SAND2020-1620C

Identifying and Clustering Activity in Seismic Waveforms

Renee Gooding, Erick Draayer, Nicole McMahon, David Stracuzzi, Dylan Anderson

Problem:

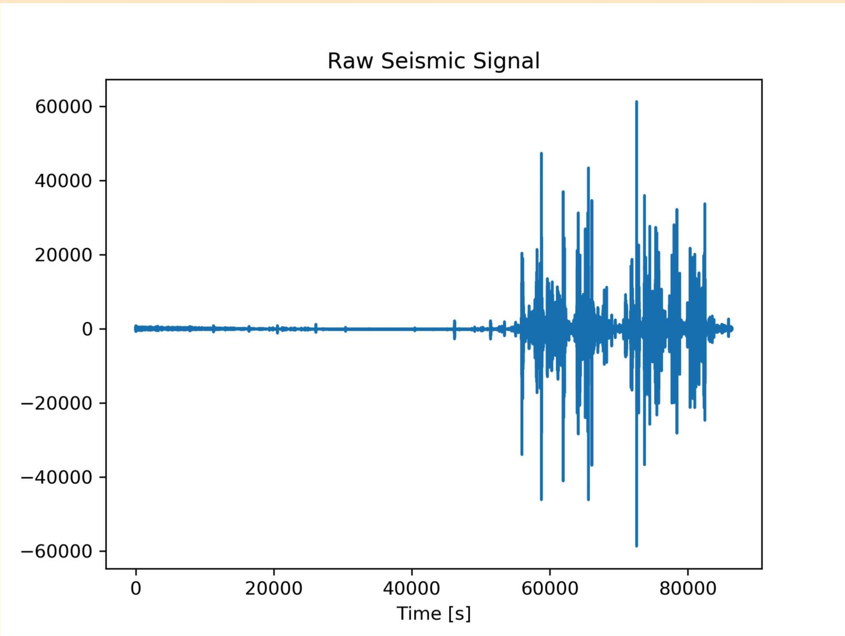
Seismic sensors can provide high temporal resolution data about nearby activity, such as use and movement of vehicles and other heavy equipment. However, traditional seismic analysis tools are designed for earthquake or other large events and do not perform well at extracting local pattern-of-life information.

Questions:

1. Can we detect pattern-of-life activities in seismic waveforms?
2. Can we cluster the detections such that similar activities are grouped together to support interpretation?

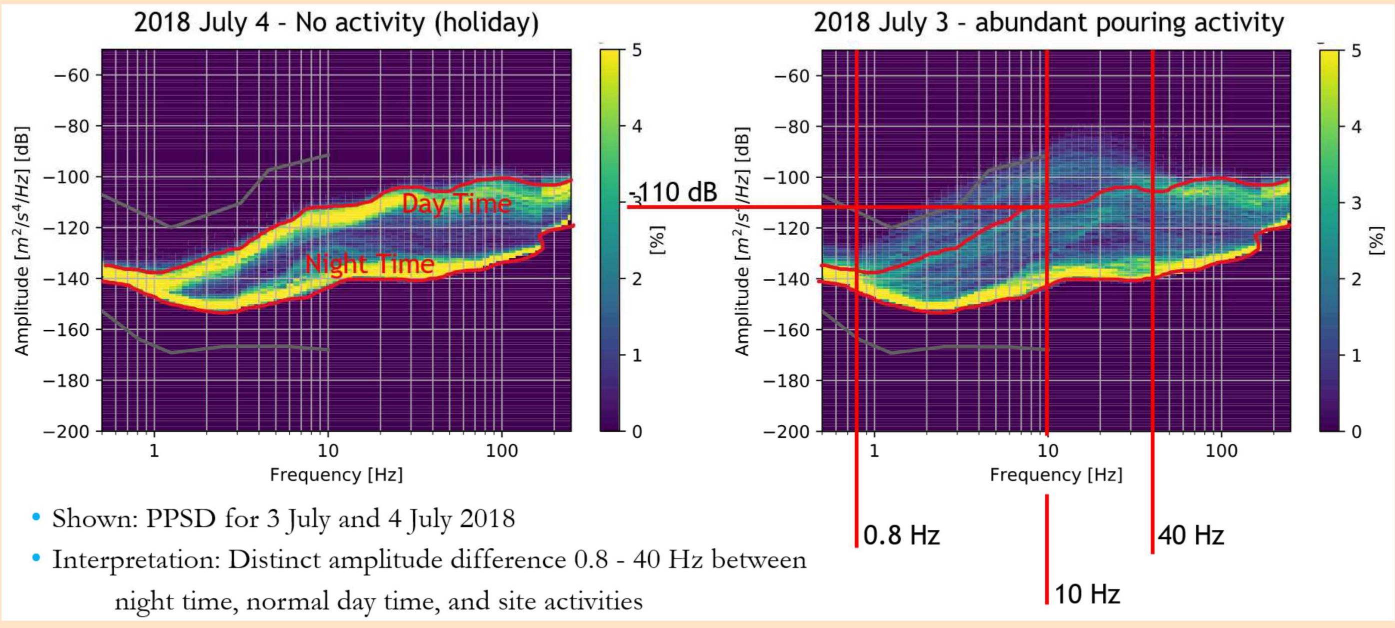
1. Raw seismic data

Real seismic sensor data recorded at 500 Hz on three channels (R, T, Z).



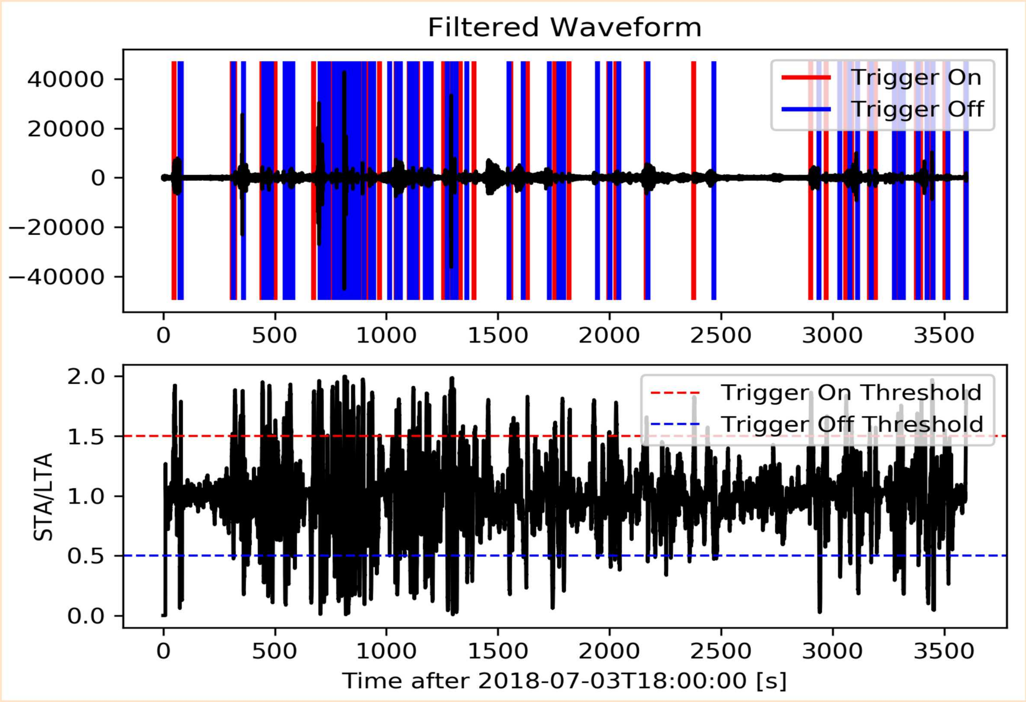
2. Probabilistic power spectral density

Define filter bands and determine periods of time where activity is high and distinguishable from noise.



3. Seismic event detection

STA/LTA to detect discrete non-traditional seismic events that occur within the full seismic waveform.



4. Method 1: Compression Based Clustering

0. Prediction by partial matching (PPM) with arithmetic coding (AC)

Statistical data compression techniques are based on Markov models of different contexts.

$$\Pr(x|c_1c_2\cdots c_n) = \frac{\Pr(c_1c_2\cdots c_n\ x)}{\Pr(c_1c_2\cdots c_n)}, \approx \frac{\mathrm{count}(c_1c_2\cdots c_n\,x)}{\mathrm{count}(c_1c_2\cdots c_n)}$$

1. Normalized Compression Distance (NCD)

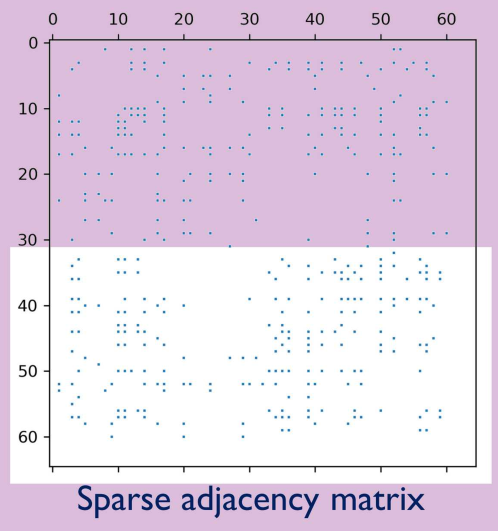
NCD between two seismic events estimates how much shared information they have.

Normalized Compression Distance:

$$NCD = \frac{C(x,y) - min\{C(x),C(y)\}}{max\{C(x),C(y)\}}$$

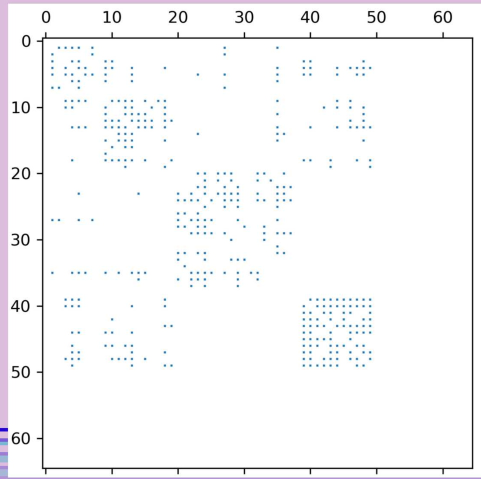
II. Weighted Adjacency Matrix and Louvain Community Detection

Construct adjacency matrix with events as nodes and NCD as edge weights, then apply Louvain community detection

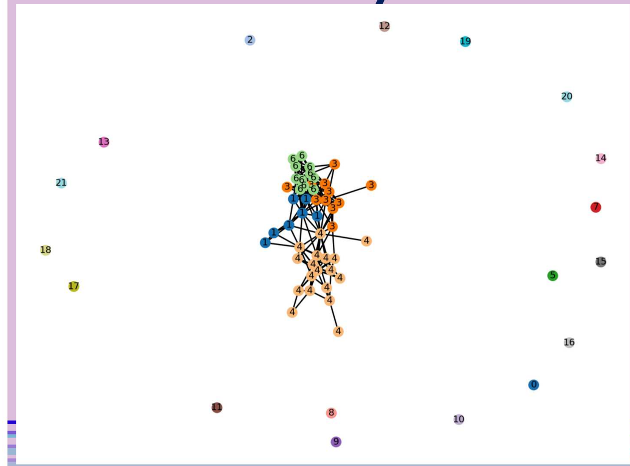


Sparse adjacency matrix

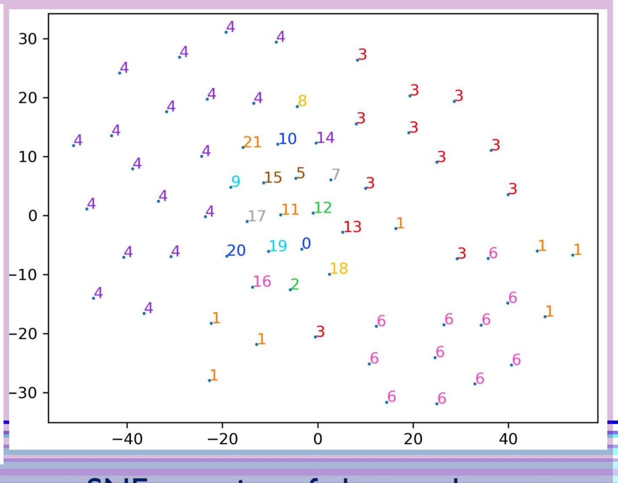
Preliminary Results



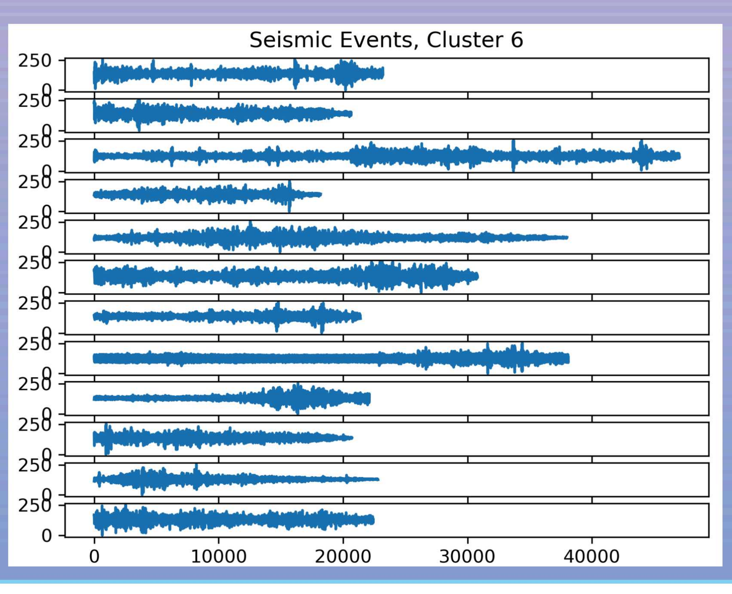
Sparse adjacency matrix reordered according to community assignment

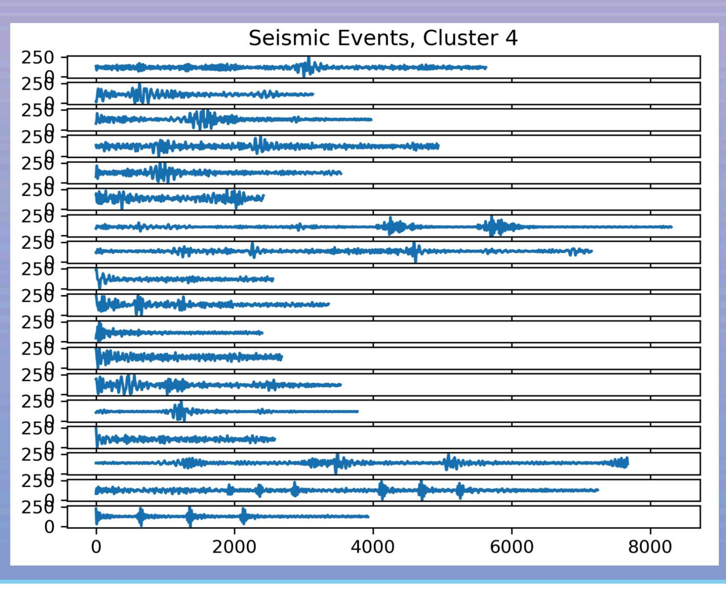


Seismic event community graph
Modularity = 0.41



t-SNE mapping of clustered events





4. Method 2: Feature Based Clustering

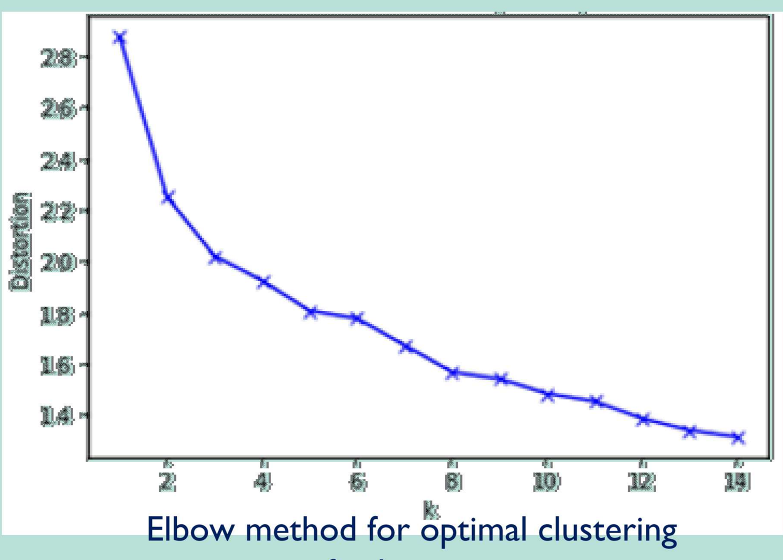
0. Feature Extraction

Top 10 Frequencies most present in Event
Power
Duration
Peak Amplitude

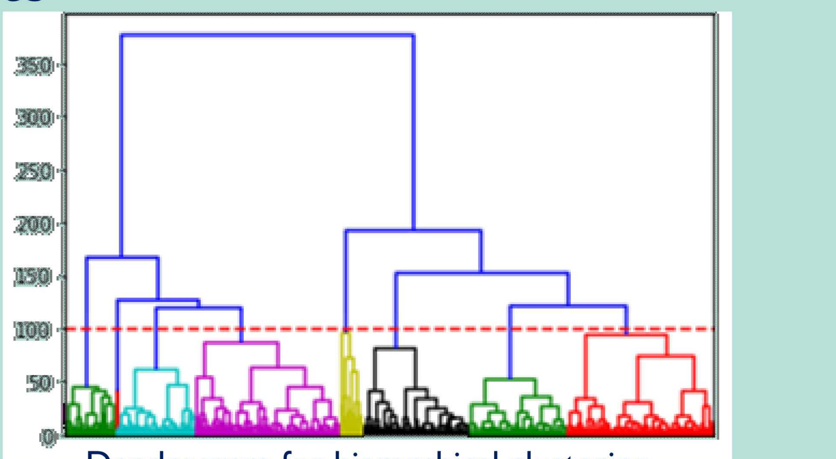
1. Dimensionality Reduction

Standardize features and use PCA to reduce number of features down to 5, which explains 74% of the variance in the features.

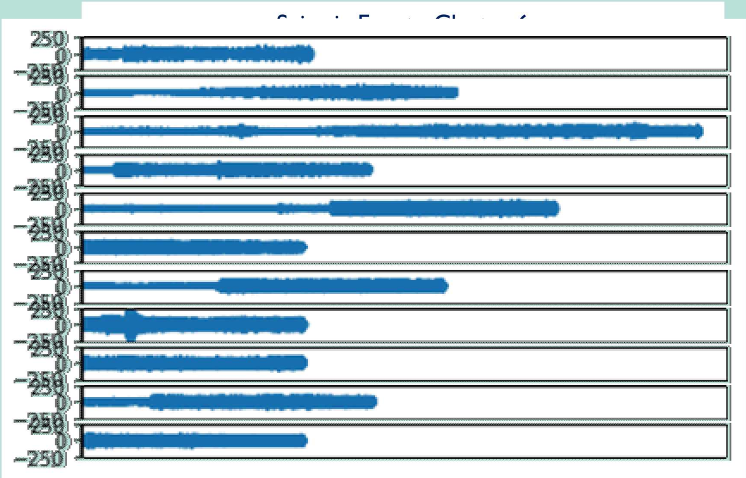
Preliminary Results

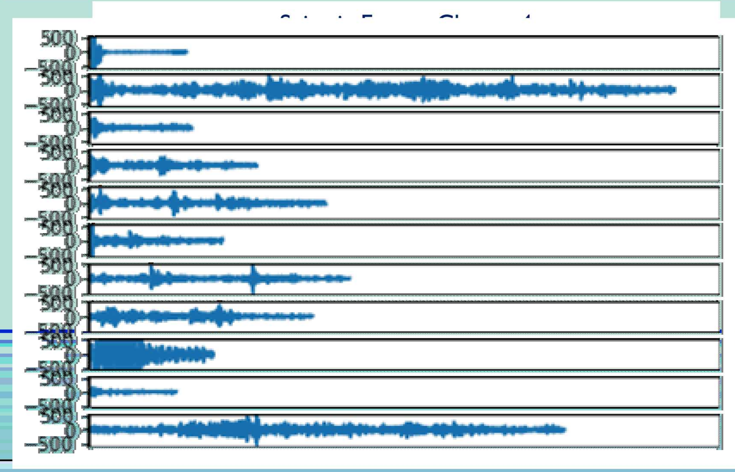


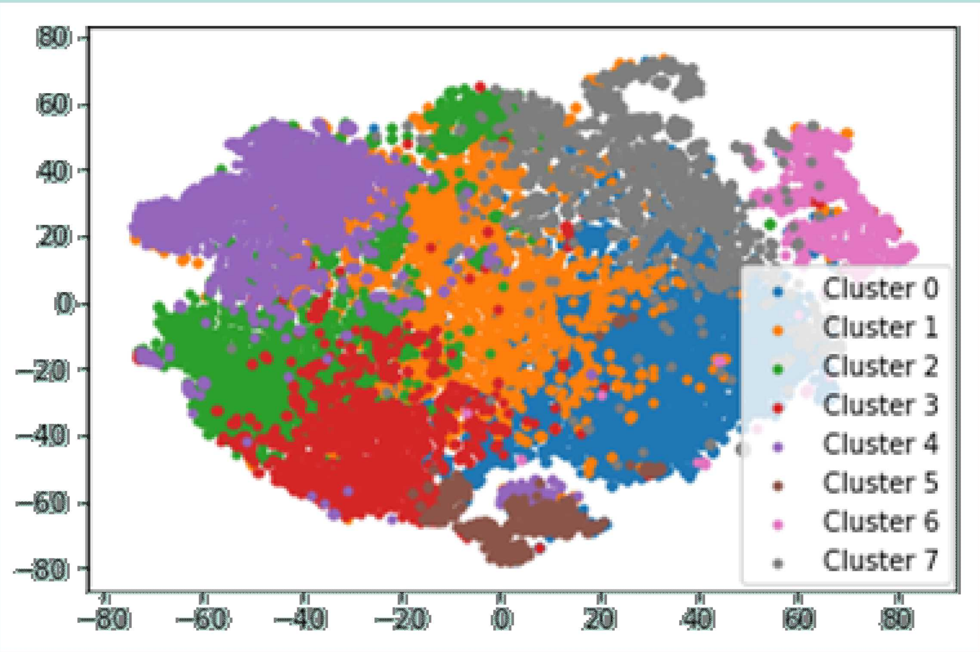
Elbow method for optimal clustering for k-means



Dendrogram for hierarchical clustering







t-SNE mapping of clustered events

Technical Challenges and Discussion

Event detection

- Need to define granularity of an event
- Events of interest are outside of traditional seismological targets
- Very poor supervision
- Tuning global parameters is challenging

Clustering

- Results depend on quality of detected events
- Very poor supervision
- Selected time period for analysis might not be sufficient for capturing similar events
- Results are hard to interpret

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.
SAND No. ____________

This work was funded by the U.S. Department of Energy National Nuclear Security Administration's Office of Defense Nuclear Nonproliferation Research & Development (NA-22).