SAND2021-12130PE

Towards Building an Automatic Array QC Tool from Signal Singular Values

C. A. Rowe, LANL

S. Heck, SNL

Los Alamos NATIONAL LABORATORY EST.1943

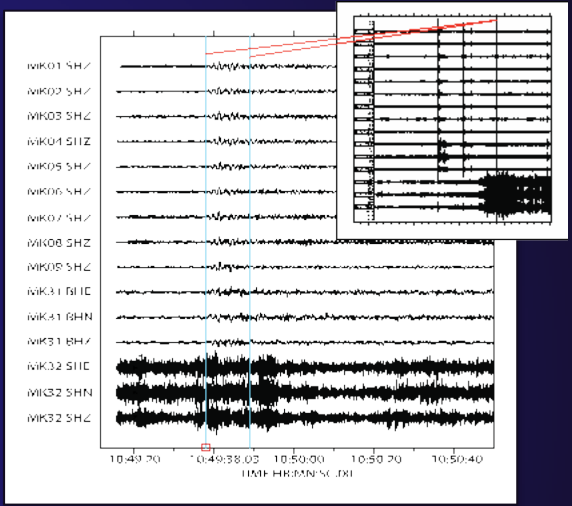
Virtual Expert Meeting on Advances in Waveform Processing and Special Studies

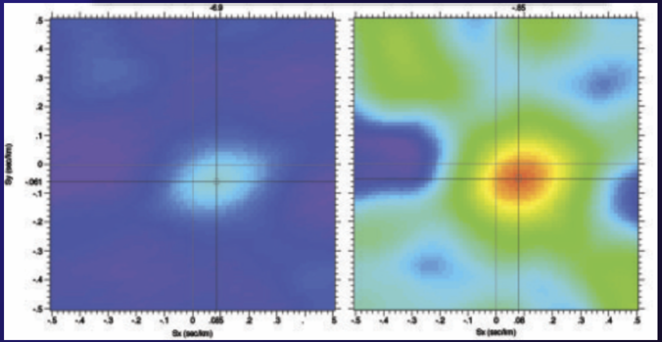
18-22 October, 2021

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.

The Problem

Array processing is a powerful tool for enhancing small seismic signals of interest, but the quality of the analysis can depend strongly upon SNR, sometimes for an individual channel. Persistently problematic channels can be eliminated a priori, but some failures or noise sources are ephemeral, and an automated array process may be compromised if the offending channel is not identified and removes.





- Top: MKAR recording of small Wenchuan aftershock.
- Bottom: FK panels for the indicated time window. Near panel: including noisy channels. Far panel: without noisy channels.

A Cybersecurity Inspiration

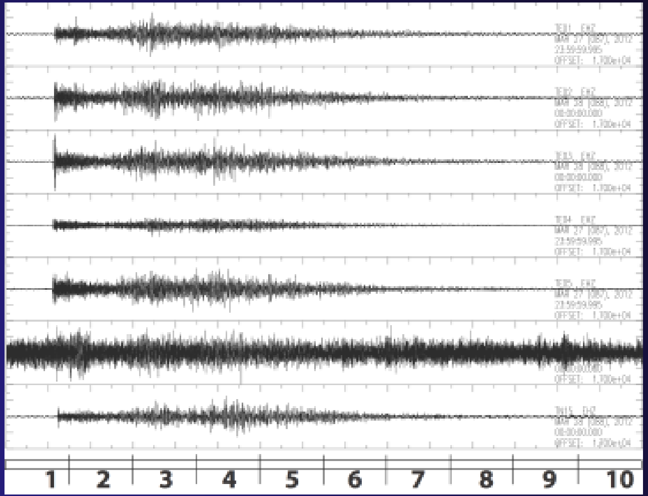
- Changes in data dimension have become key in the world of cybersecurity and monitoring of node traffic on large servers. The underlying approach relies upon the treatment of N nodes as a time-varying matrix of N vectors, each corresponding to the data passing through a node.
- Routine activity across the network of nodes is characterized by a level of independent quasi-random behavior for traffic through each node, but examination of the actual data is not necessary.
- A principal components analysis (PCA), singular value decomposition (SVD) or eigenvalue analysis of the traffic matrix will yield a fairly consistent suite of singular values or eigenvalues and a fairly stable dimension of the system.
- Sudden changes in activity on a single node or collection of nodes will perturb the matrix system and may contribute to a change in the associated eigensystem or PCA.

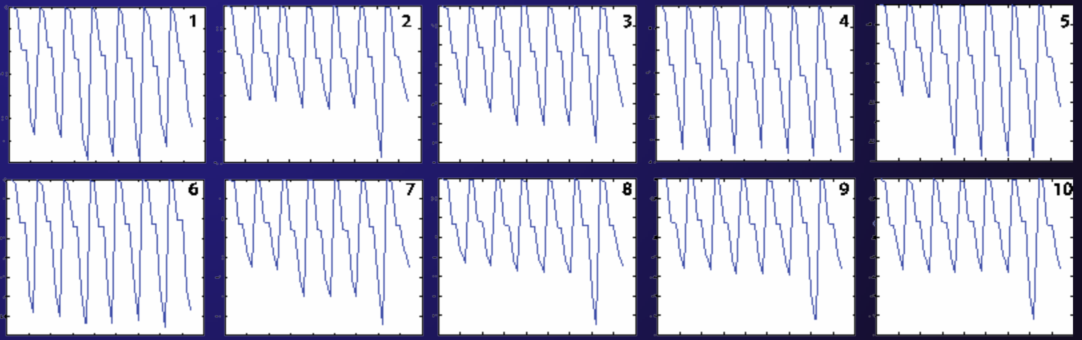
Adapting the Subspace Detector

- We modify the template-based detector approach by considering a time-varying system of basis vectors for a seismic array.
- At any given time, waveform data for the full array can be described by a system of eigenvectors and eigenvalues or singular vectors and associated singular values. "Typical" background noise will be manifested as a relatively stable eigensystem over time.
- This can, however, be perturbed by ground motion (or electronic) disturbances that affect the entire array or only a portion of the array.
- Because synchronous independent failures are a rarity, we focus here on cases where a single array component is impacted by seismic or electronic events.

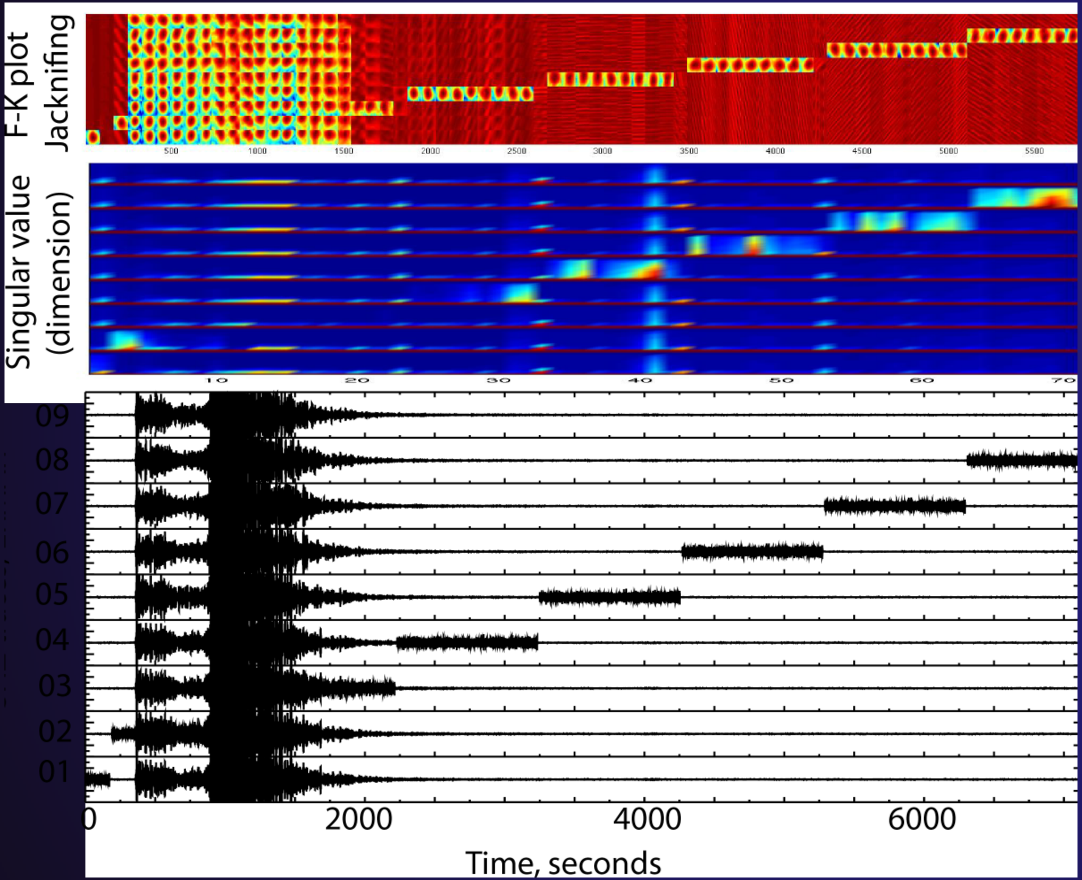
Jackknifing

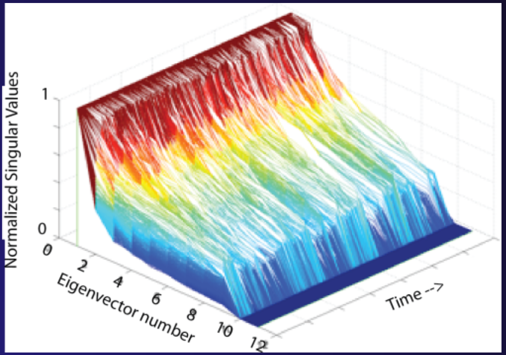
The question is how to leverage the eigensystem to identify the corrupted channel. We employ a jackknifing process, in which each station is removed and an N-1 system is evaluated for its singular values. Removal of the uncontaminated traces produces a suite of singular values in N-1 cases that does not vary significantly for any of the jackknife steps. Removal of the contaminated station, however, produces a significant change to the eigensystem (e.g., Rowe et al, 2014).





An Example IMS Array

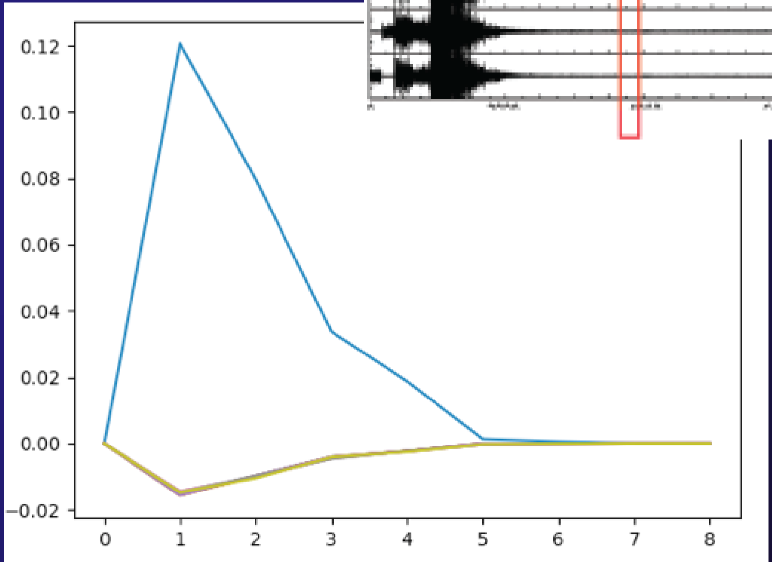


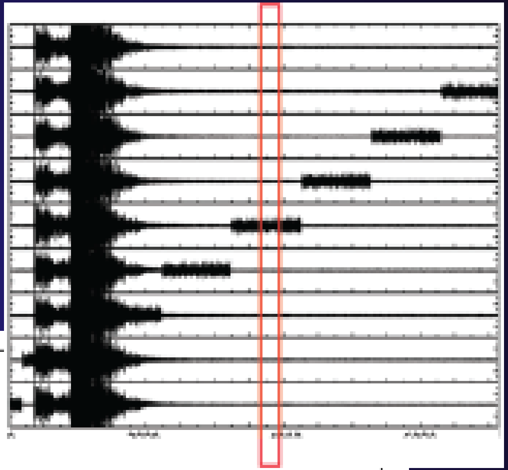


Above: normalized singular value functions progressing as a time window moves through traces. Left: Moving through array in a jackknife method, dropping one trace at a time. SVD changes track the calibration pulse well. Running FK windows likewise exhibit improved integrity.

Quantifying

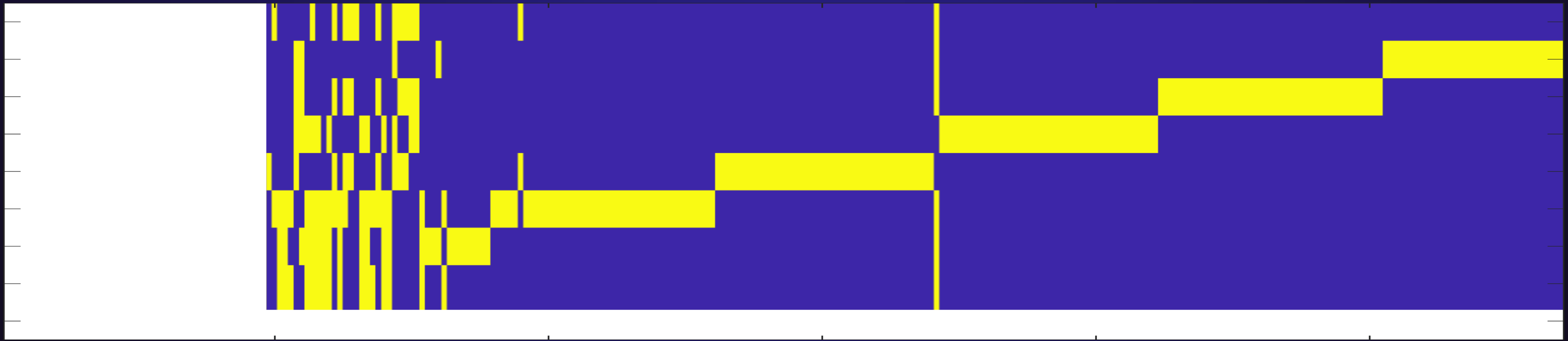
We quantify by comparing the normalized SV functions for all N jackknife realizations of a time step. We determine a median function at a time step. Example at right, in which nine normalized SVD functions have each been subtracted from the median function and their differences are shown for a time window during a calibration sweep when the fifth channel is calibrating. Singular values are clearly perturbed when the fifth channel is eliminated from the eigensystem.

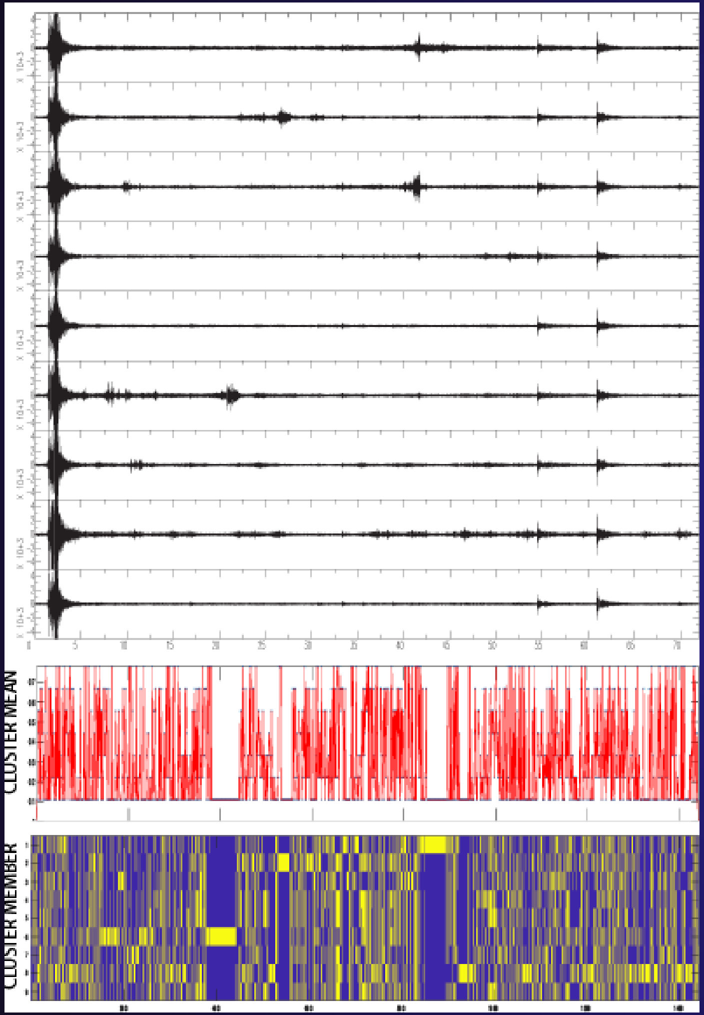




Isolating the Bad Channel

- We apply an agglomerative pair-group clustering algorithm to separate the median distance functions into two populations at each time step, using mean Euclidean separation to assign membership.
- Here we see outlier cluster membership accurately assigned to the calibrating channel, except when the large earthquake randomizes the eigensystem variations.
- Although this assumption ignores the possibility of two bad channels, we plan to address this possibility in later versions of the application.

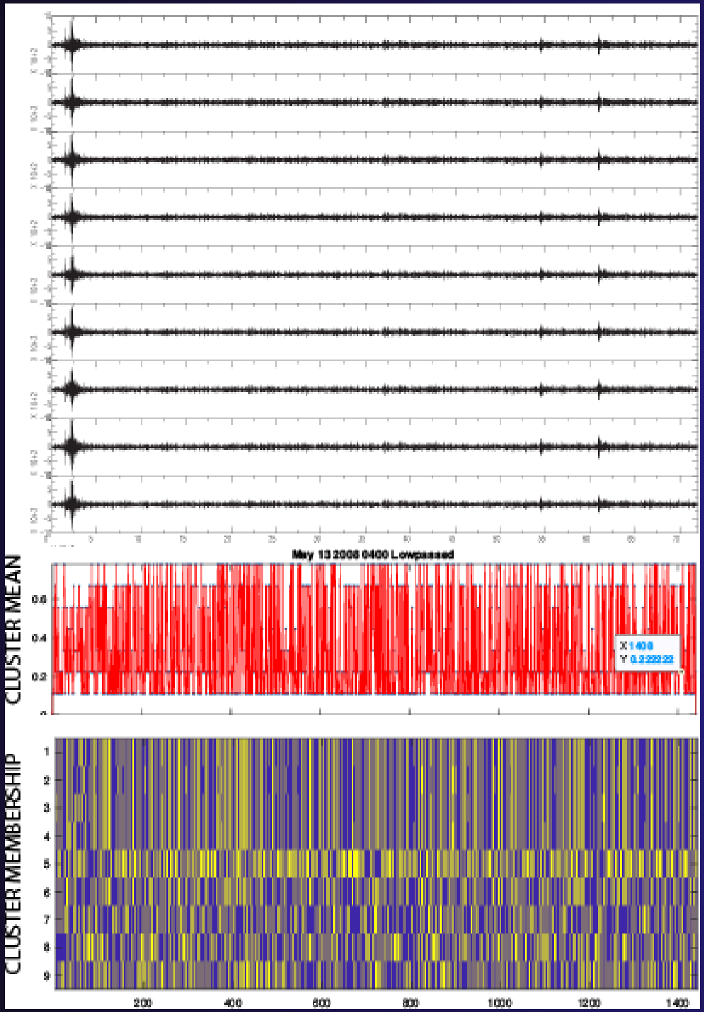




Another Example

Top: Two-hour time segment of MKAR SHZ channels.
Middle: Running average of cluster summation for a two-hour segment of array MKAR.
Bottom: associated cluster membership showing which channels have contributed to extended time periods of anomalous behavior. In this example, Ch 6 has intermittent issues in the first third of the example, whereas Ch 1 is problematic for a brief time about 2/3 through the scan. Ch 8 exhibits frequent noise issues, which is true over many days at this time (May 2008).

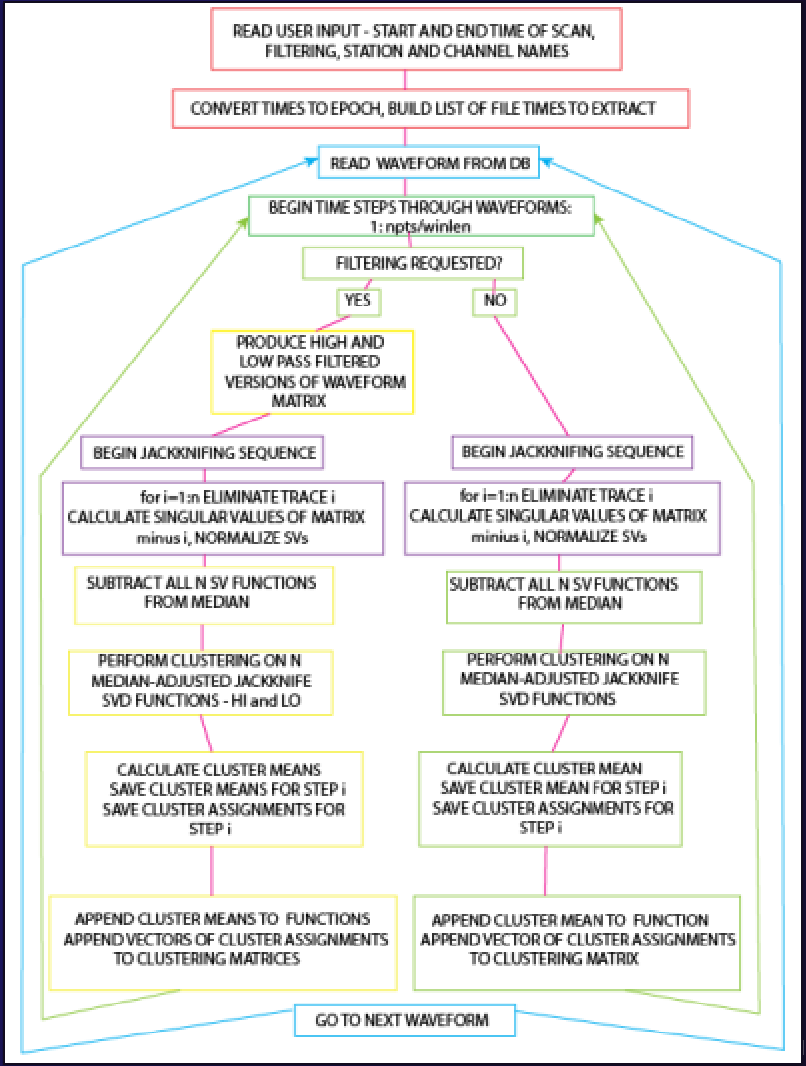
Frequency Dependence



Same data as shown in the previous slide, but the traces have been low-pass filtered at 0.5 Hz. Note random, intermittent noise does not appear to be manifested on the channels in general and very few time periods exhibit a single channel isolated for multiple time measurements into its own cluster. At low frequencies, Ch 5 emerges as the primary (but minor) offender, whereas in the unfiltered example it seldom perturbed the eigensystem.

Code Details

- Code written in Python within Anaconda, leveraging Pisces (MacCarthy and Rowe, 2014) database interface
- Command-line driven: arguments include start and stop time of analysis, stations and channels to include.
- User-defined window length, filter corners (or none) and time skip.
- We plan to build a GUI to reduce typing (and typos) with the option for real-time graphical feedback.



Issues and Additional Thoughts

- One known issue is the lack of a clustering threshold, such that minor noise issues do not produce cluster definitions that suggest a problem when the presence of noise is not enough to meaningfully compromise a study. We will explore additional clustering tools and metrics toward this end.
- Another issue is the possibility of more than one problem channel. We are exploring a nested jackknifing approach.
- Other potential useful applications of this method is data QC for very large experiments, such as active source multi-fold seismic projects in which bad channels can degrade analysis.