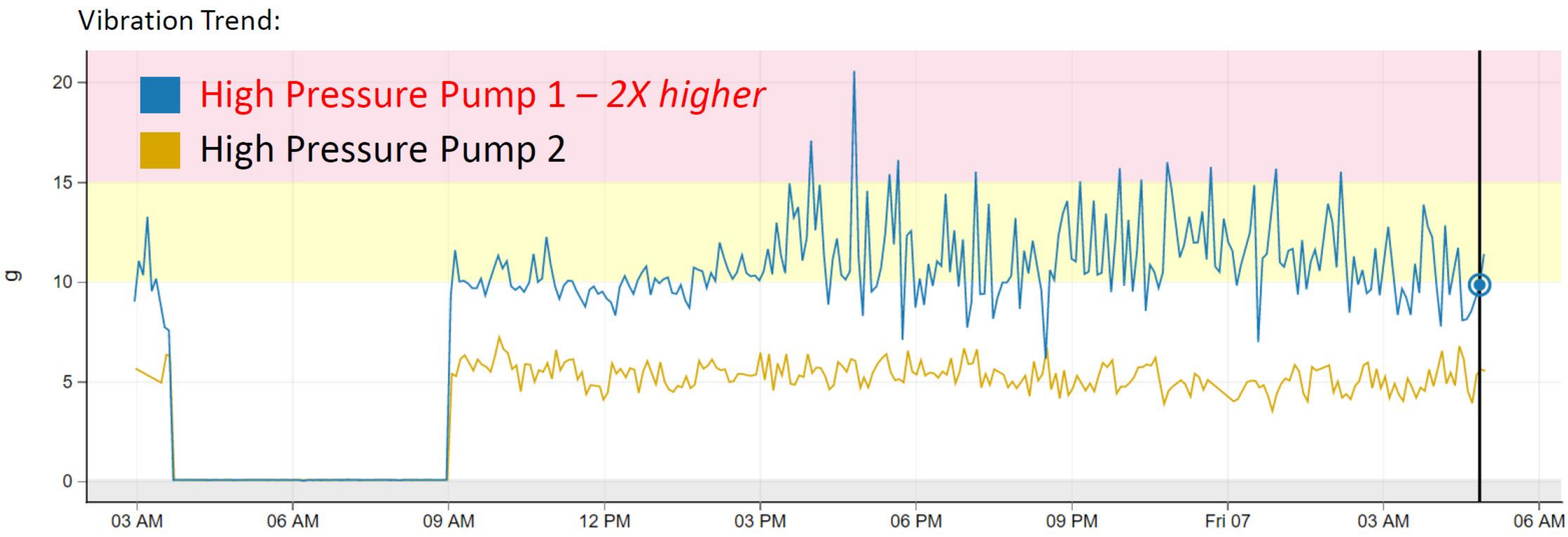
Consider this interview both ways. If this problem excites you as much as it excites us, we'd love to have you, as you'll be doing it everyday!

Good luck, happy thinking and happy coding!

Pump Example #1 : Piston Wear

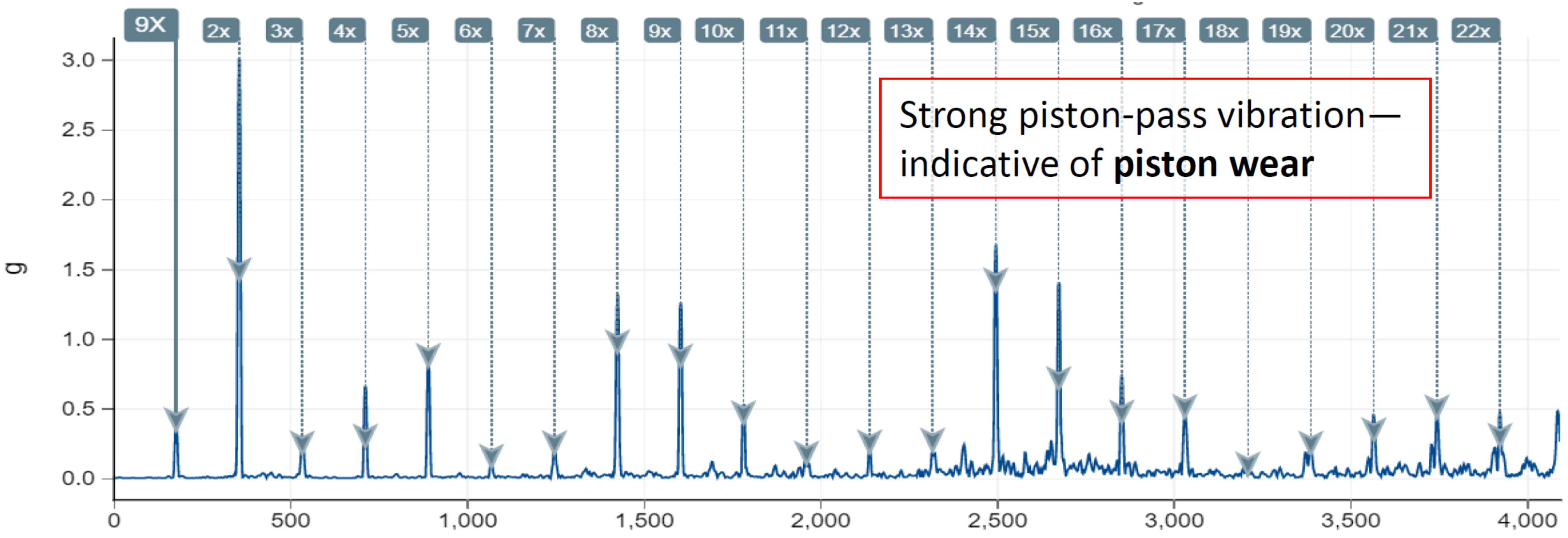
Time Domain

<u>Time Domain</u> comparison of two similar pumps with one pump showing a <u>2x</u> increase in vibration compared to its sibling …



Frequency Domain

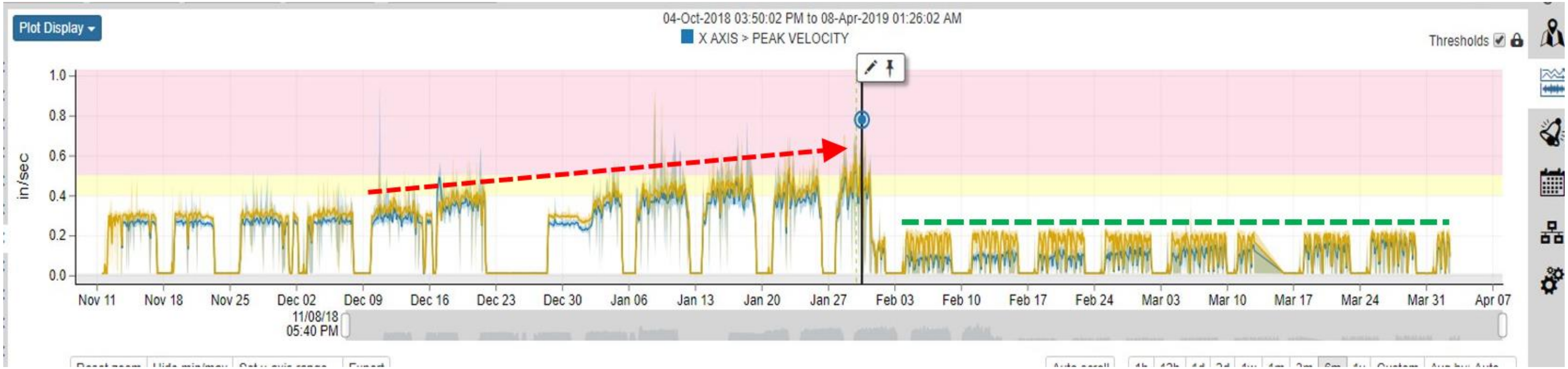
Typically, the time-domain will alert us that an issue exists, while the frequency-domain often provides some more insight into the potential problem. Taking a closer look, it is determined that Pump 1 is showing a repetitive increase in its piston-pass vibration (denoted by the "x" per pass) …



Pump Example #2 : Imbalance

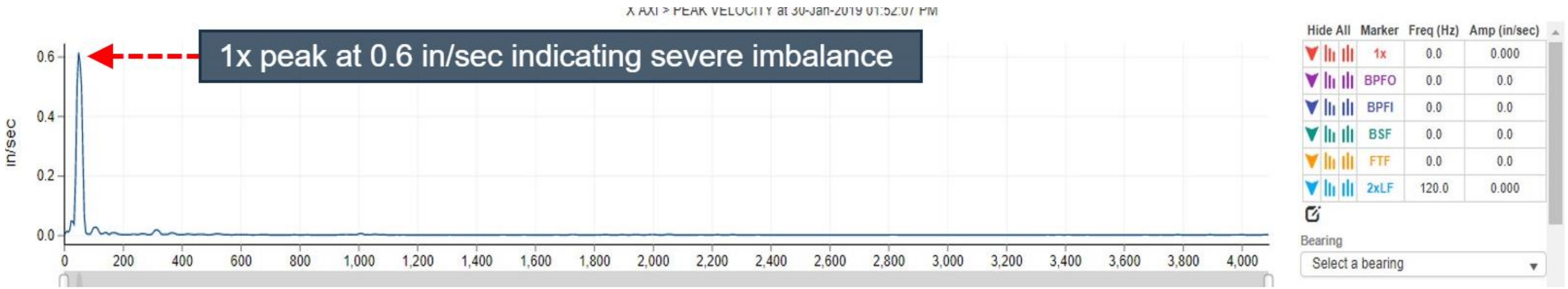
Time Domain

Time Domain view of both peak velocity and peak acceleration for a suspected fault condition …



Frequency Domain

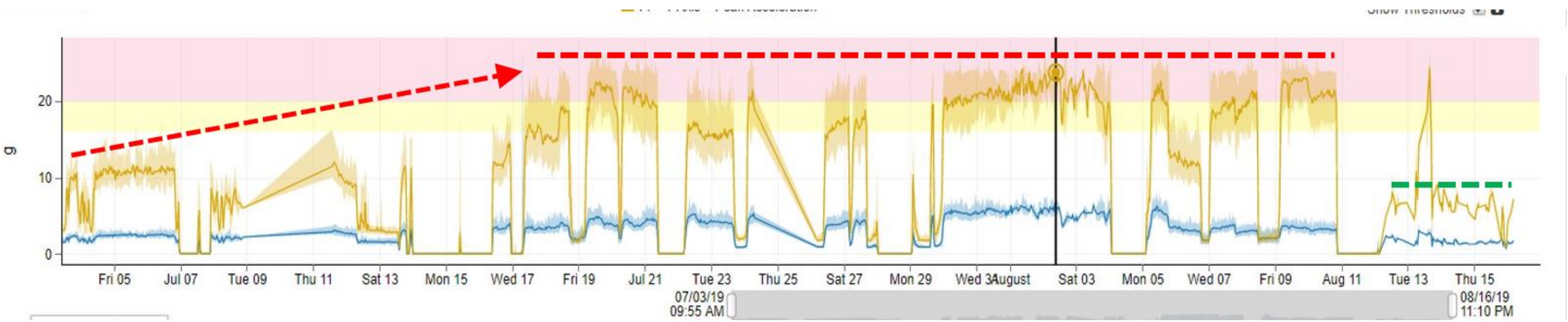
The frequency domain view shows us a high peak at the 1x frequency indicative of severe imbalance within the pump motor …



Pump Example #3 : Bearing Wear

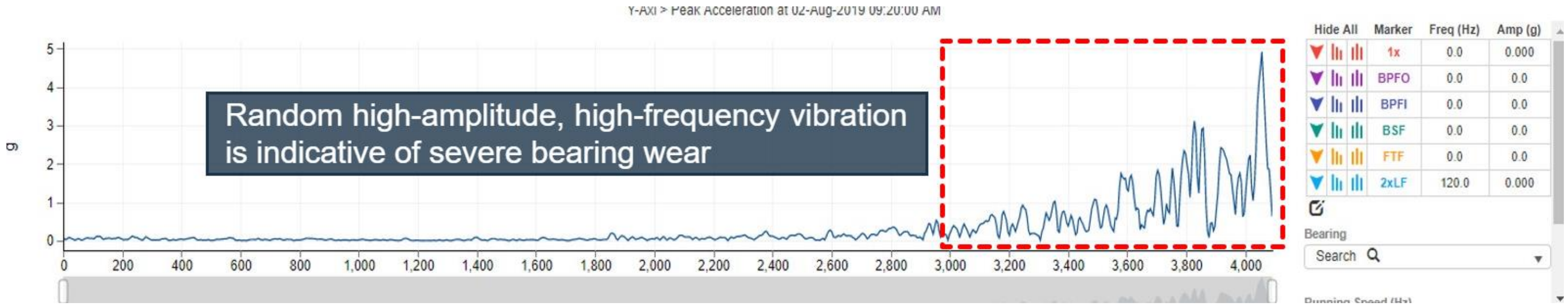
Time Domain

<u>Time Domain</u> view of peak acceleration on both the x-axis and y-axis for a suspected fault condition …



Frequency Domain

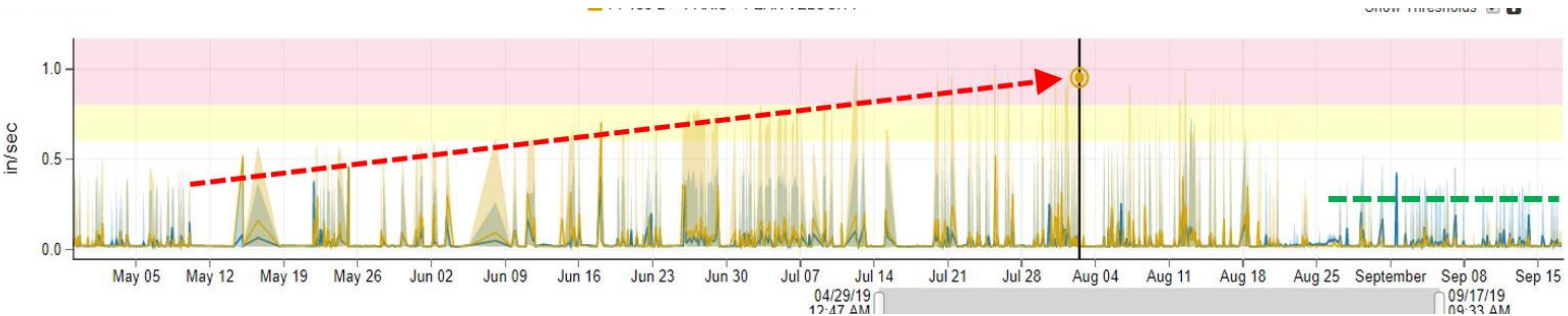
The <u>frequency domain</u> view shows us a high-amplitude and high-frequency vibration which is indicative of bearing wear …



<u>Pump Example #4 : Pump Overloading</u>

Time Domain

<u>Time Domain</u> view of peak velocity on both the x-axis and y-axis for a suspected fault condition …



Time Domain Deep Dive

The frequency domain shows us a main operating vibration at 48Hz that is being modulated by 10Hz which points to an overloaded pump. This modulation is showed again in an individual data burst as shown below in the time waveform.