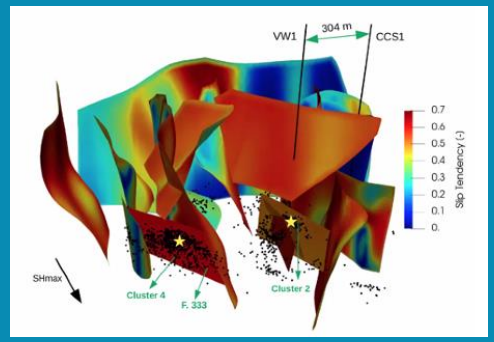
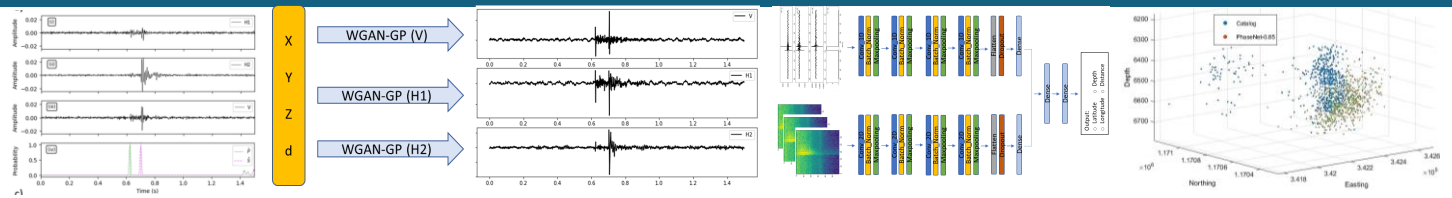
SAND2025-01144C

An Integrated ML Approach for Analyzing Micro-Seismicity during Geologic Carbon Storage at the Illinois-Basin Decatur Project site





Hongkyu Yoon

Daniel Lizama, Rachel Willis, and Jennifer L. Harding

Sandia National Laboratories

ML in Solid Earth Geoscience 2025

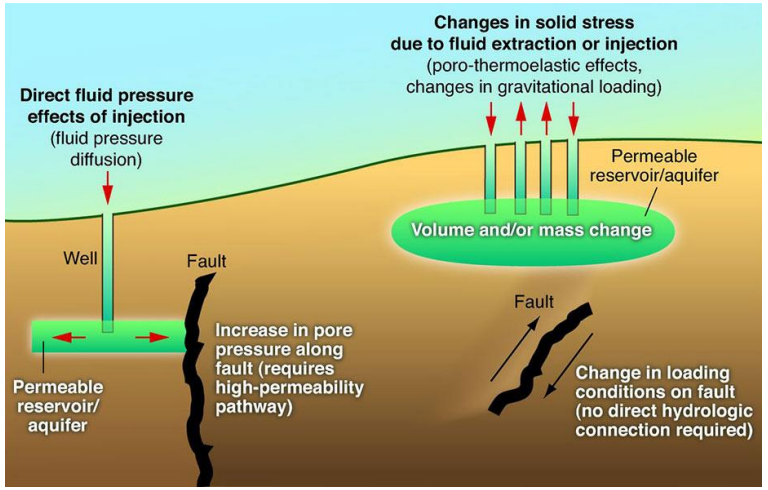
This work was supported by DOE Office of Fossil Energy and Carbon Management project -Science-informed Machine Learning to Accelerate Real Time (SMART) Decisions in Subsurface Applications-Carbon Storage.

Motivations

♦ Motivations

- Fluid injection or withdrawal causes changes in pore pressure and local stresses, resulting in induced seismicity (IS) during subsurface energy activities
- Machine learning (ML) has been successfully developed and applied for data analysis of (micro-)seismic data (e.g., event detection, phase arrival time, source locations)

Induced seismicity



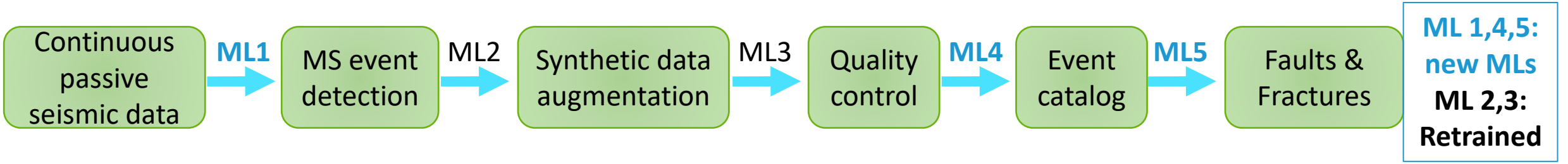
USGS: http://earthquake.usgs.gov/Research/induced/modeling.php

♦ Goals

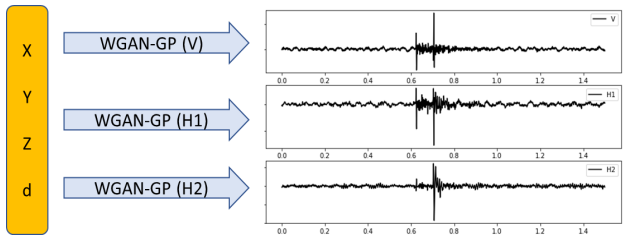
(1) Develop/apply machine-learning models for seismic wave data analysis and source location estimation at Illinois Basin Decatur Project (IBDP) site

(2) Delineate slip mechanisms associated with microseismic (MS) data

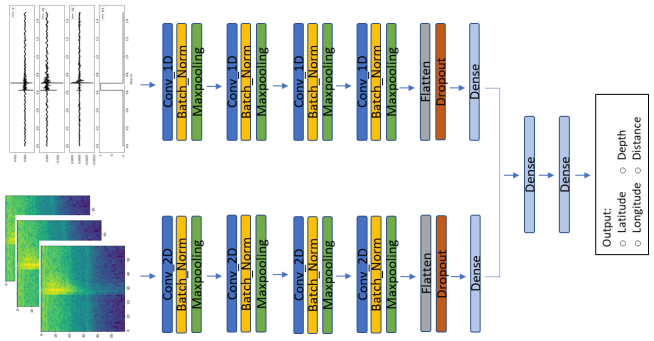
Overview of integrated ML methods



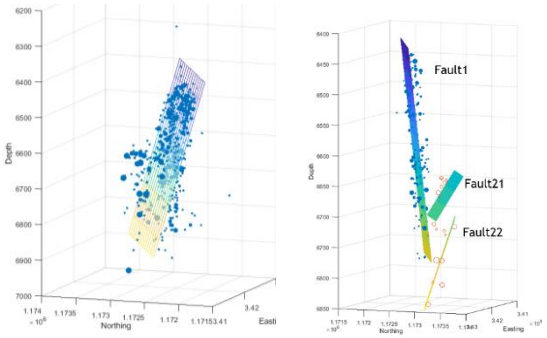
- **ML1**: Developed a new convolutional neural network (CNN) model to detect seismic-event signals out of raw continuous passive seismic waveform
- ML2: Generate **synthetic waveform data** using generative adversarial networks (WGAN-NP) to estimate source location of newly detected events accurately due to a small number of events in the catalog
- ML3: Down-select synthetic data with high quality of waveform using PhaseNet phase picks
- **ML4**: Developed a new **multi-modal** CNN model to locate seismic events with both original catalog and synthetic waveform data
- **ML5**: Used a unsupervised clustering and a least-square algorithm to infer fault planes from the seismic event catalog



ML2: Waveform generative model



ML4: Source location model



ML5: Clustering-based fault plane construction

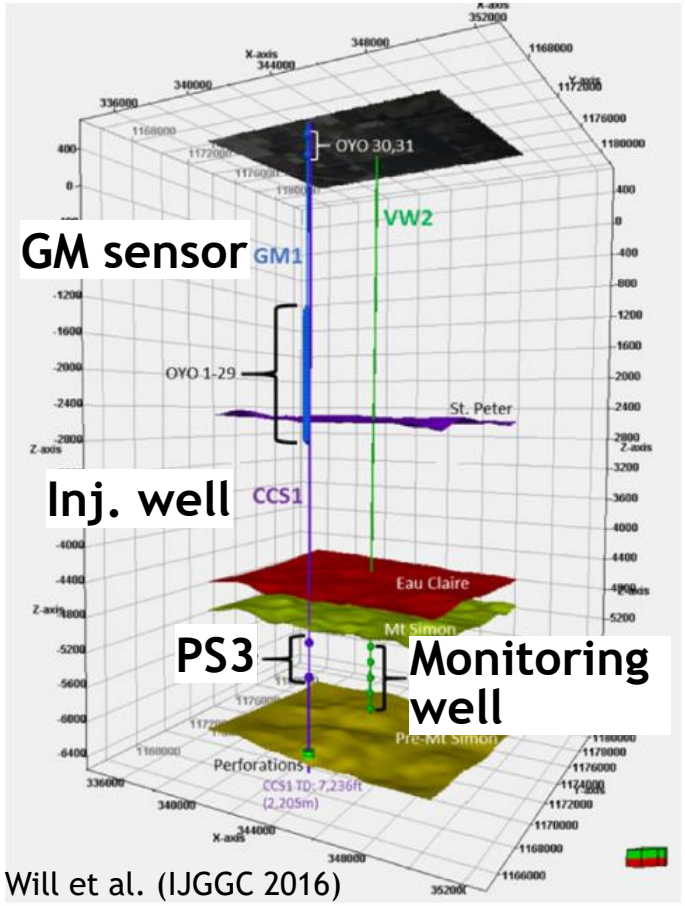
MS Waveform Data at the IBDP Site

Raw (unprocessed) continuous data

- Big data (~70 TB for 3+ yrs)
- 2 kHz sampling rate
- # of traces: 84-94 (inconsistency at an early injection period)
- Only vertically oriented sensors at an early phase

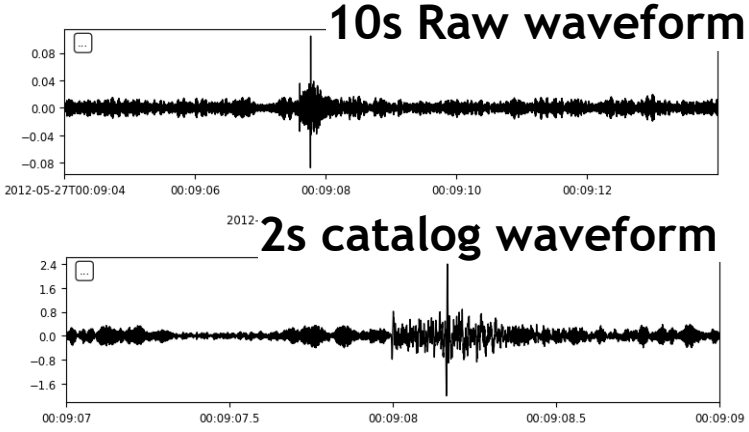
Processed data & catalog (~3 yrs injection)

- Detected event (processed 2s window, ~ 19K events, 3 channels (Z,H1,H2)
- A small # of located events (~ 5K events with source locations)
- Relatively low magnitude (mostly <0, max magnitude = ~1.5)
- **Processed 2s window data have been shifted from original data** (needed to generate event data for machine learning separately)

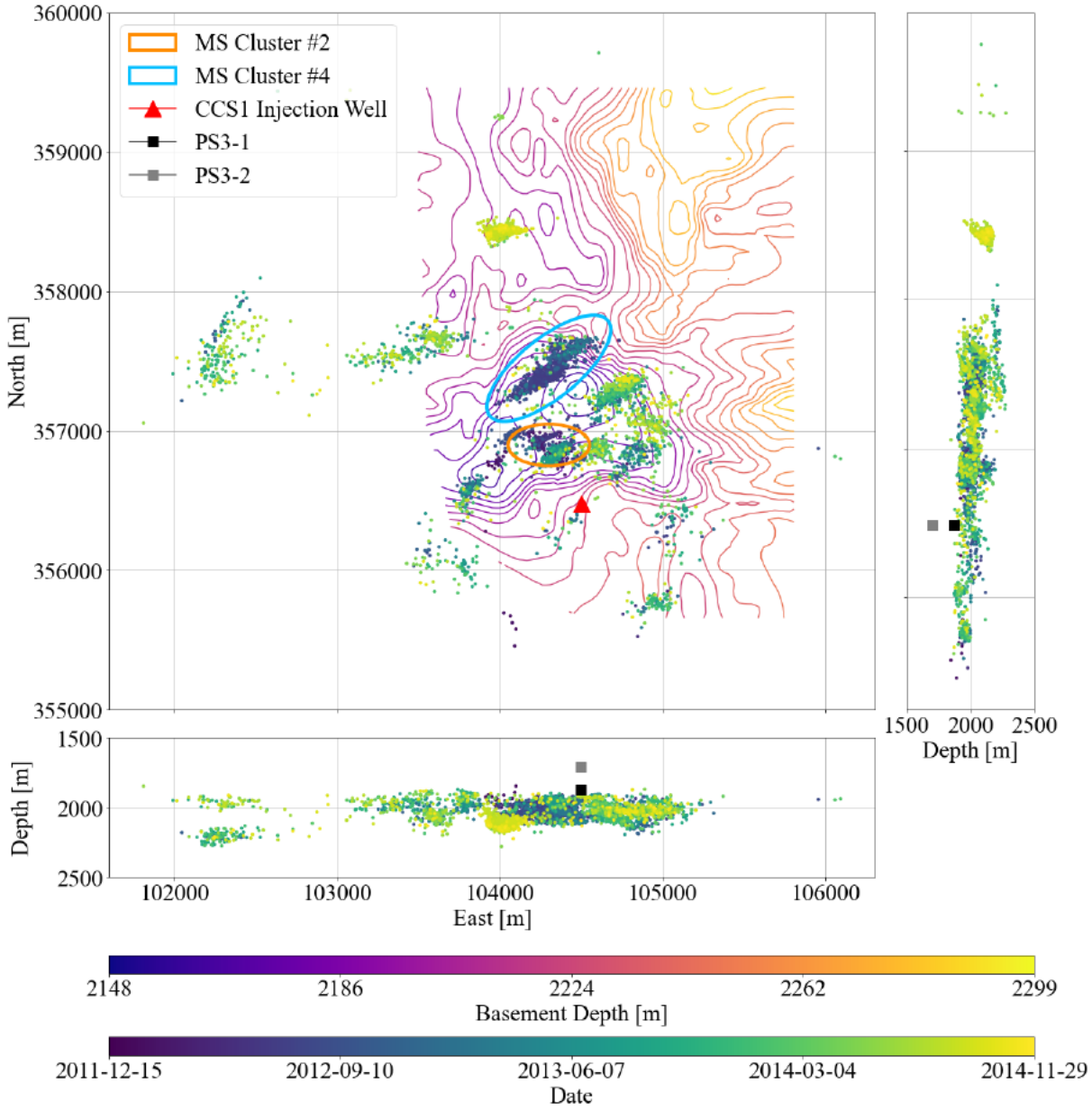


Will et al. (IJGGC 2016)

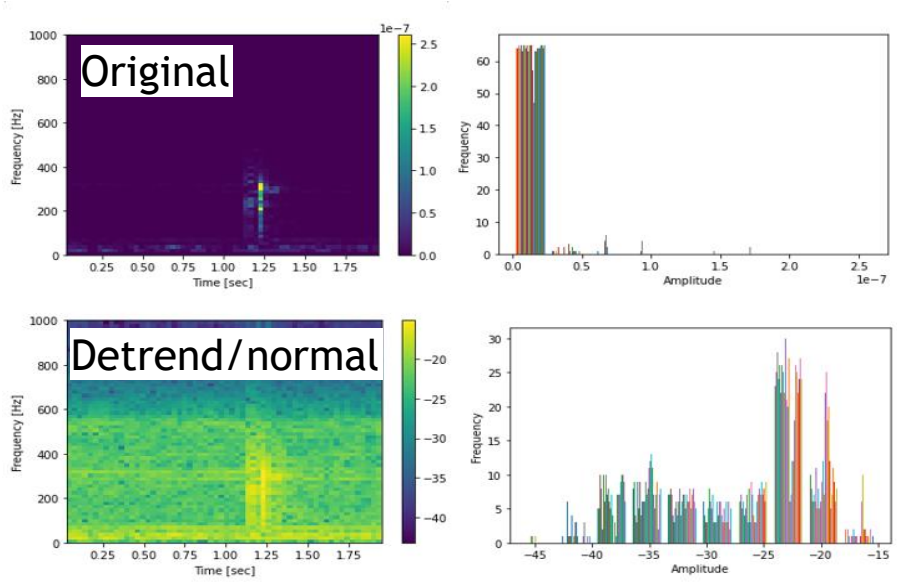
Fig. 1. Subsurface array configuration. Distance units are feet, *Z* axis is referenced to mean sea level.

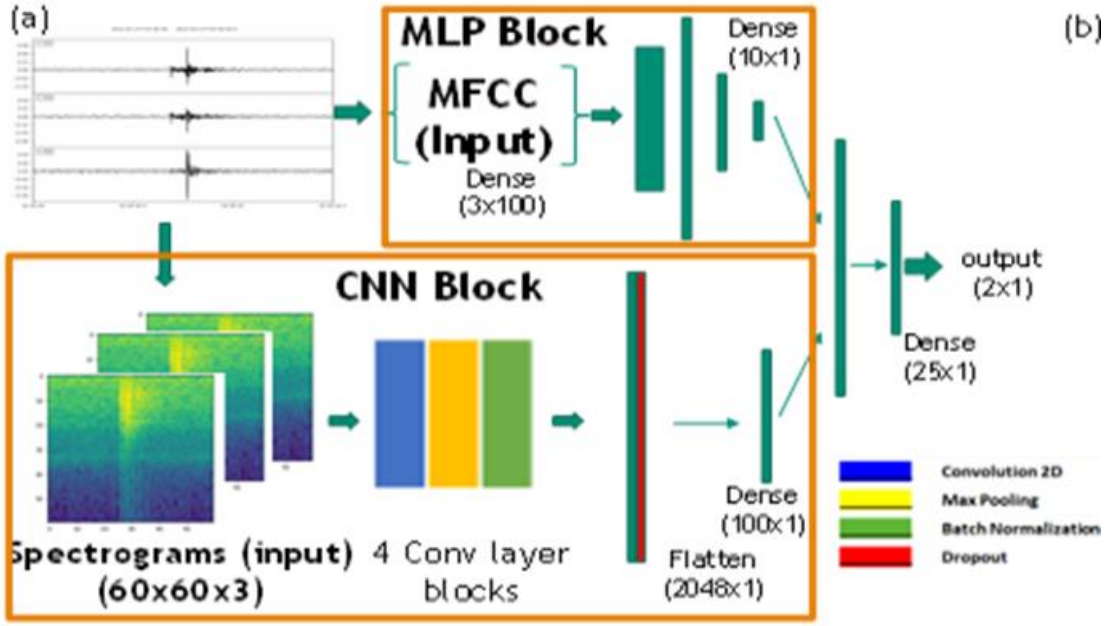


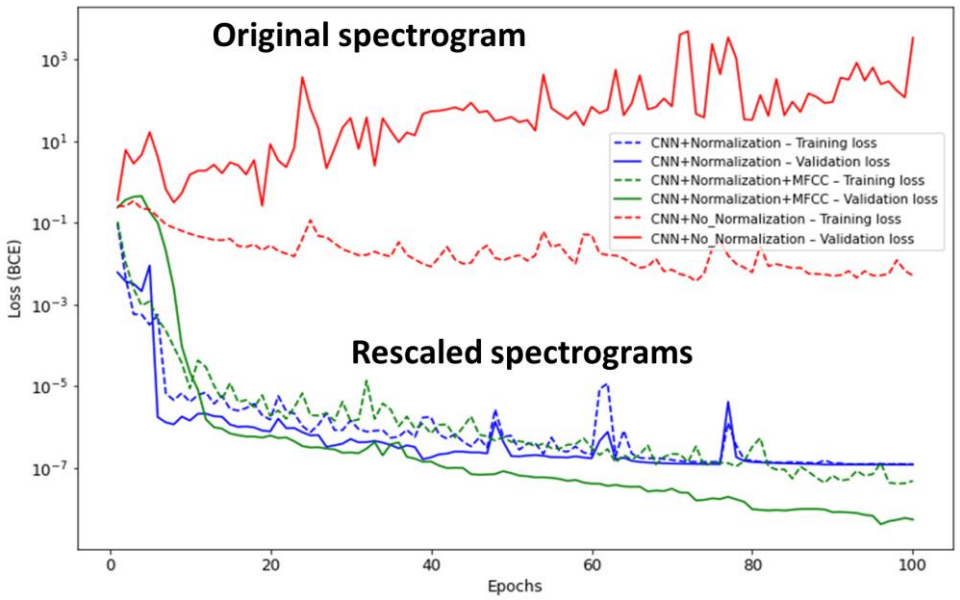
Map of induced seismicity at IBDP (catalog)

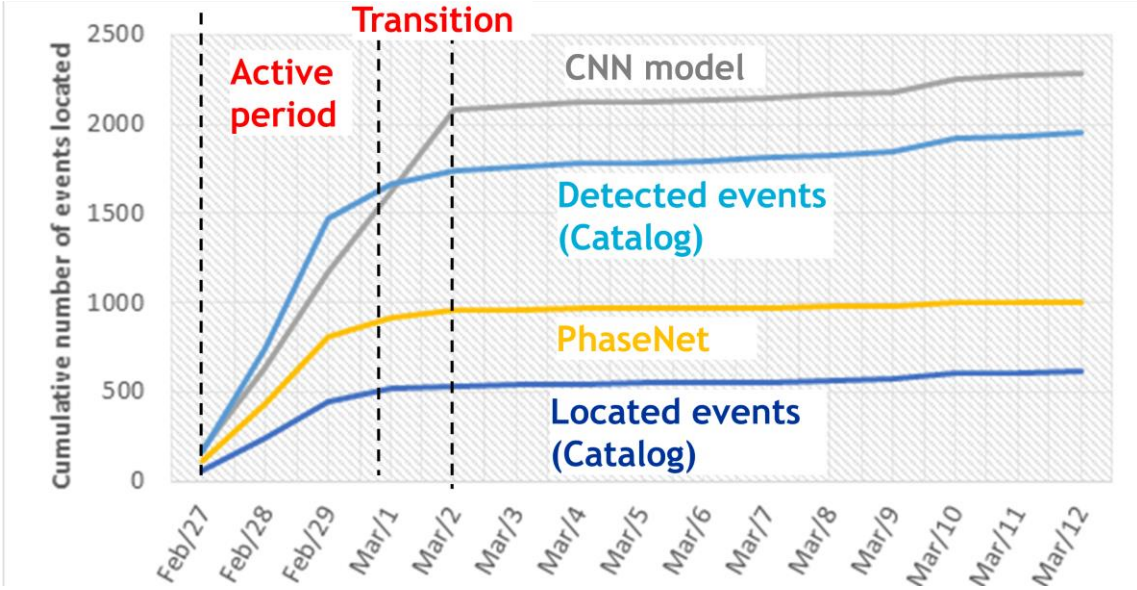


CNN-based event detection



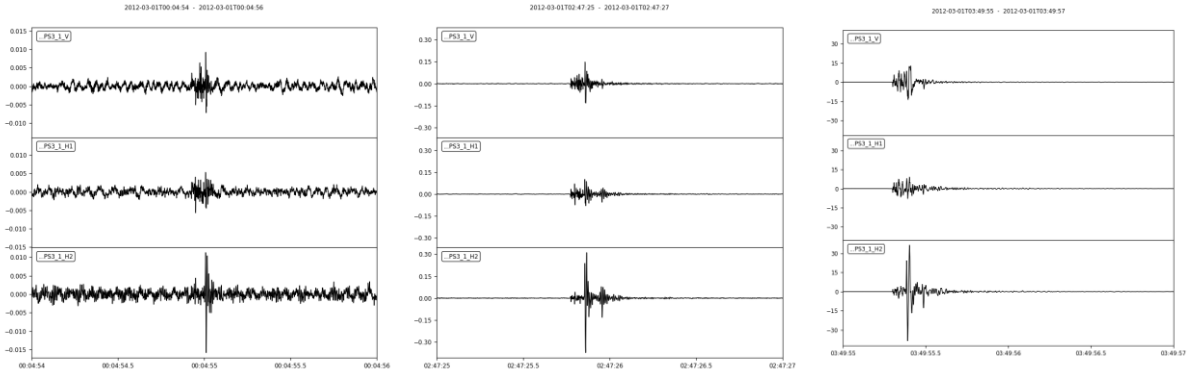




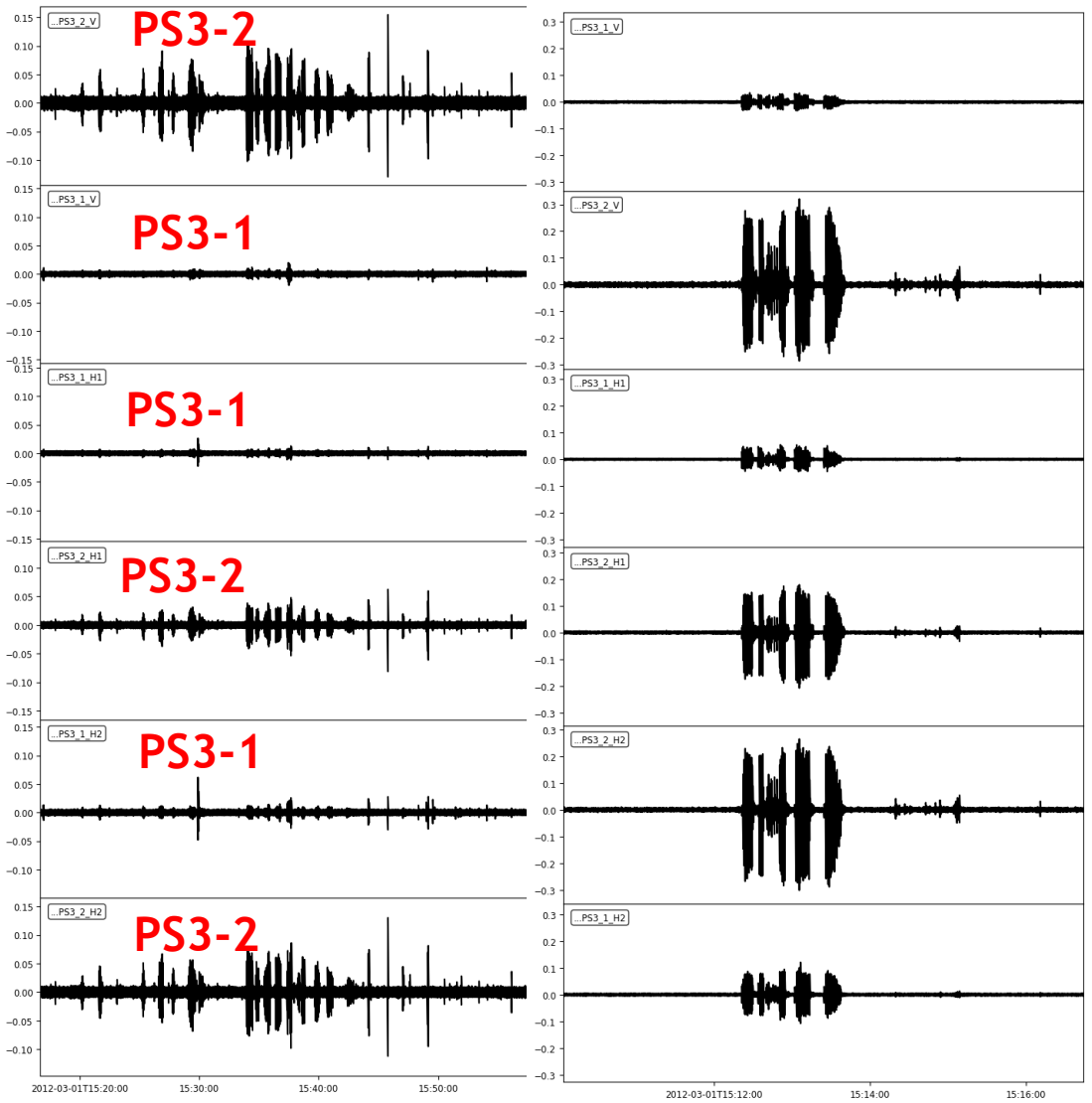


Waveform characteristics: Active vs. transition periods (March 01-02)

Active period (Feb27-29): 2s window events



Transition period

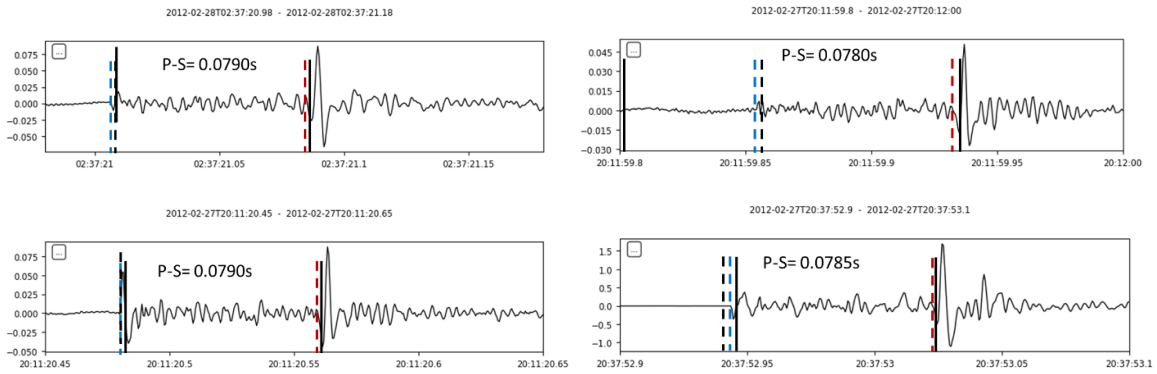


Long-period long-duration (LPLD) seismic events

- Represent slow shear slip (e.g., hydraulic fracturing)
- Observed in the literature (e.g., Das and Zoback, 2013) where natural fracture density is high, likely caused by high pore pressure and/or high clay contents (i.e., low permeability) => slow slipping
- Tend to be observed “only on faults large enough to produce a sequence of slow slip events”
- This observation needs to be used to parameterize the thickness of fault zone in inverse modeling

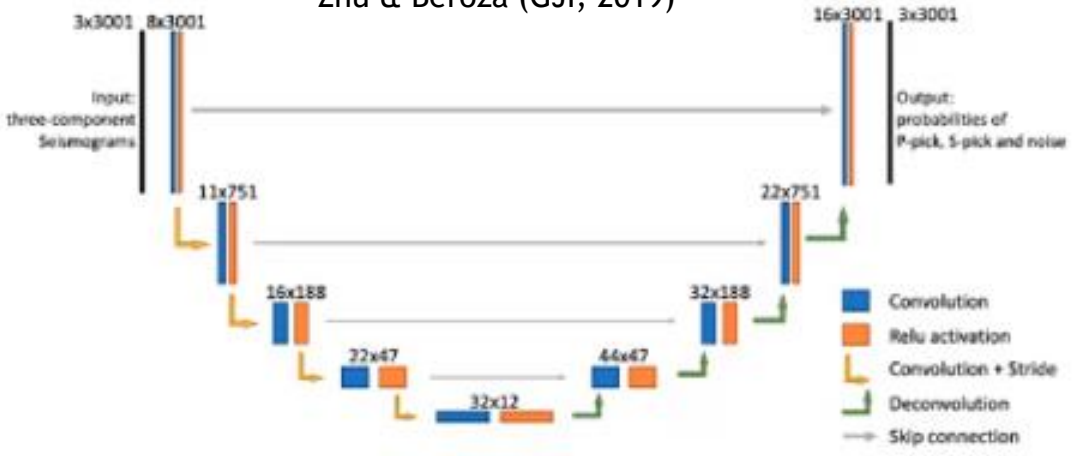
Arrival time estimation (retraining of PhaseNet)

Manual pick vs. automatic picker
(AR-picker (S-arrival), AICD (P-arrival) for training data)



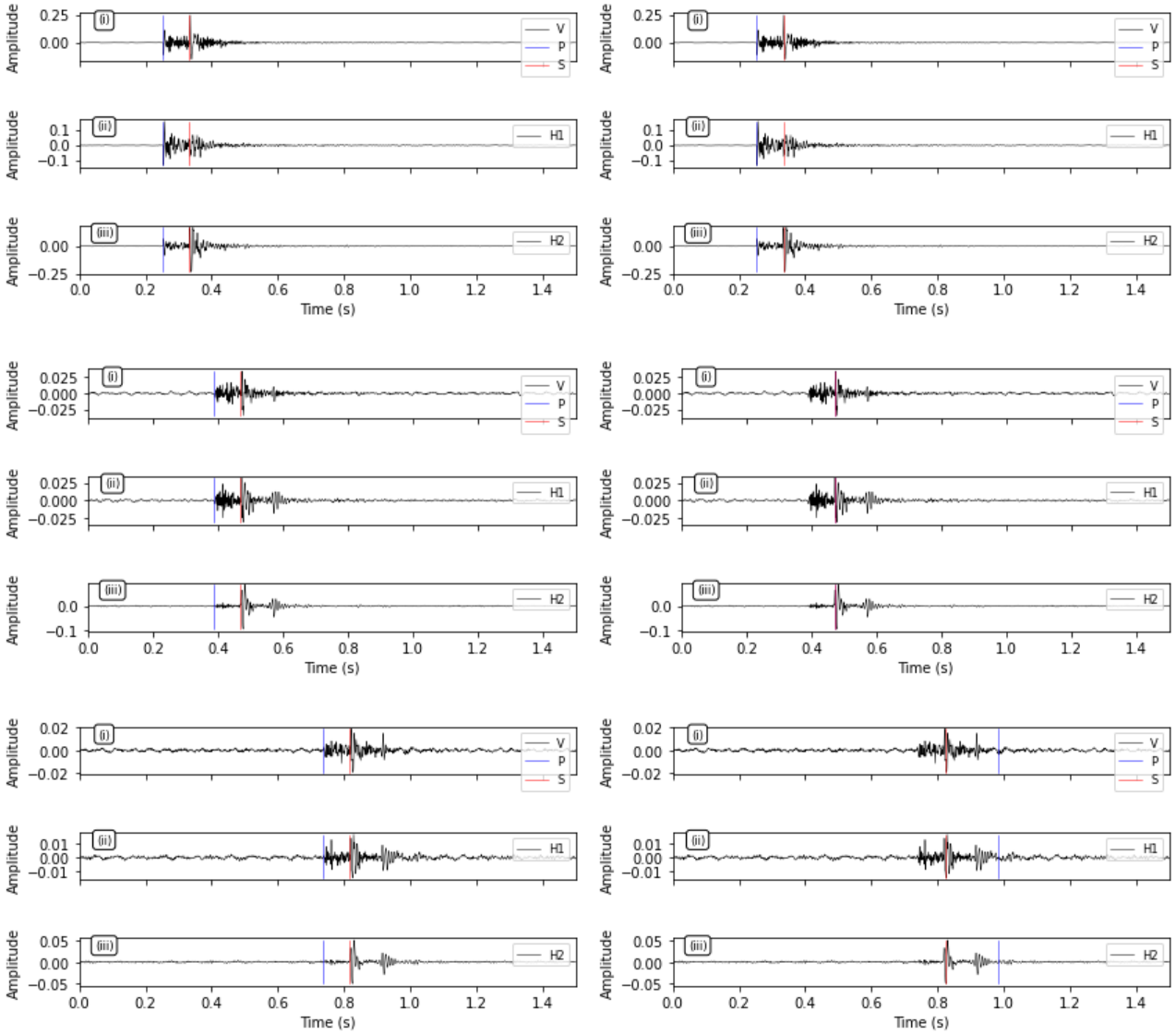
Phase Picking (PhaseNet)

Zhu & Beroza (GJI, 2019)



PhaseNet Picking

AICD &AR-picker



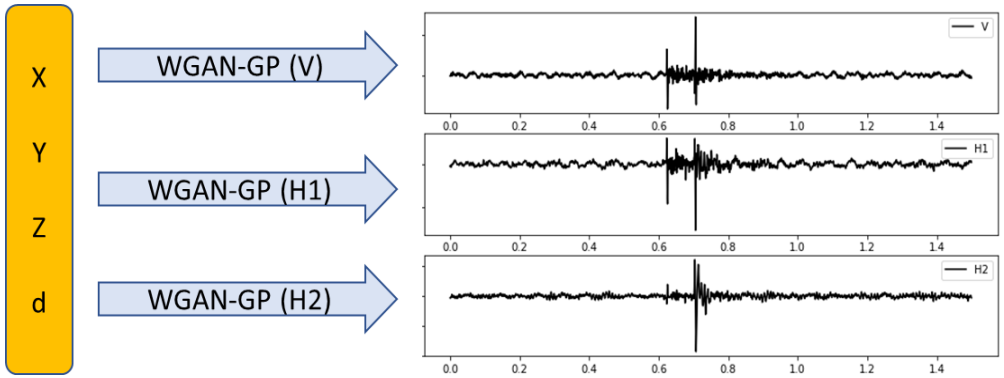
Lizama & Yoon (In Prep)

Source location estimate with data augmentation

SeismoML (WGAN) for synthetic waveform data generation → PhaseNet for arrival times → Multi modal CNN for source locations → Source location estimate of newly detected events

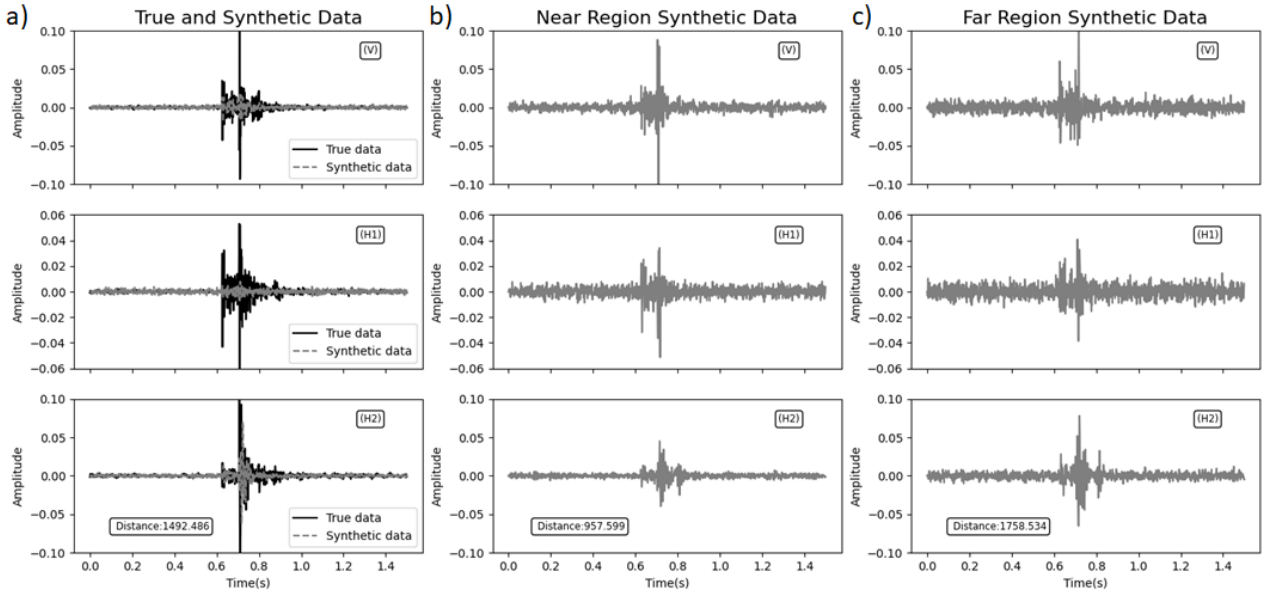
Data augmentation: To overcome a lack of event data for machine learning training:

- WGAN is employed to generate synthetic waveform data for a range of source locations (x,y,z) and distance to the geophone location.
- Waveform modulations and phase arrivals show a good quality of synthetic waveforms.



Seismogram generative model using multiple WGAN-GP.

SeismoML: A. Spurio Mancini et al. (Solid Earth, 2021)



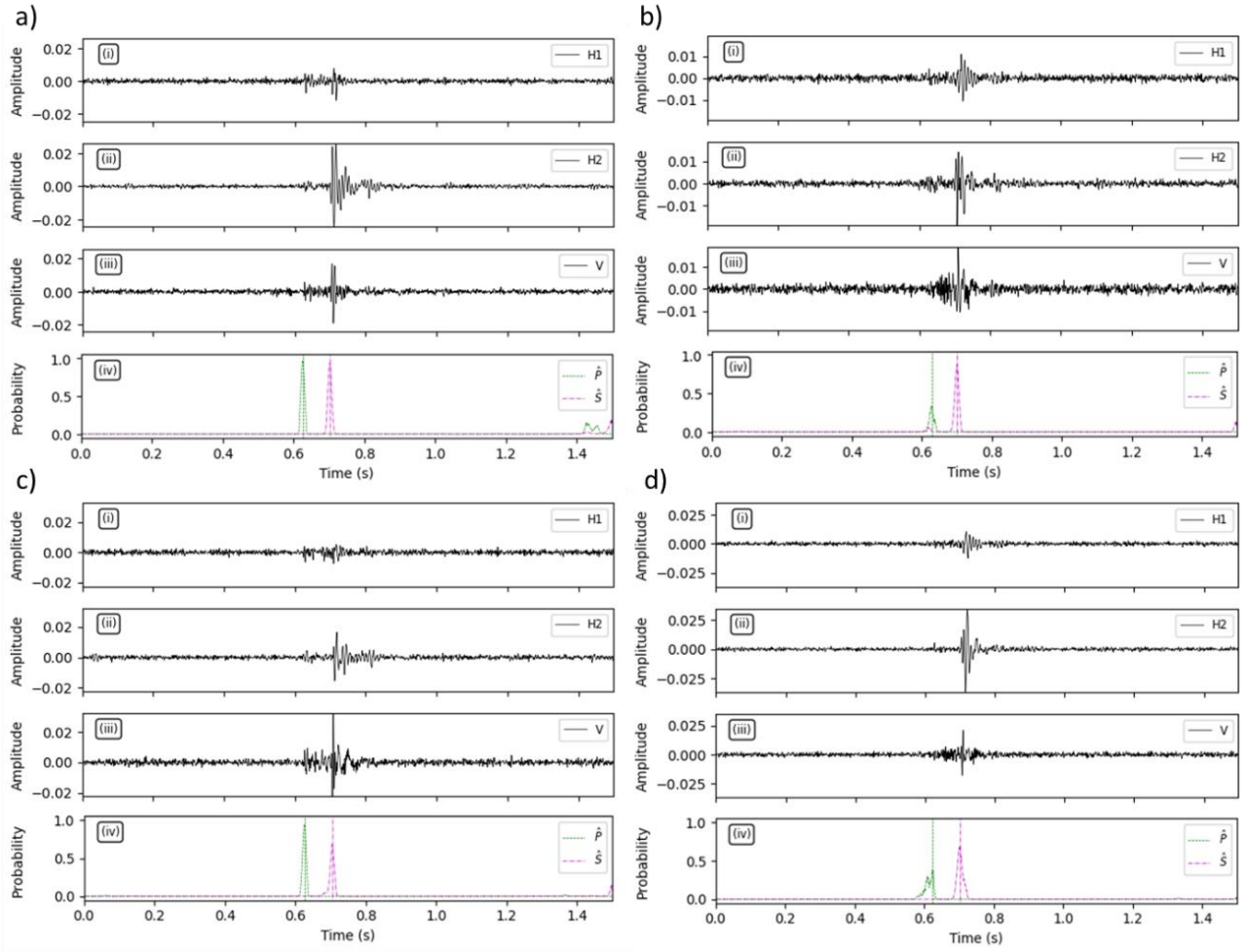
Real vs. synthetic event waveforms

Source location estimate with data augmentation

SeismoML (WGAN) for synthetic waveform data generation → **PhaseNet for arrival times** → Multi modal CNN for source locations → Source location estimate of newly detected events

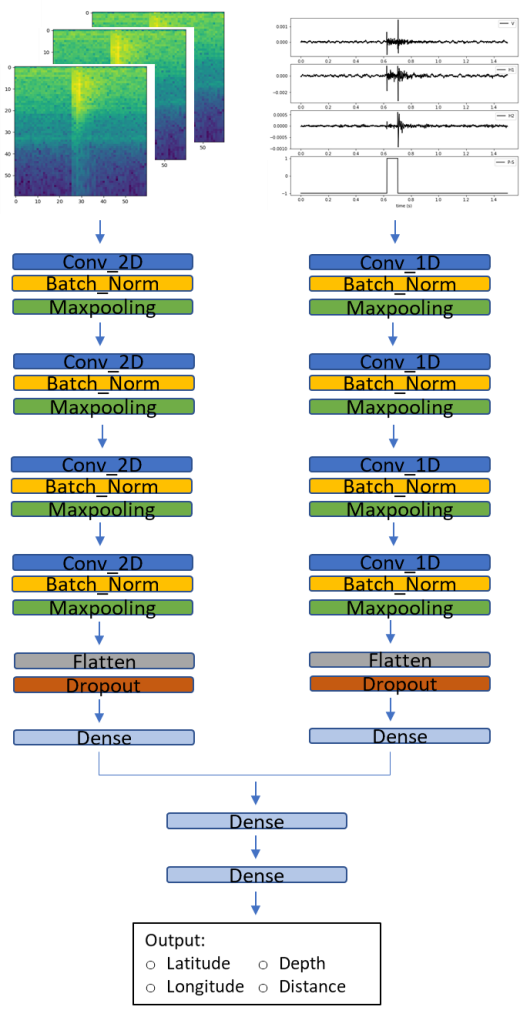
Phase Picking: To obtain phase arrival of synthetic data and evaluate sample quality for source location training data:

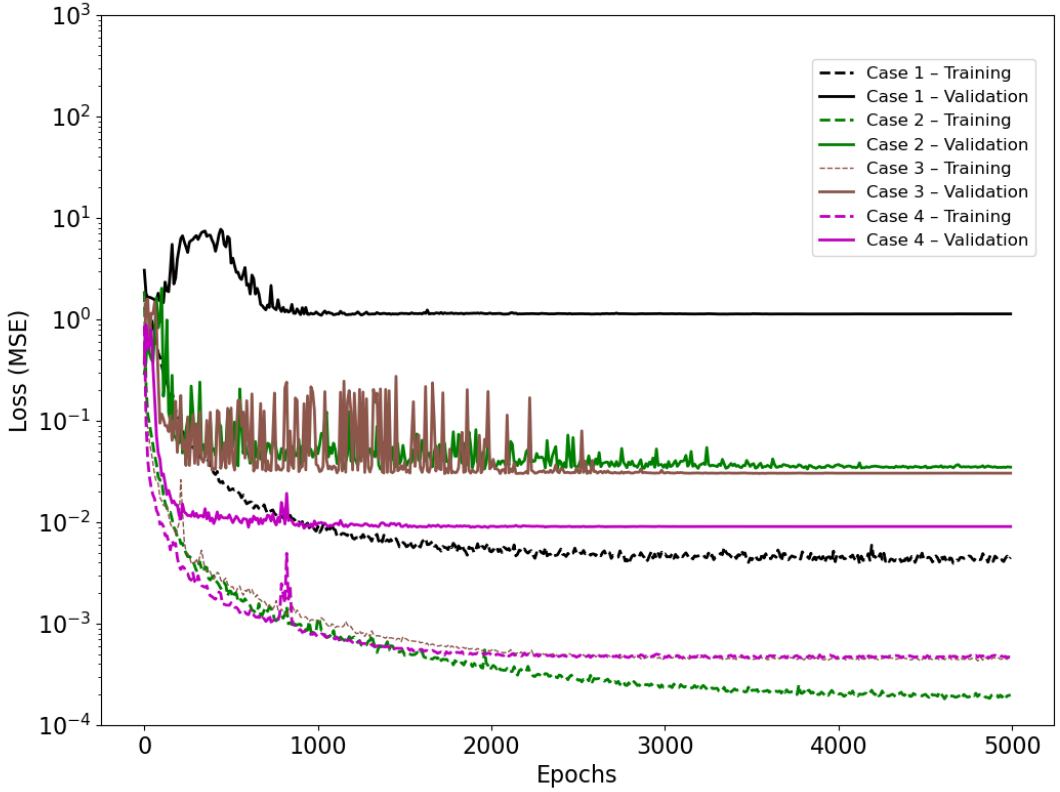
- PhaseNet model to pick P & S phase arrival times of synthetic waveforms with probabilities greater than 85% for source location estimation.
- Training samples increased from 419 to ~5,000 samples with known P & S arrival estimates
- Note that any number of new waveform data can be generated.



Multimodal CNN (MCNN) model for source location

SeismoML (WGAN) for synthetic waveform data generation → PhaseNet for arrival times → **Multi modal CNN for source locations** → Source location estimate of newly detected events

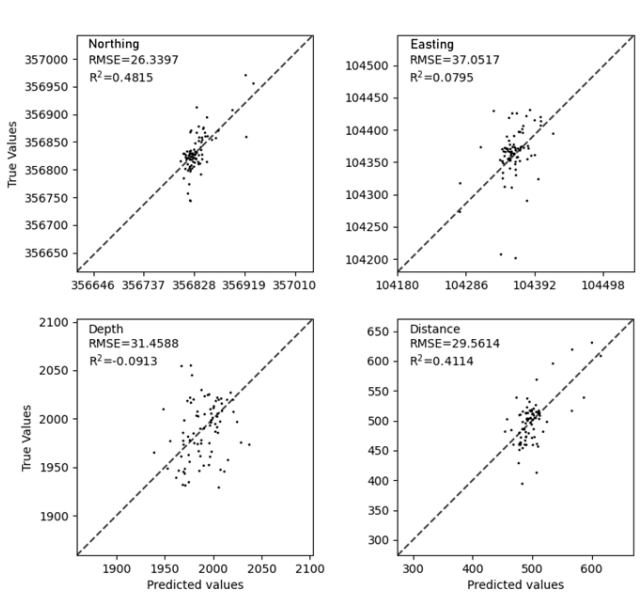




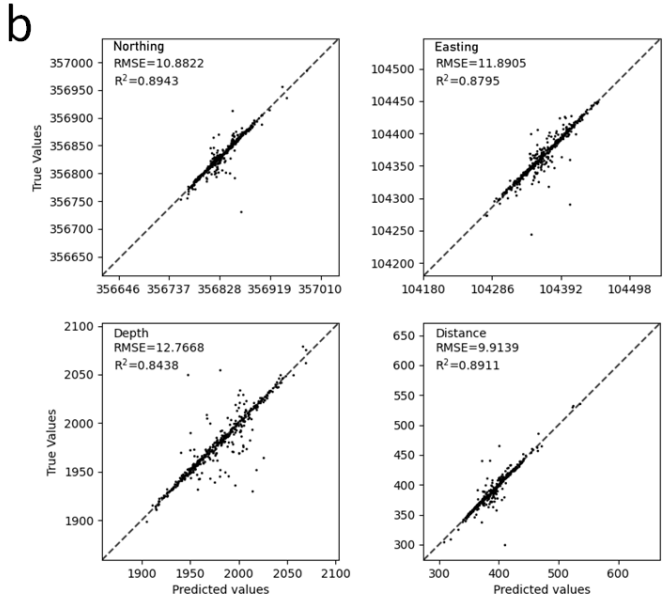
- Case 1 (417 events) : no synthetic data
- Case 2 (2132 events)
- Case 3 (3682 events)
- Case 4 (4956 events)

Impact of synthetic data on ML training

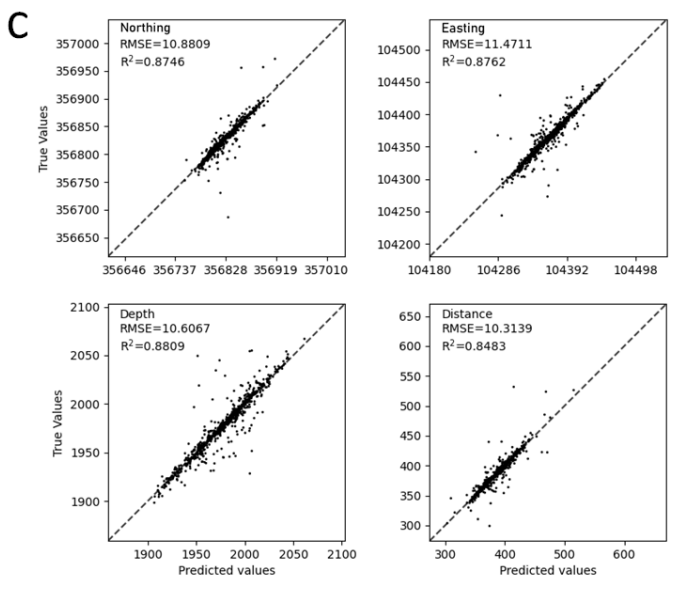
- Case 1 (417 events) no synthetic data



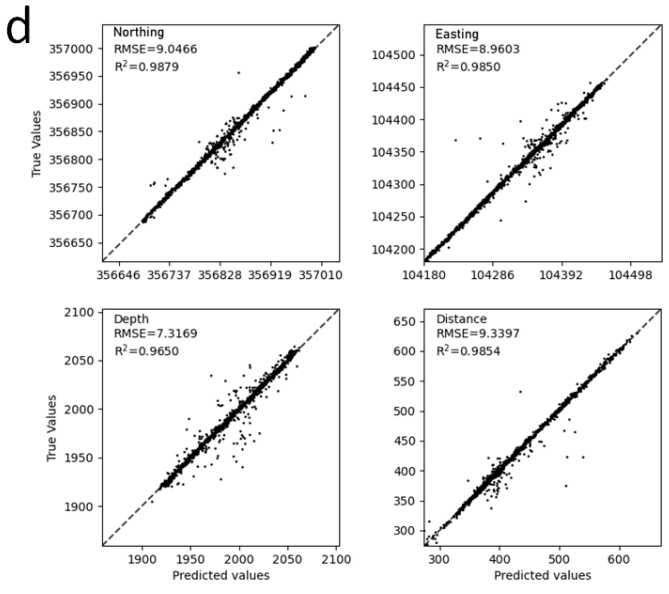
- Case 2 (2132 events)



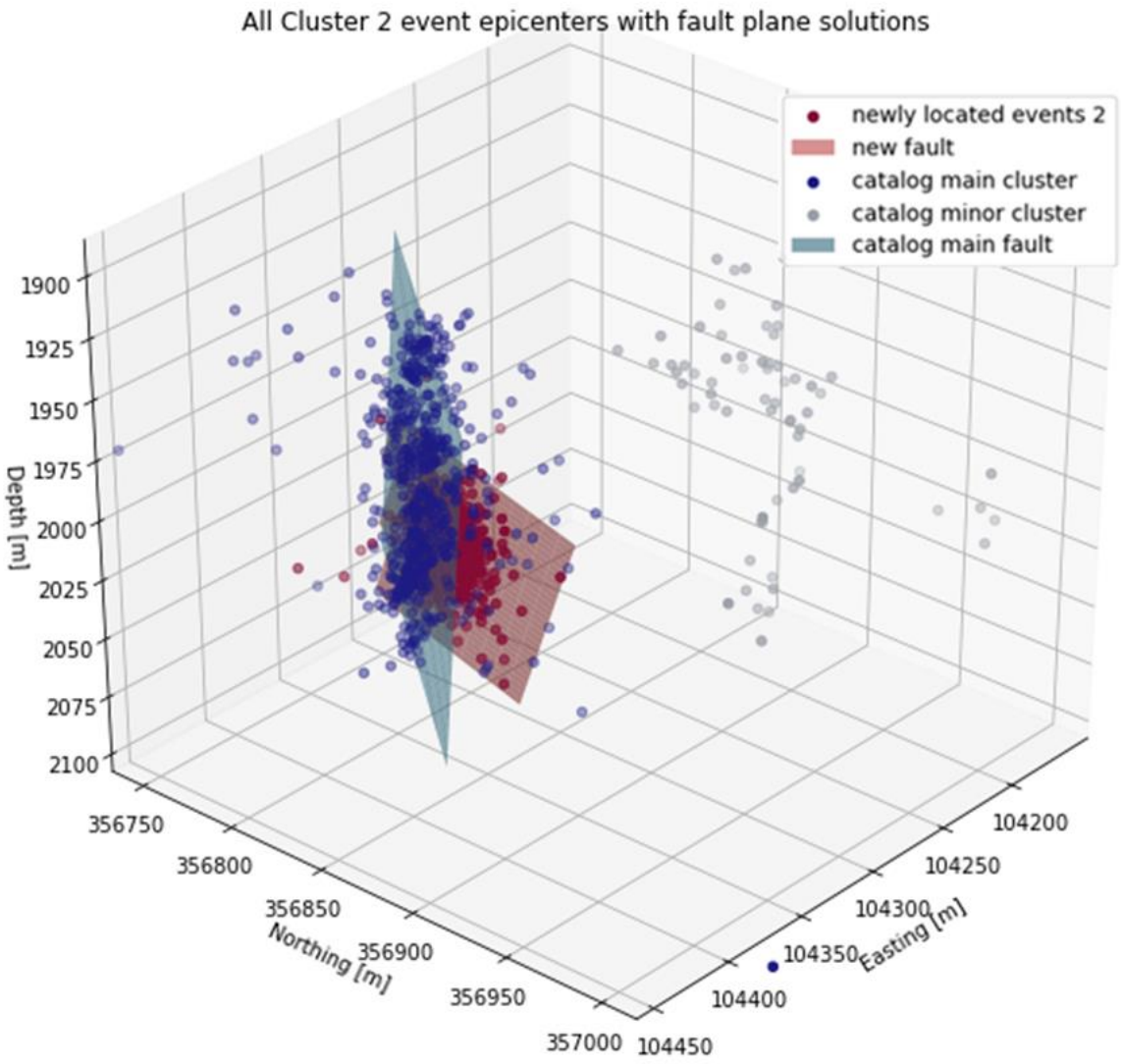
- Case 3 (3682 events)

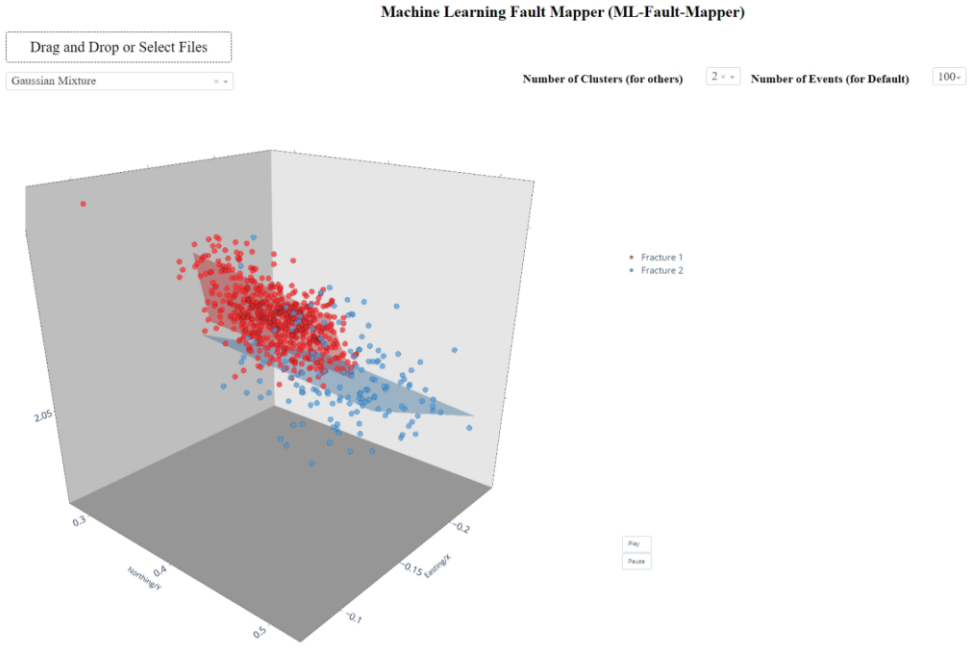


- Case 4 (4956 events)



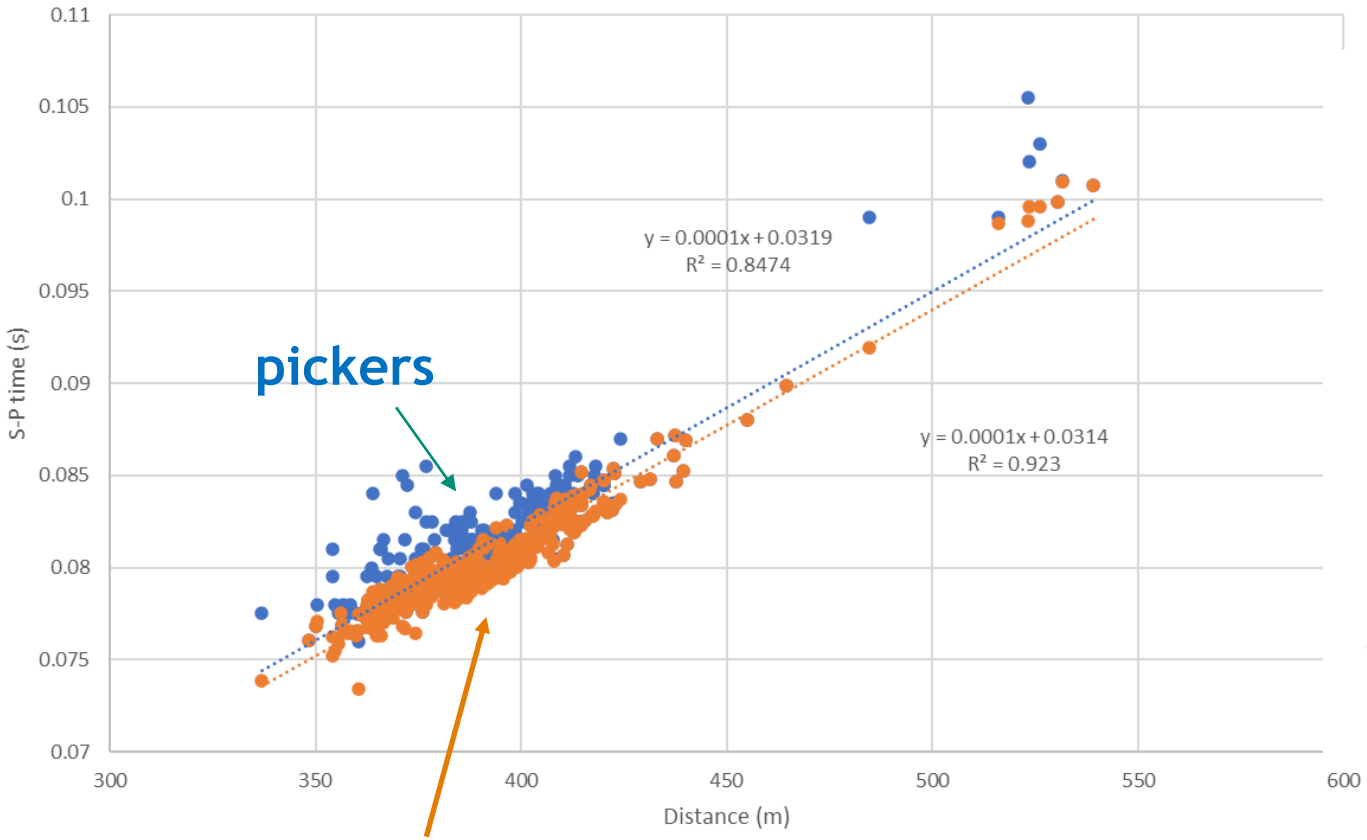
Fault/fracture plane construction

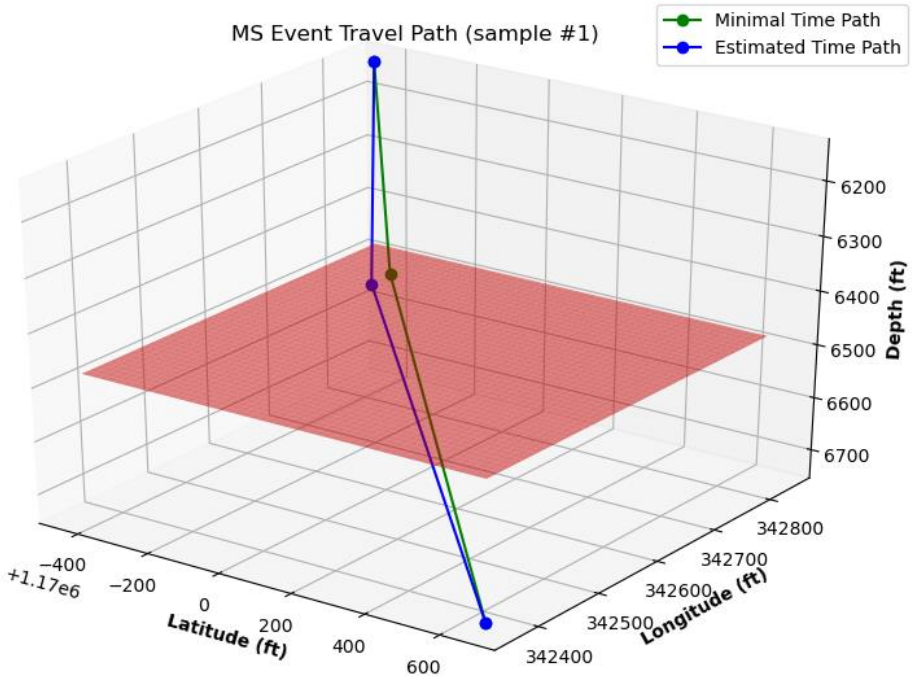




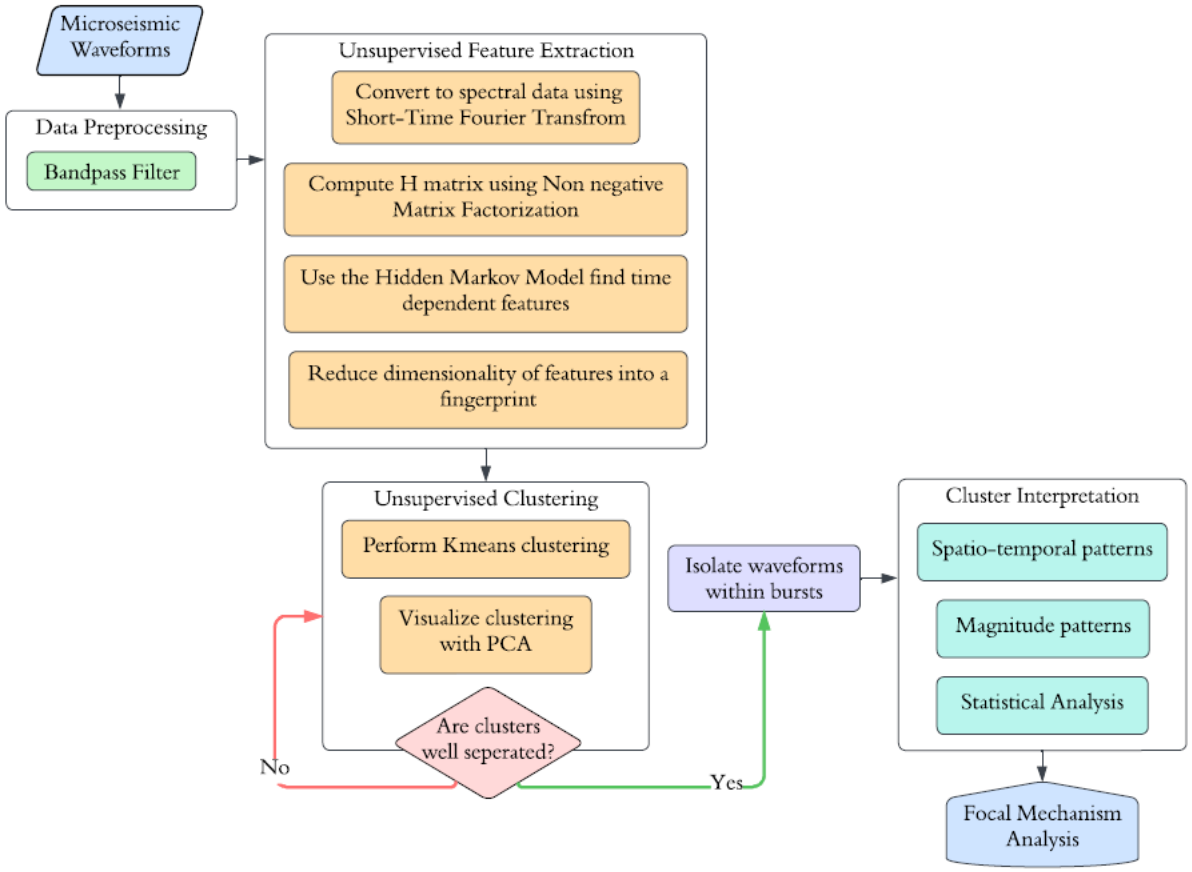
Evaluation of source location accuracy for newly detected events

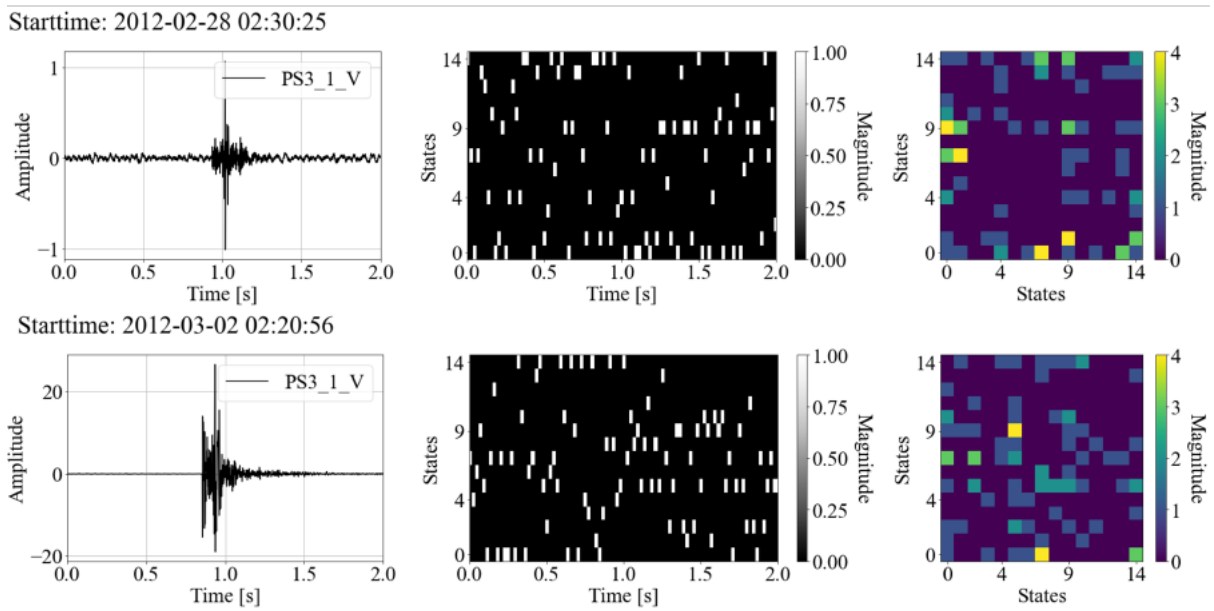
♦ S-P arrival time difference





Unsupervised ML clustering of MS events



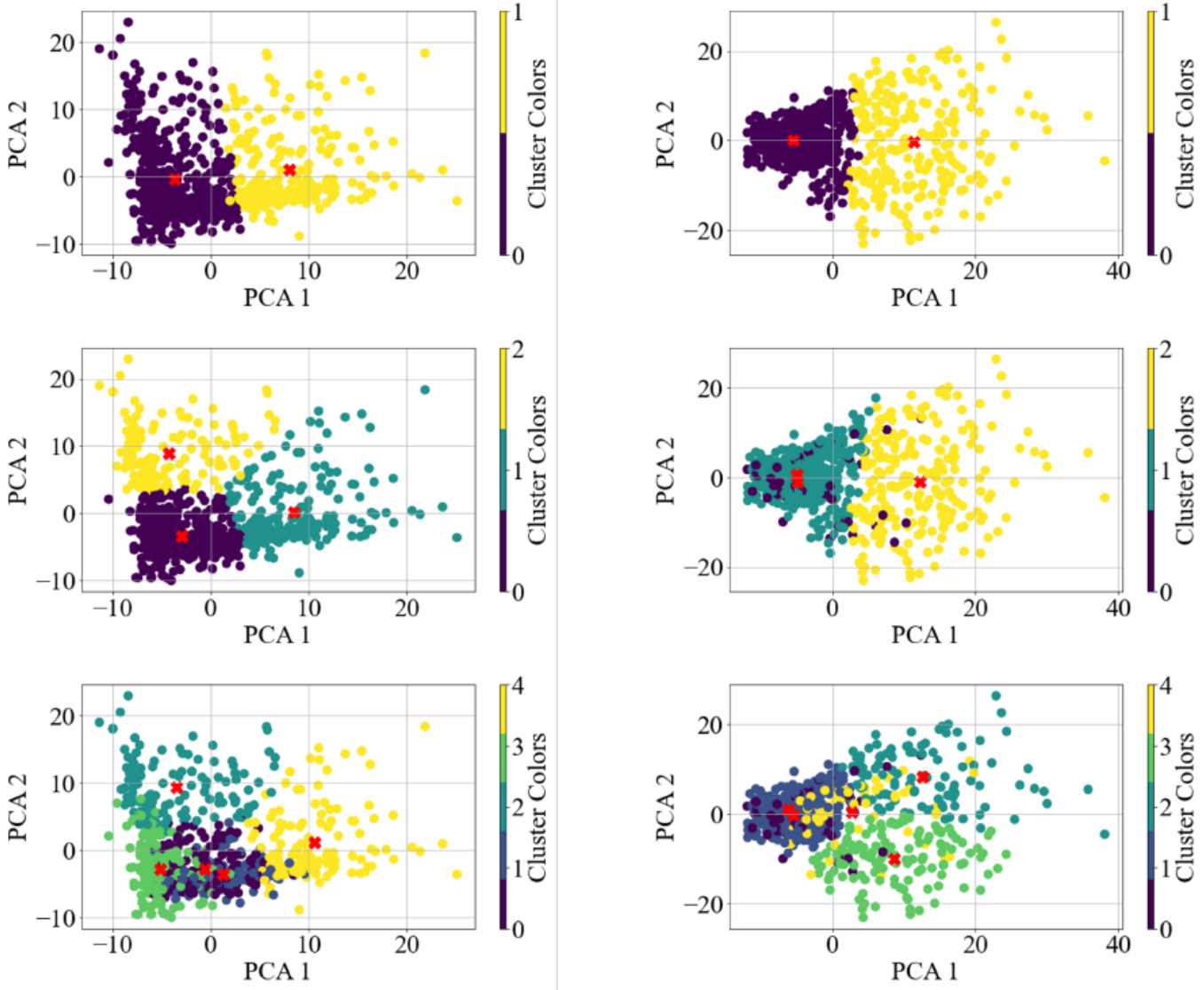


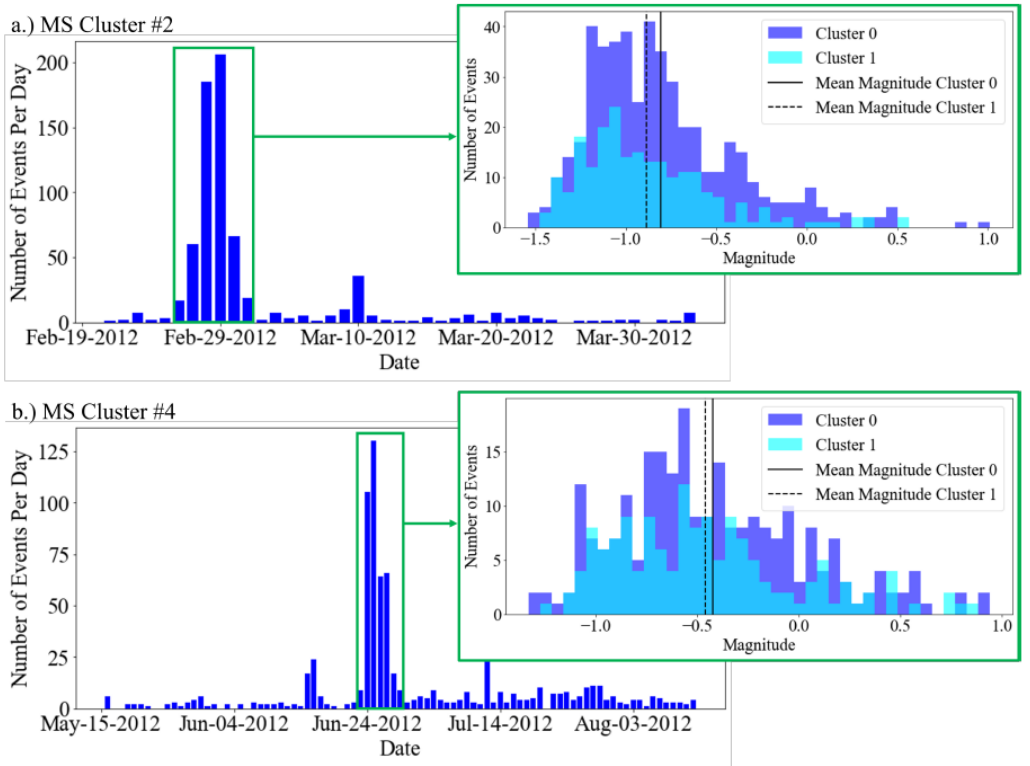
Willis et al. (JGR-ML in review)

Unsupervised Clustering (PCA)

a.) MS Cluster #2

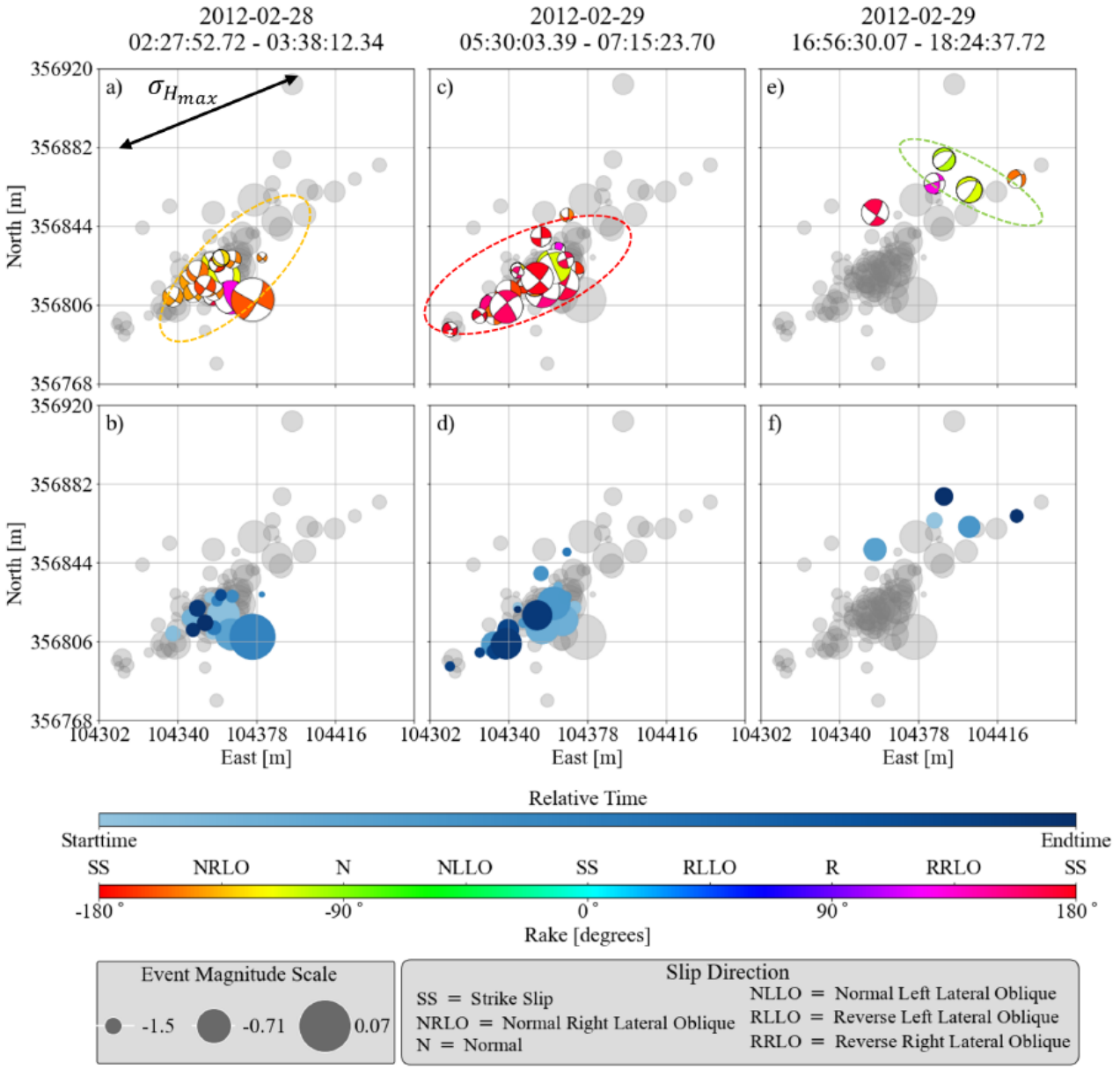
b.) MS Cluster #4



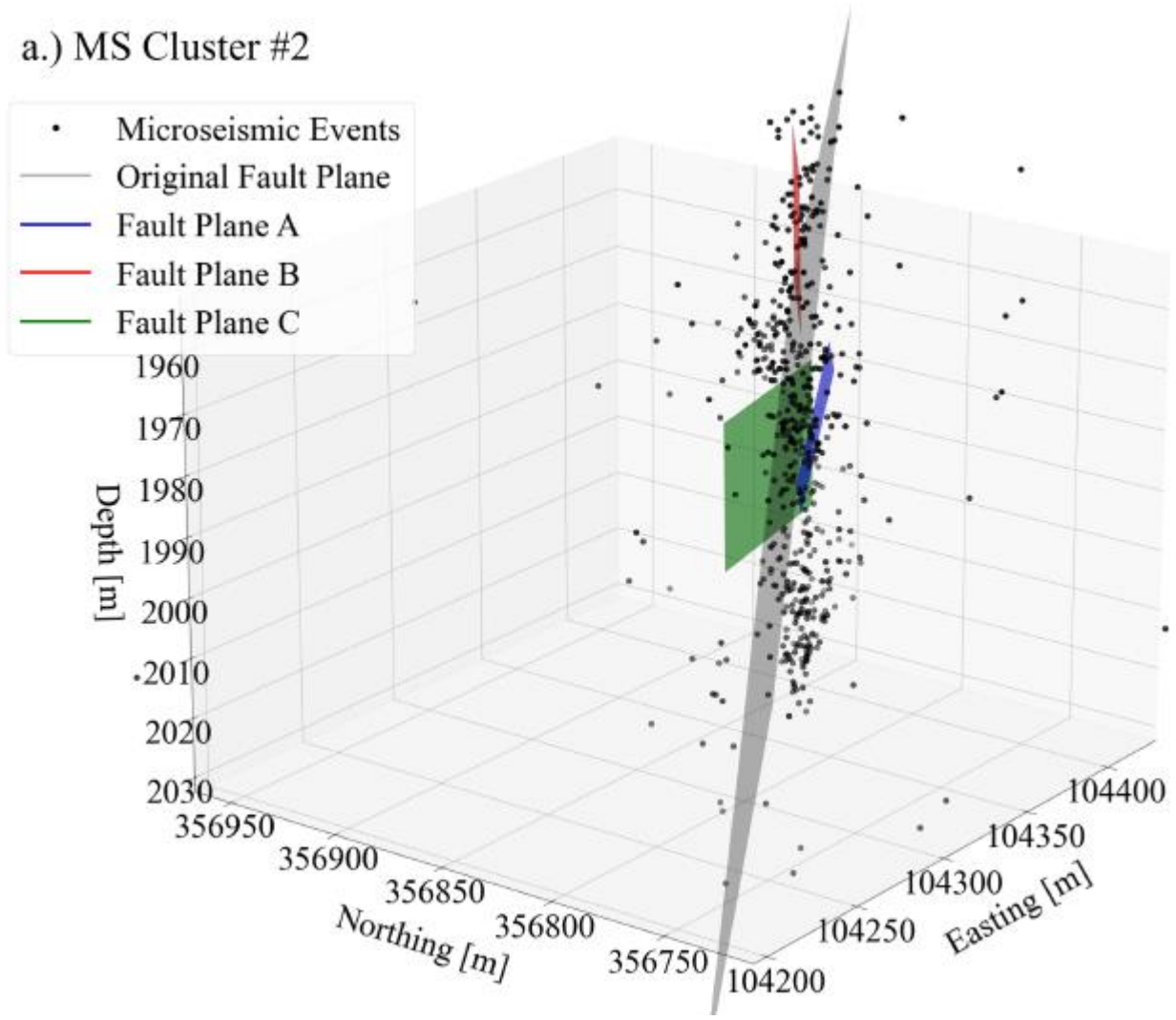


Focal Mechanism Analysis

- HASH (USGS)
- Focal mechanism analysis of ML cluster 0 events
- A general trend of north easterly propagation movement with distinct groups of events based on location and time.
- Earlier events occur towards the center with the major events being Normal Right Lateral.
- The next group happens shallower and towards the west with the main events being Right Lateral Strike Slip.
- The last group is located further East with the main events being both Reverse Right Lateral and Normal



Fault plane construction based on clustering



Summary

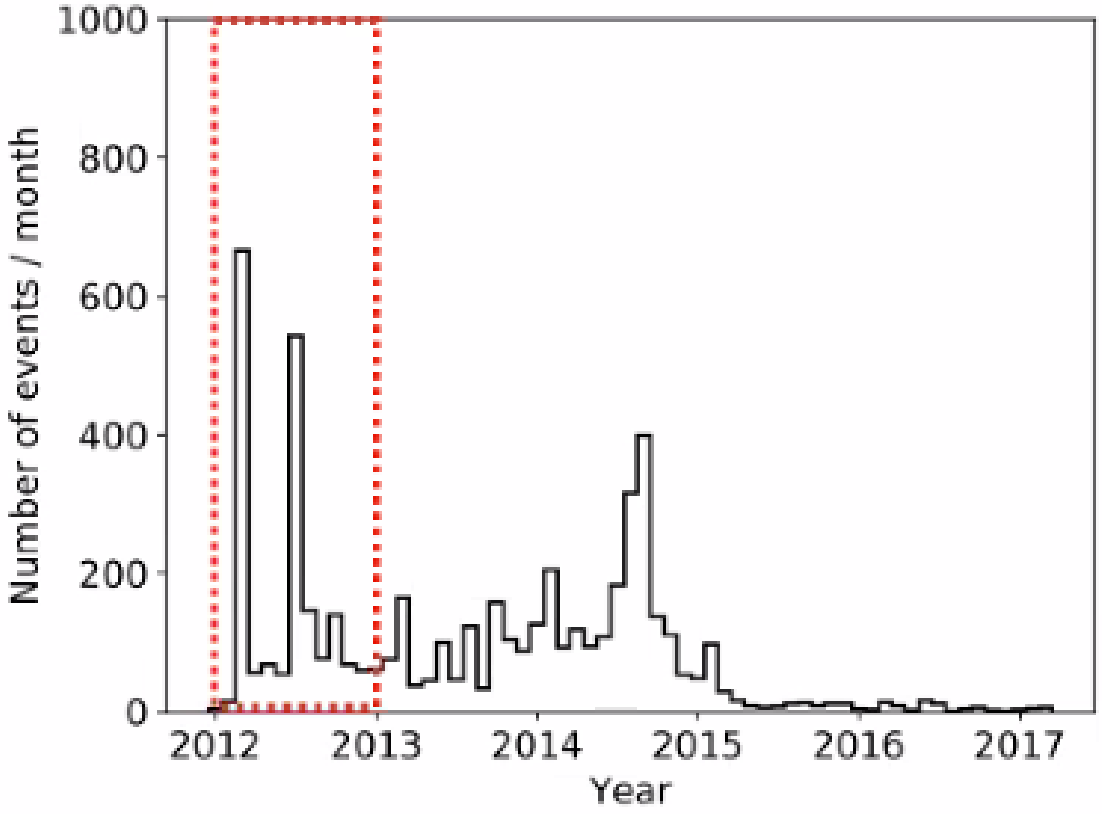
- Rescaled spectrograms as input to ML training dramatically improved ML accuracy
- Re-trained PhaseNet has a relatively high accuracy of phase arrival time picking
- During the transition period, seismic events tend to be long and overlapped (i.e., slow slip and multiple events) and PS3-2 tends to be higher amplitude than PS3-1 ➔ very distinctive from active and post periods
- Synthetic data with quality control (WGAN-GP and PhaseNet) increased high quality events volume with known P & S phase arrivals. Increasing training samples allowed the MCNN to improve the accuracy of source location estimation dramatically
- Clustering and focal mechanisms analysis provided in-depth analysis of slip mechanisms
- Sequence of sub-clusters of MS events indicates the directional stability within the fault architecture, which matches focal mechanism analysis results

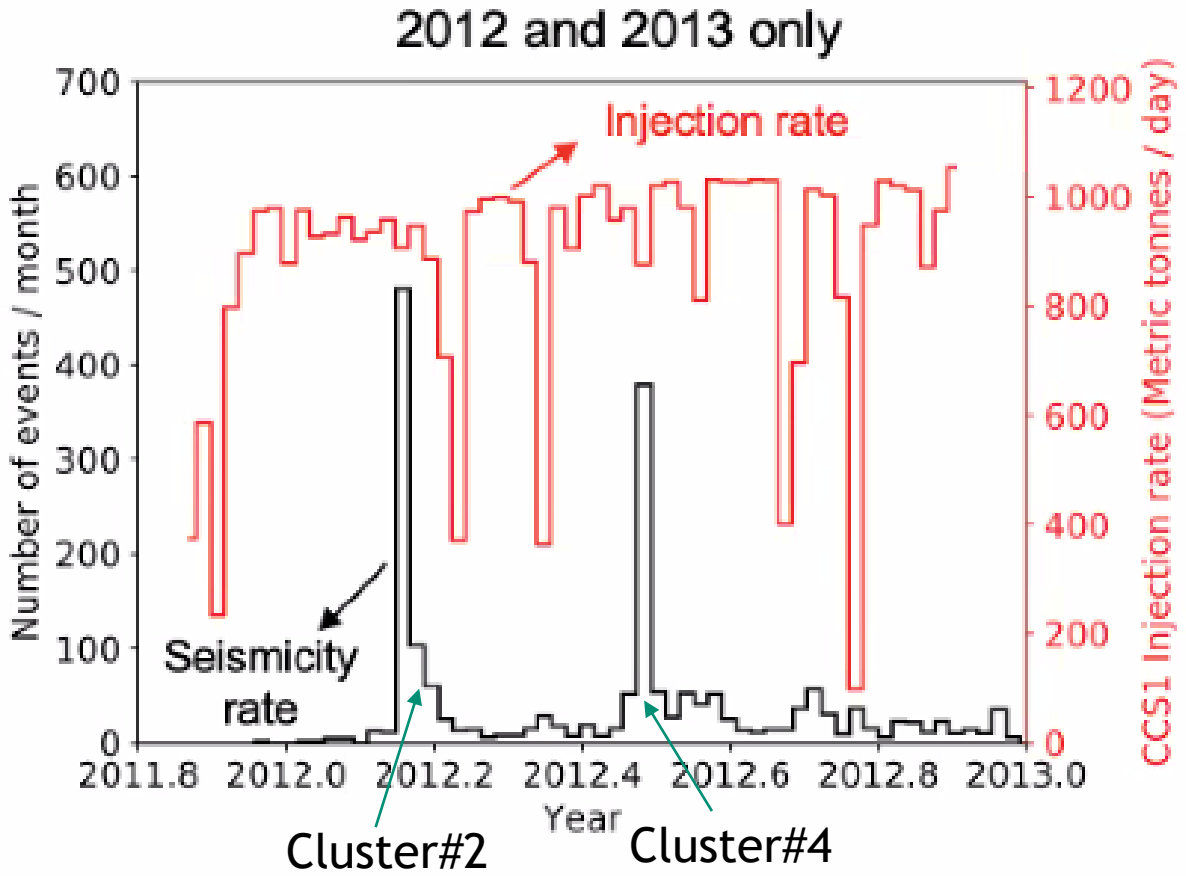
Thank You!

?

No clear relation between seismic rates and injection rates

Key observation:
Poroelastic response is time-dependent





Silva et al. (BSSA, 2024)

Evolution of microseismic events in MS Cluster #2

strike (blue),
dip (green),
rake (red)
angles