SAND2019-8557PE

Machine Learning Based Analysis Of Acoustic Emissions During Induced Seismicity

Rachel Willis – University of North Carolina, Wilmington

Hongkyu Yoon – Org. 8864

PRESENTED BY

2019 Summer Student Symposium

This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories.

Outline

1. Introduction

2. Experimental Setup

3. Acoustic Data

4. Machine Learning Setup

5. Machine Learning Analysis

6. Complications

7. What's Next??

Personal Background

- B.S. in Physics and B.S. in Mathematics and Statistics.
- M.S. in Mathematics.
- Summer 2018: NSF Funded REU at UNM
  - Gravimetric Study for Detecting Subsurface Density Structures Beneath Volcanoes
- 2018 – 2019: Honors Thesis
  - Applications of Spherical Harmonics for Elastogravitational Deformations of the Earth
- PhD in Physics or Geophysics

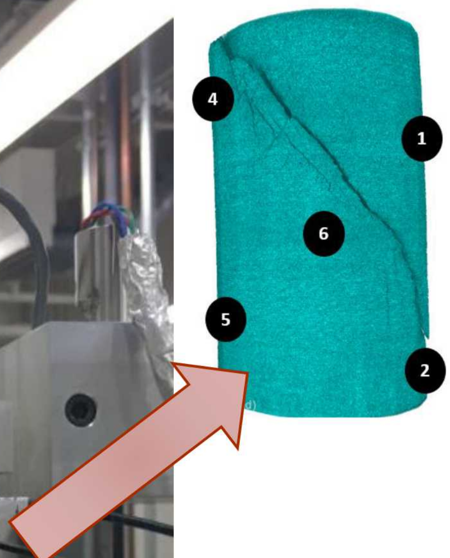
Introduction

We are using machine learning methods to analyze the acoustic emissions of induced seismicity. Acoustic/Seismic precursors may be used to predict fault failure.

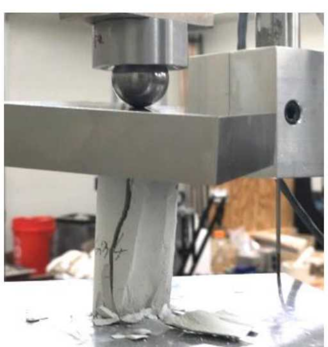
Background

- Seismic forecasting focuses on 3 key points: when, where, and how large.
- Advances in instrumentation resulted in new discoveries of slip processes.
- Acoustic/seismic precursors to failure in materials.
- Precursors may exists in most if not all seismic events

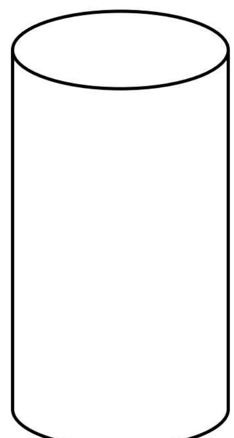
Experimental Setup



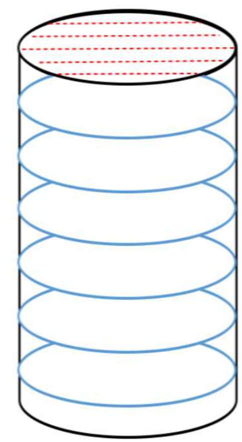
- Six F15α sensors from Physical Acoustics (Channels)
- 200-400 kHz filter to get rid of noise



Cast Sample
C22, C33, & C23

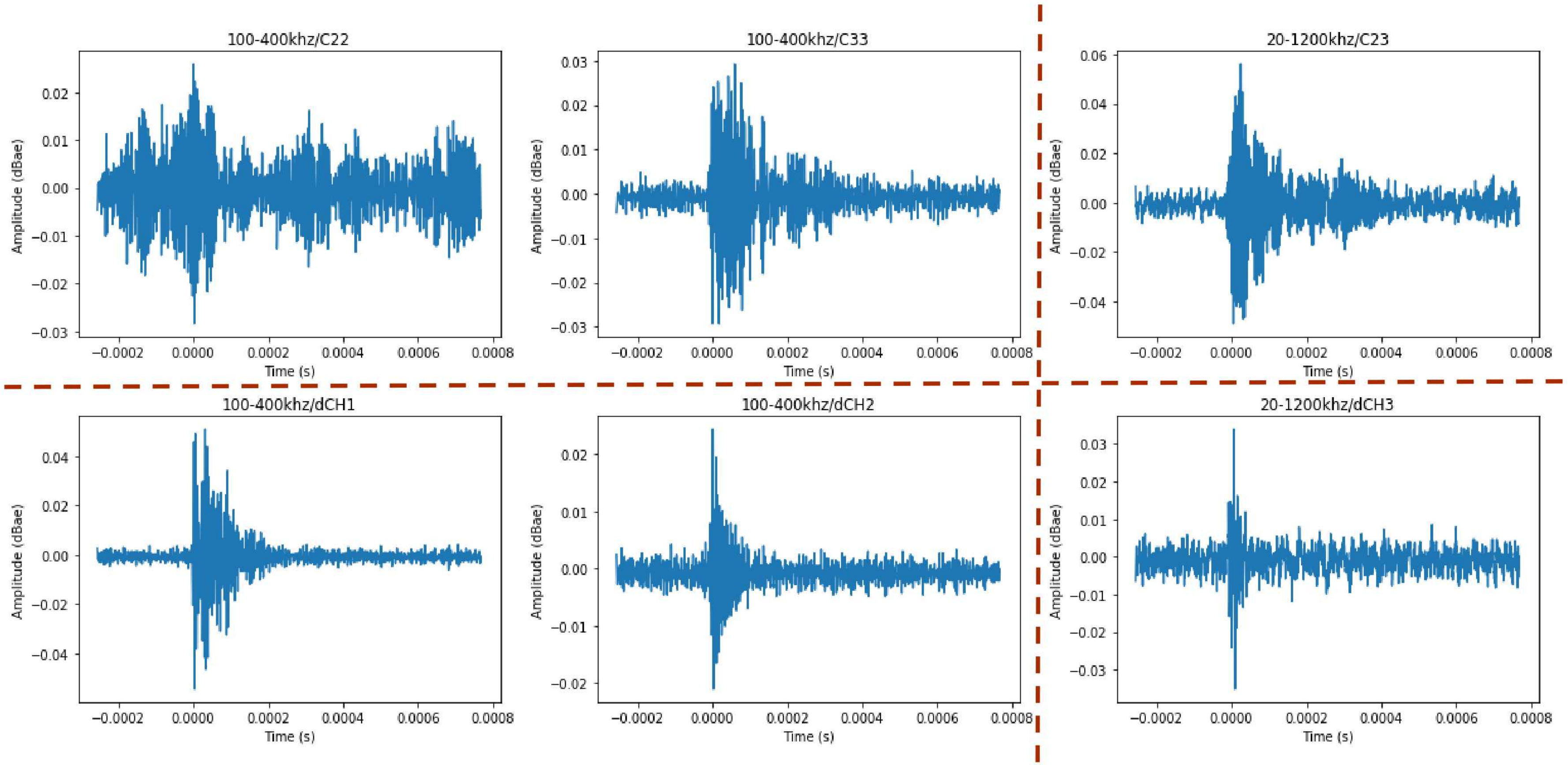


H Sample
dCH1, dCH2, & dCH3



Acoustic Data

6 Samples → 2 Types of Samples → 6 Acoustic Channels Per Sample → 2 Different Frequency Ranges



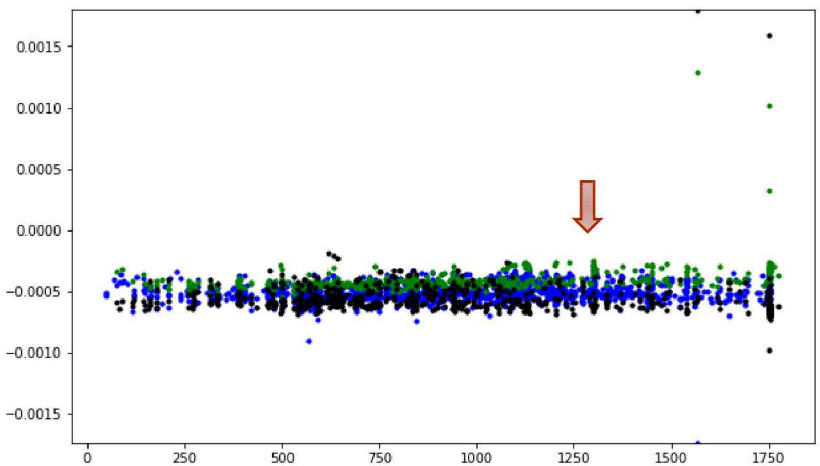
Machine Learning Setup

We apply different machine learning features to find patterns or precursors to predict failure.

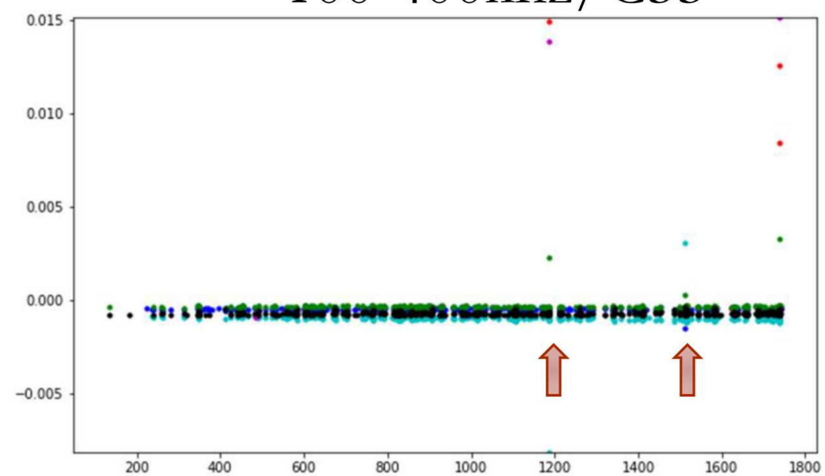
1.
   - Load one set of data.
   - Assign each sensor a different color.
2.
   - Apply and analyze basic features.
   - Ave, Std, Skew, Kurtosis, & Energy.
3.
   - Apply and analyze complex features.
   - mfcc_mean4, mfcc_mean18, percentile_roll50_std_20, & trend_error

Machine Learning Analysis -- Ave

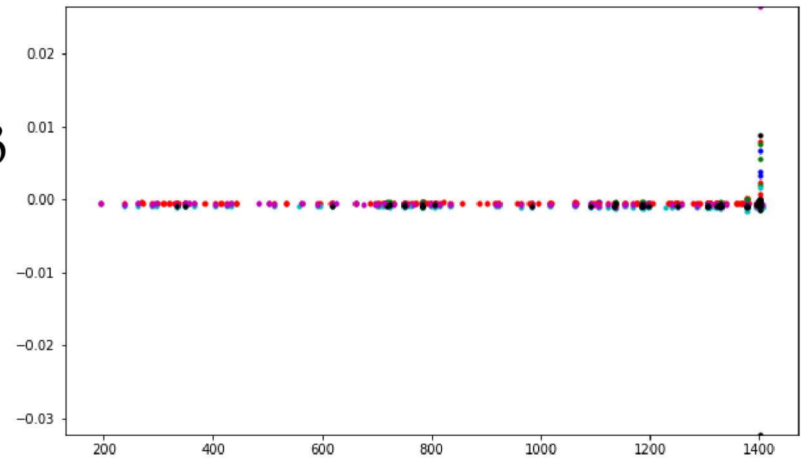
100-400khz/C22



100-400khz/C33



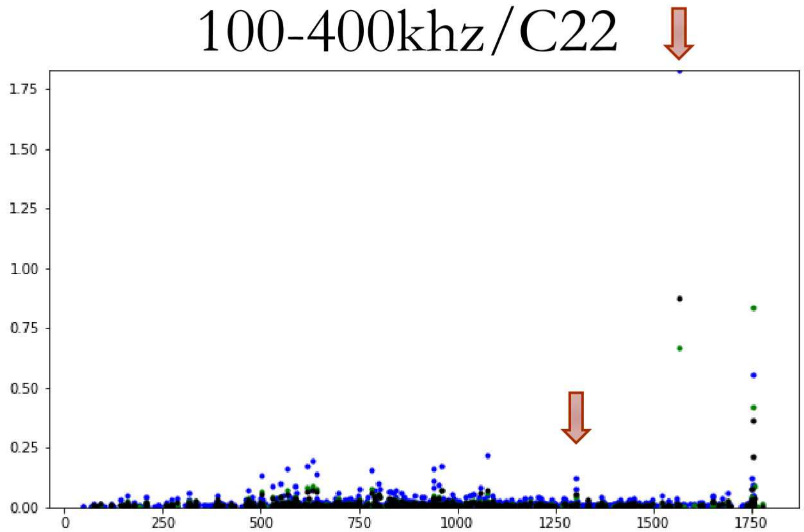
20-1200khz/C23



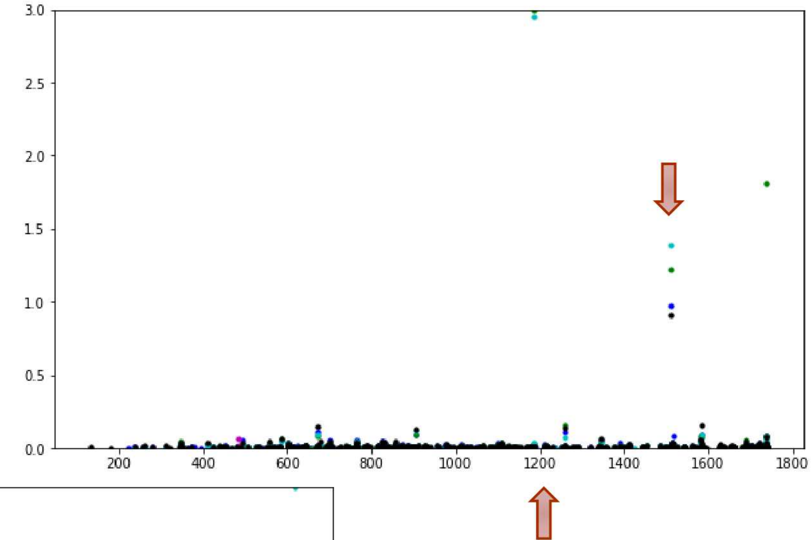
All averages are fairly close to zero. There are a few outliers present in each sample. Possible significant events for C22 around 1300us and 1600us. Possible significant events for C33 around 1200us and 1500us.

Machine Learning Analysis -- Std

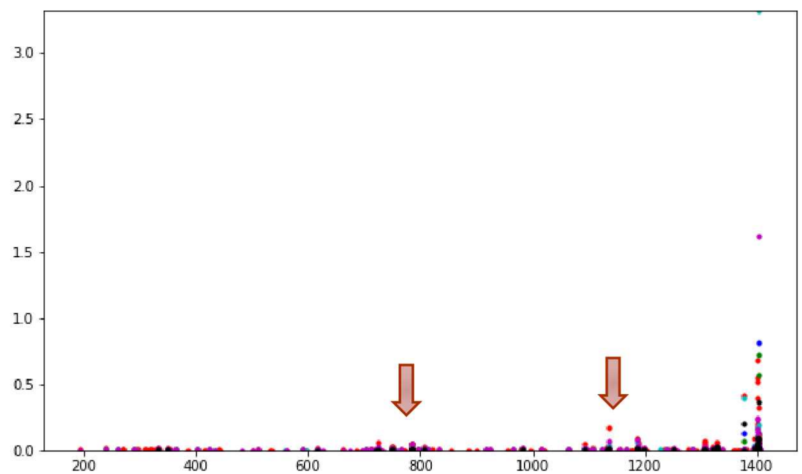
100-400khz/C22



100-400khz/C33



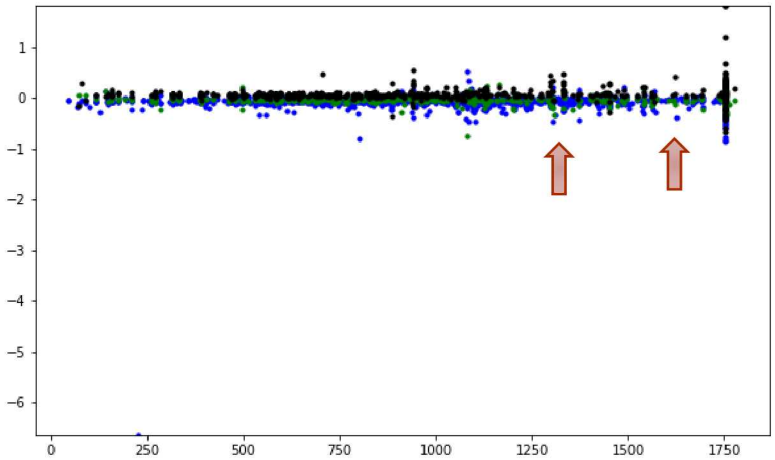
20-1200khz/C23



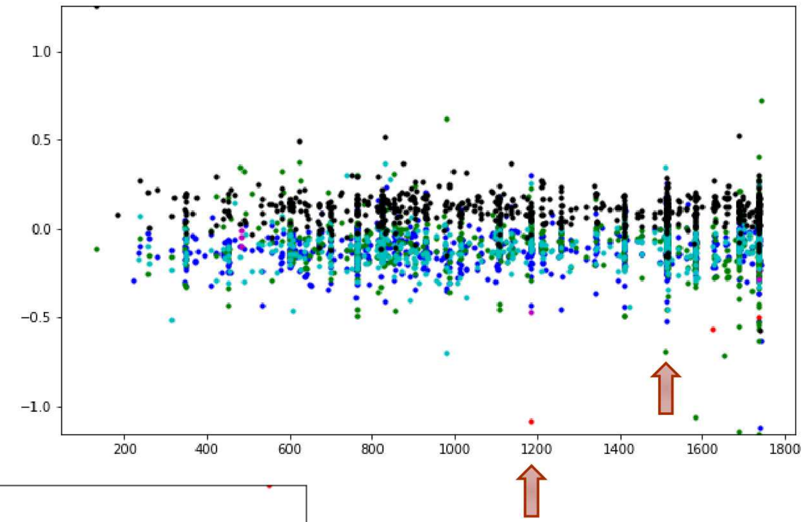
For all samples it seems that the standard deviation is increasing until failure. This means the acoustic signal is changing more drastically as we get closer to failure which is expected. Some filtering may allow us to get a better representation.

Machine Learning Analysis -- Skew

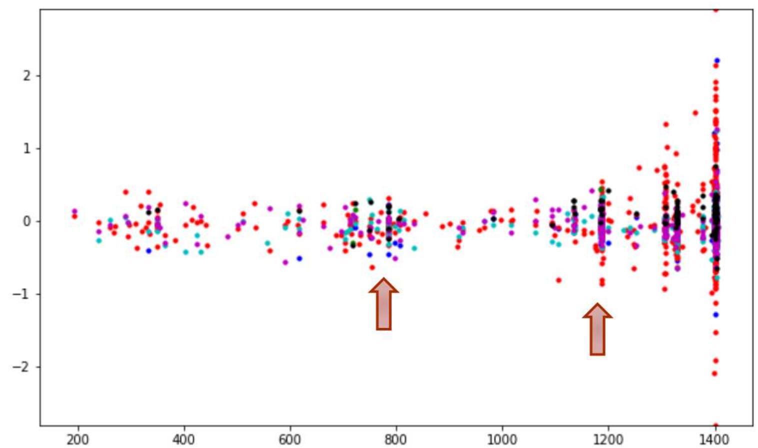
100-400khz/C22



100-400khz/C33



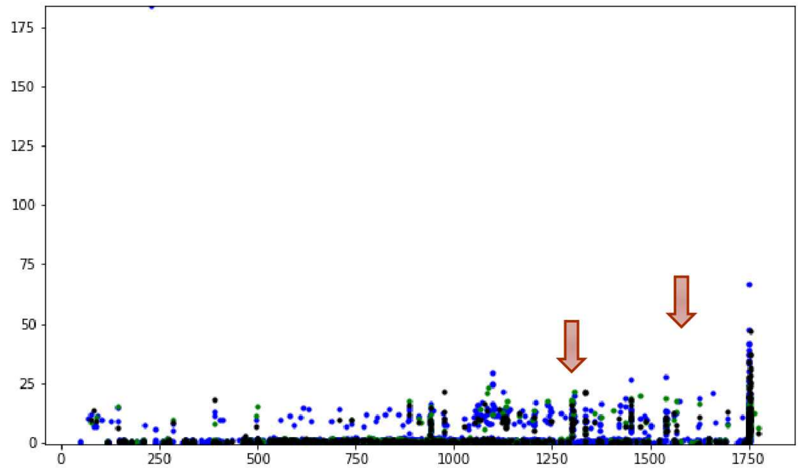
20-1200khz/C23



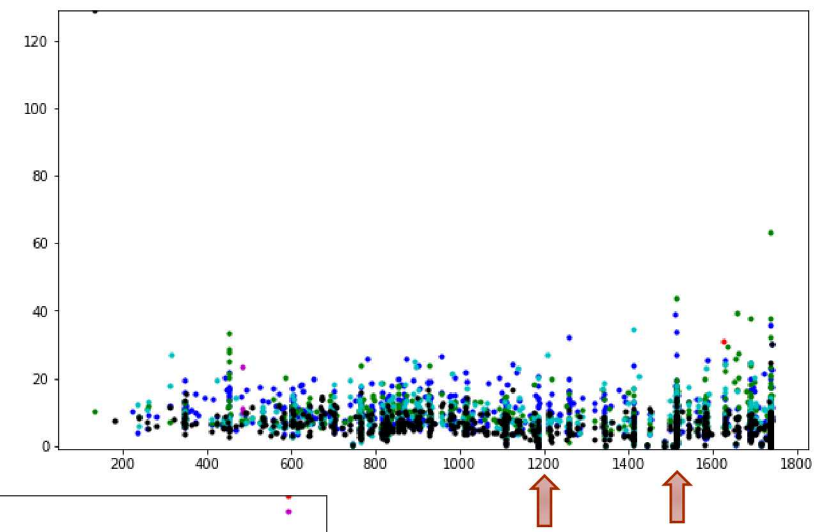
All samples show an increase in skewness leading up to failure which is what we would expect. There are also outliers present in C22 and C33.

Machine Learning Analysis -- Kurtosis

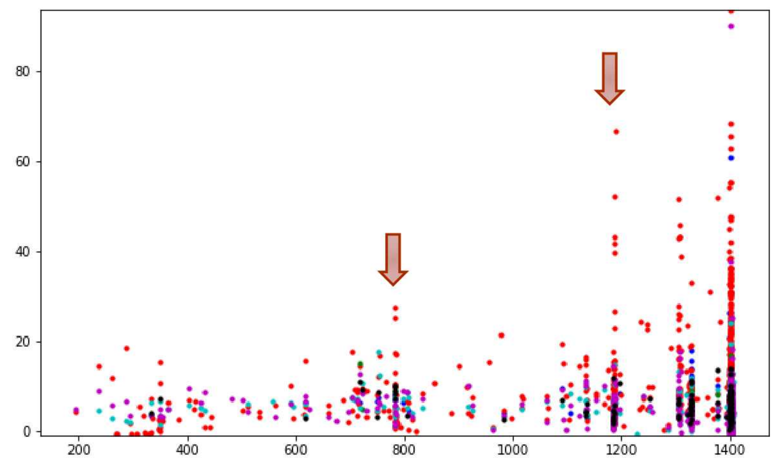
100-400khz/C22



100-400khz/C33



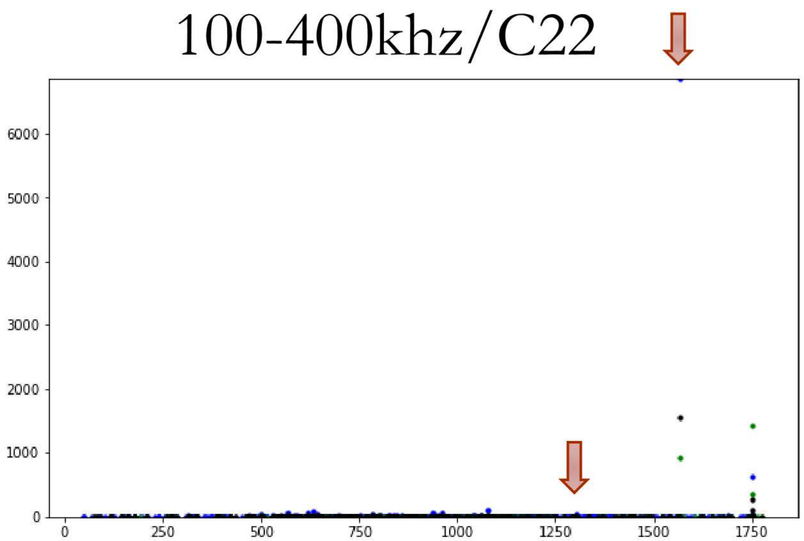
20-1200khz/C23



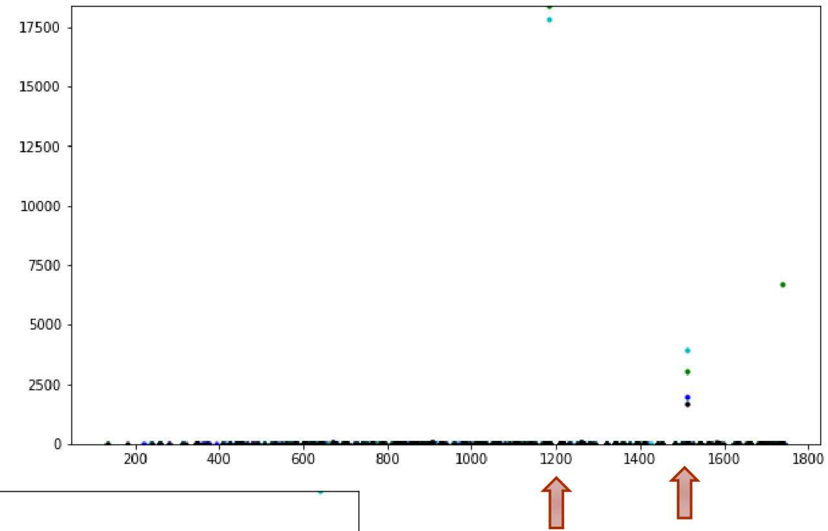
There is an upward trend present in all samples meaning as we approach failure there is an increase in peaks so more events are recorded. Again there are outliers present in C22 and C23. The behaviors between Kurtosis and Skew seem to be similar.

Machine Learning Analysis -- Energy

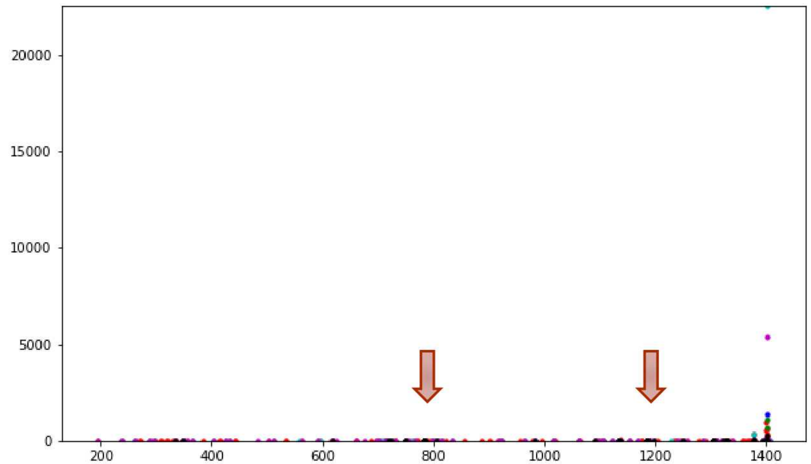
100-400khz/C22



100-400khz/C33



20-1200khz/C23



All samples have significant outliers present. All outliers are from different channels. We could infer that the energy is increasing towards failure. Note energy is the sum of magnitudes squared.

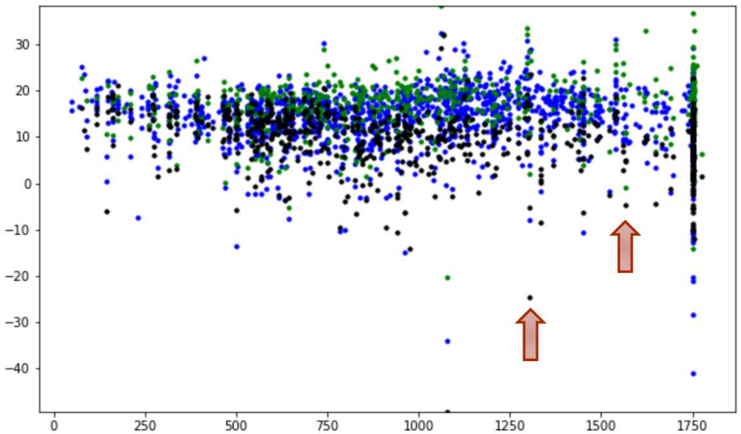
Machine Learning Analysis

We can conclude the following from our analysis of the common machine learning features:

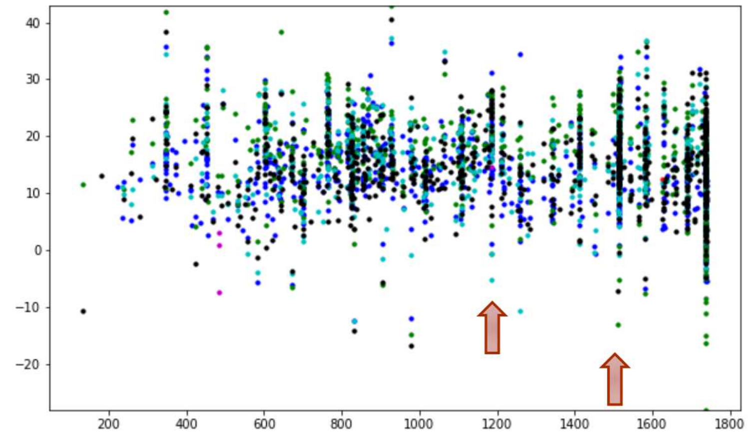
- Several significant outliers, need to filter.
- Skew and kurtosis have a similar behaviors.
- Possible significant events:
  C22 1300 s & 1600 s;
  C33 1200 s & 1500 s ;C23 1200 s.
- Best performing features: skew & kurtosis.

Machine Learning Analysis mfcc_mean4

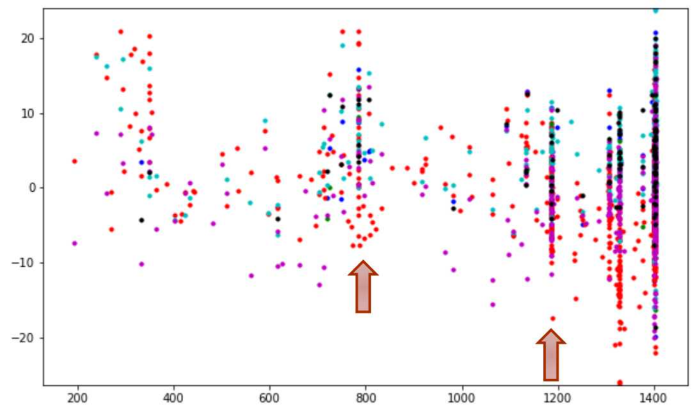
100-400khz/C22



100-400khz/C33



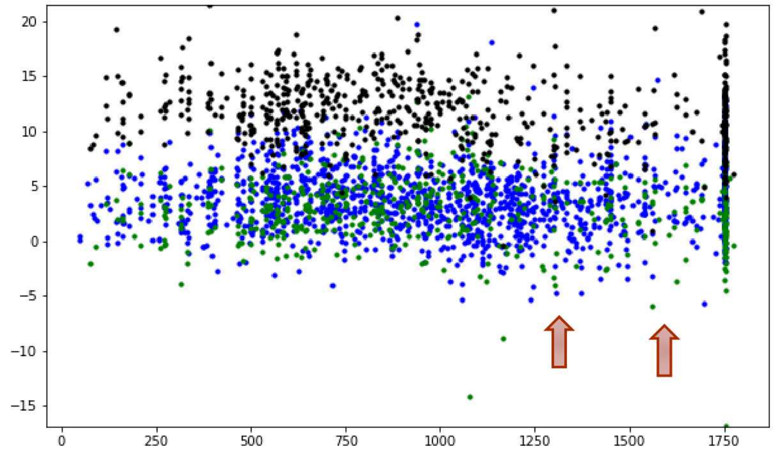
20-1200khz/C23



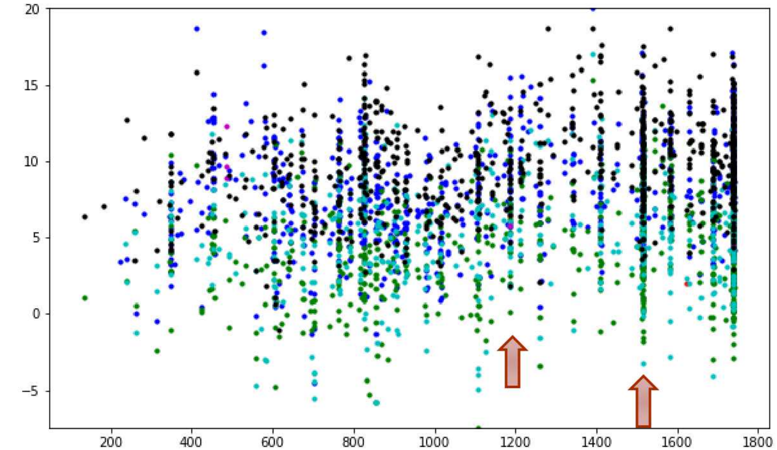
Mfcc represents filterbanks for energies and frequencies. This is targeting low frequencies. When a significant event occurs the data points either jump above or below the mean.

Machine Learning Analysis mfcc_mean18

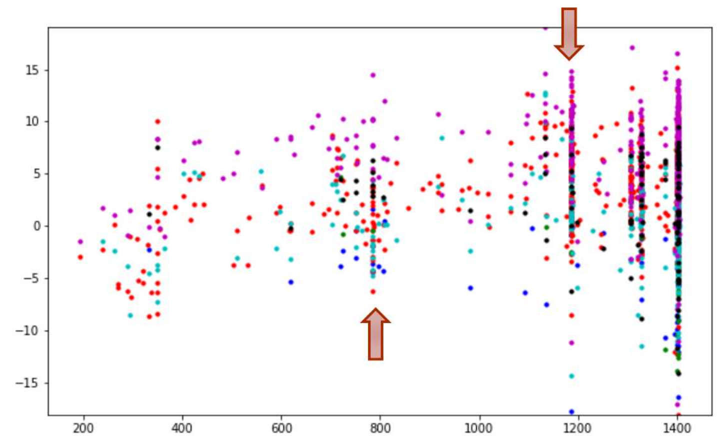
100-400khz/C22



100-400khz/C33



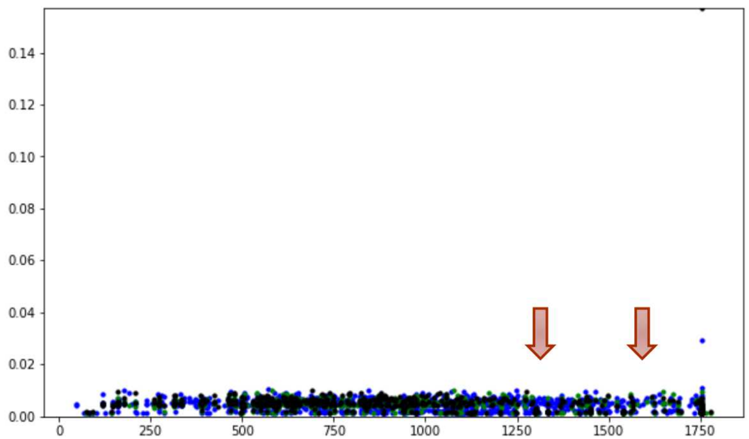
20-1200khz/C23



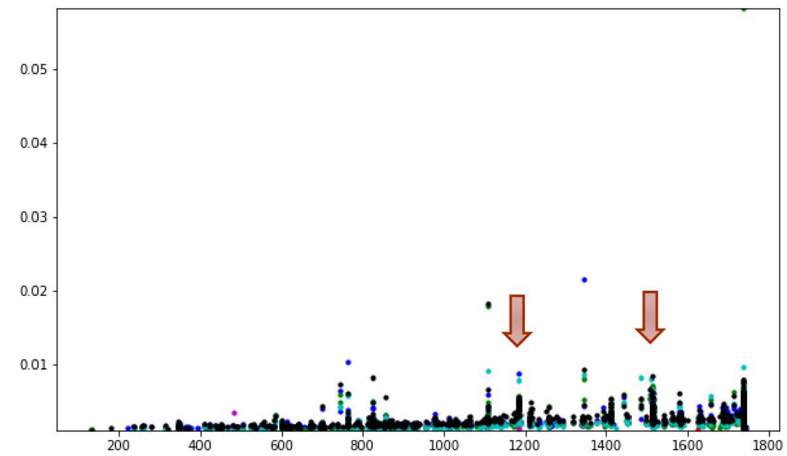
In this mfcc we are targeting higher frequencies and it appears to be less stable than the lower frequency mfcc. This could be due to the large amount of noise present in the data and acoustic propagations.

Machine Learning Analysis percentile_roll50_std_20

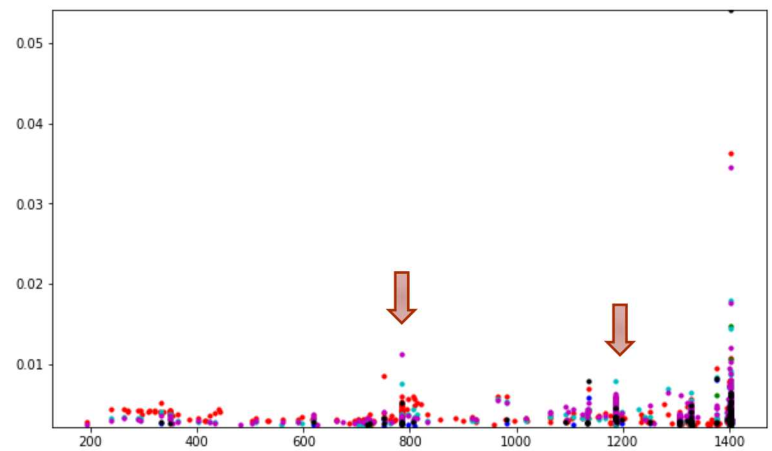
100-400khz/C22



100-400khz/C33



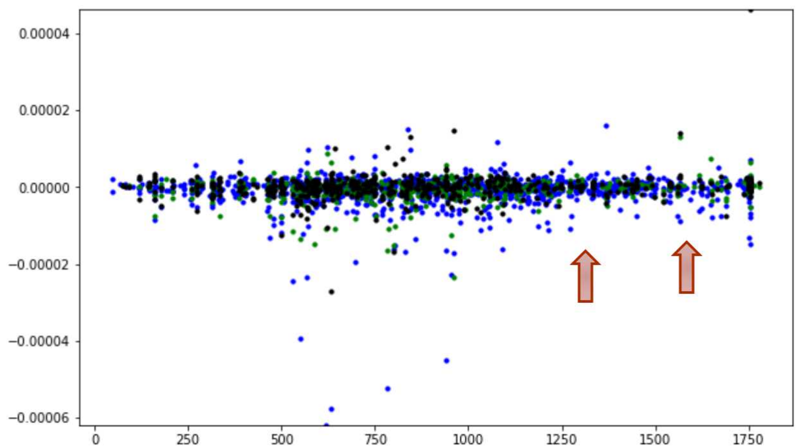
20-1200khz/C23



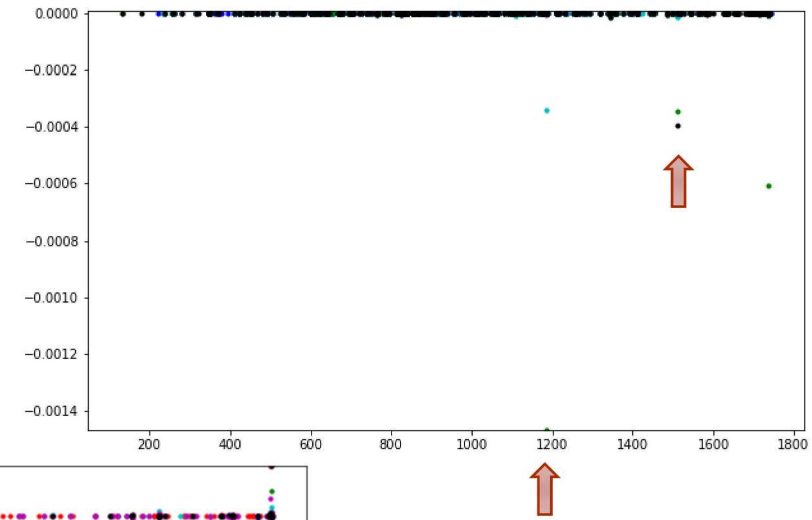
There is somewhat of an upward trend present in all samples; however, there are very significant outliers present as well.

Machine Learning Analysis trend_error

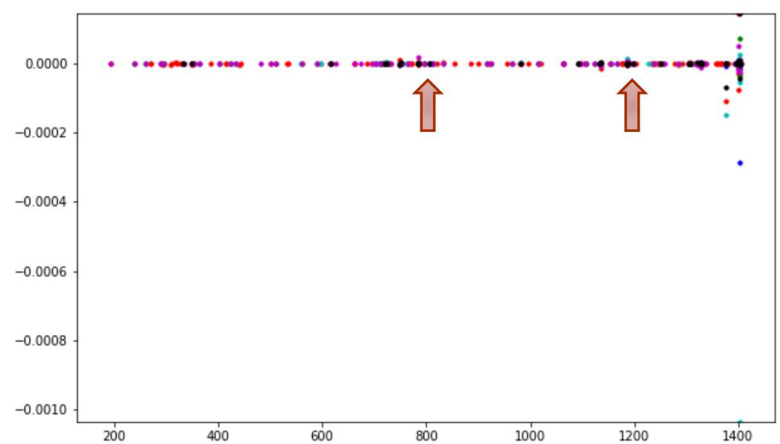
100-400khz/C22



100-400khz/C33



20-1200khz/C23



Again there are outliers present that need to be filtered out. C22 shows a lot of trend error towards the middle of the data which could be caused by noise or events.

Machine Learning Analysis

We can conclude the following from our analysis of the complex machine learning features:

- Several significant outliers.
- mfcc_mean4 has a more stable distribution that mfcc_mean18.
- Possible significant events:

  C22 1300 s & 1600 s;

  C33 1200 s & 1500 s; C23 1200 s.
- Best performing features: mfcc_mean4 & mfcc_mean18.

Complications

- Only analyzed Cast samples due to lack of data available for the H samples.
- There are 2 Cast samples in the 100-400kHz range and only one Cast sample in the 20-1200kHz range.
- There are 6 signal channels; however, not all 6 channels are present in the data files.
- The Cast sample produces a lot of noise making it difficult to find significant events.
- The outliers are in all files and for different channels.
- No visible pattern or definite precursors to failure.

What's Next???

- Apply EMD (Empirical Mode Decomposition) to filter out the noise.
- Add more features to find possible precursors.
- Look at specific sections of time.
- Apply ML Algorithms such as Kernel Ridge.

References

Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, 44, 9276–9282. https://doi.org/10.1002/2017GL074677

Kaggle, Inc. (n.d.). LANL Earthquake Prediction. Retrieved from https://www.kaggle.com/c/LANL-Earthquake-Prediction/overview