Deep Learning Applied to Earthquake Signals

Gregory C. Beroza, Weiqiang Zhu, Mostafa Mousavi, Yixiao Sheng, Clara Yoon, Yihe Huang, Miao Zhang, Yen Joe Tan, William Ellsworth, Karianne Bergen, Yongsoo Park

(Stanford University)

Why are earthquakes an interesting target?

Earthquakes are important – deadly and destructive

Earthquakes tell us about processes that are otherwise inaccessible

Earthquakes are incompletely understood – a scientific frontier

Interesting signals
- Broadband
- Characteristic attributes
- Adjacent events are similar

Lots of data
- continuous: 100 sps, 24/7/365
- Many sensors: 1 – 10,000+
- Many small events (power law)
- Digital since ~1990
- Continuous since ~2000

Interesting Noise
- Pervasive
- Non-stationary
- Most data is noise
- Noise is useful

I'm going to talk about studying small earthquakes – why?

(1) Lots of them!

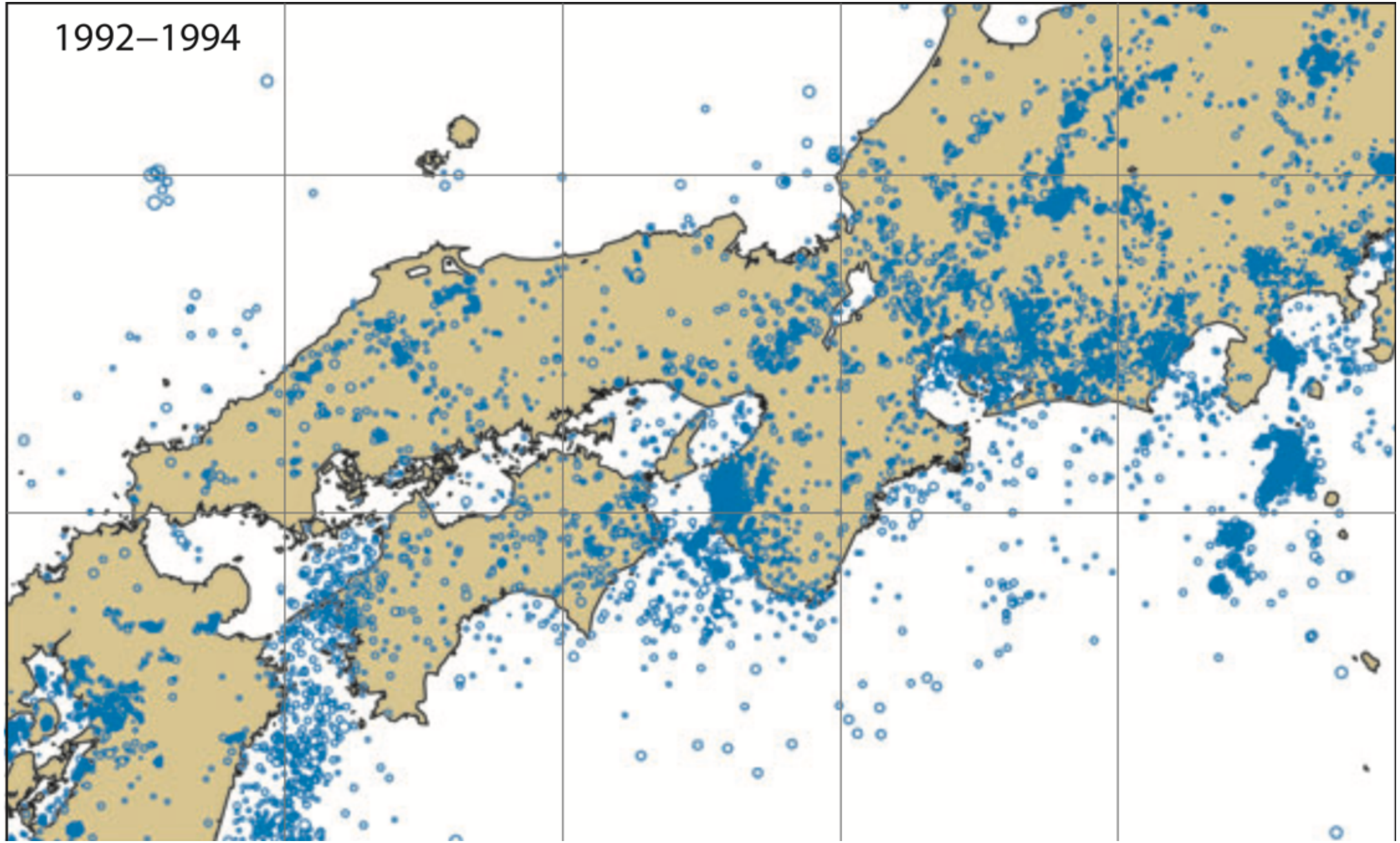
Gutenberg-Richter Statistics: $N(M) = 10^{a - bM}$ (where $b \sim 1$)

Reducing detection threshold by 1 magnitude unit yields 10x more earthquakes.

(2) They inform fault geometry by illuminating the same fault systems as large earthquakes.

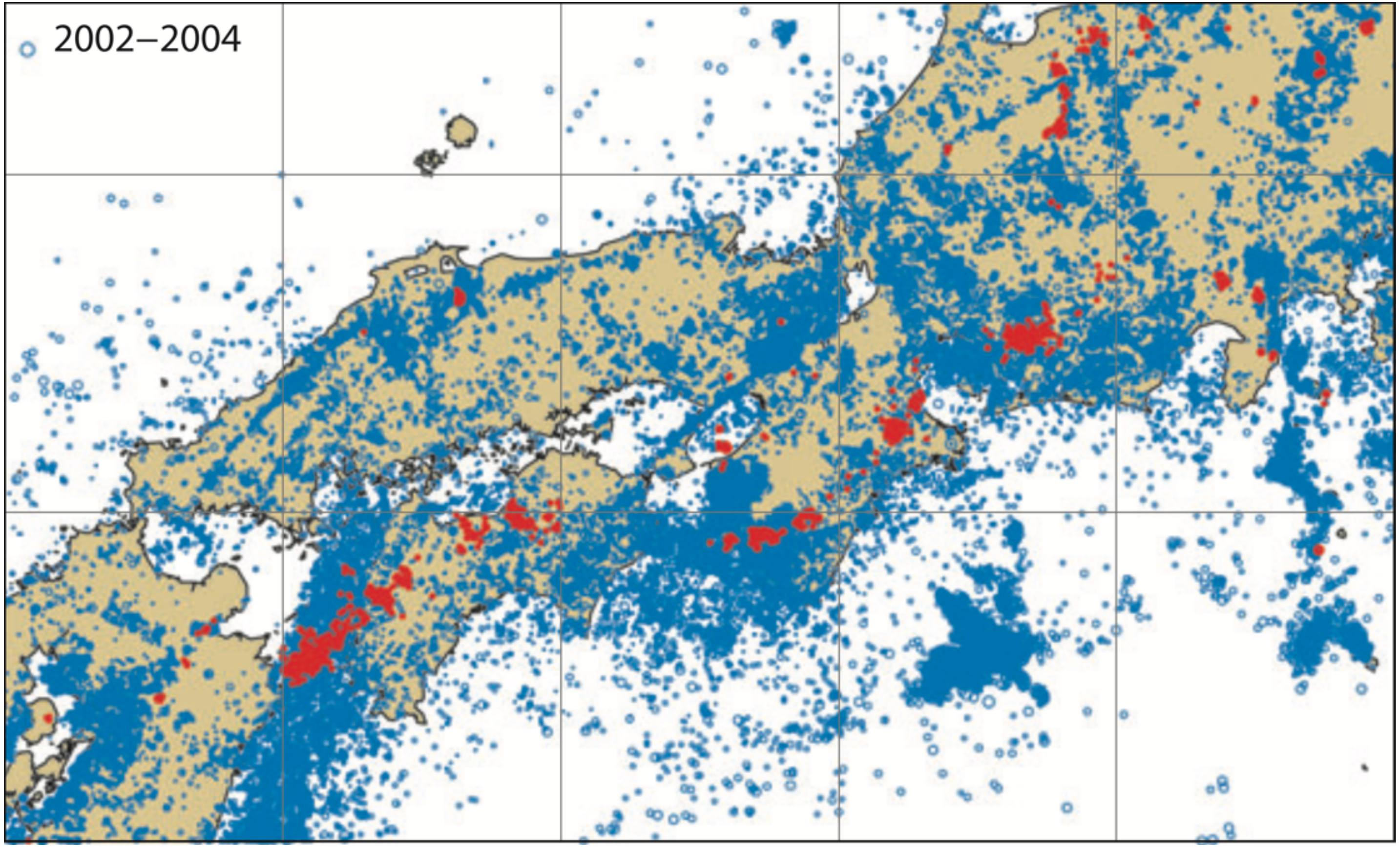
(3) They appear to behave like large earthquakes, suggesting that understanding small earthquakes will help us understand large ones.

Clearer View of Fault Systems and Discovery of New Behaviors



Beroza and Ide [2011]

Clearer View of Fault Systems and Discovery of New Behaviors



Beroza and Ide [2011]

Earthquake Monitoring Across Scales

Global (10^7)

National (10^{6-7})

Regional (10^{5-6})

Local (10^{4-5})

Reservoir (10^{3-4})

Mines (10^{1-3})

Lab ($< 10^0$)

Decreasing Array Aperture

Increasing Sensor Density

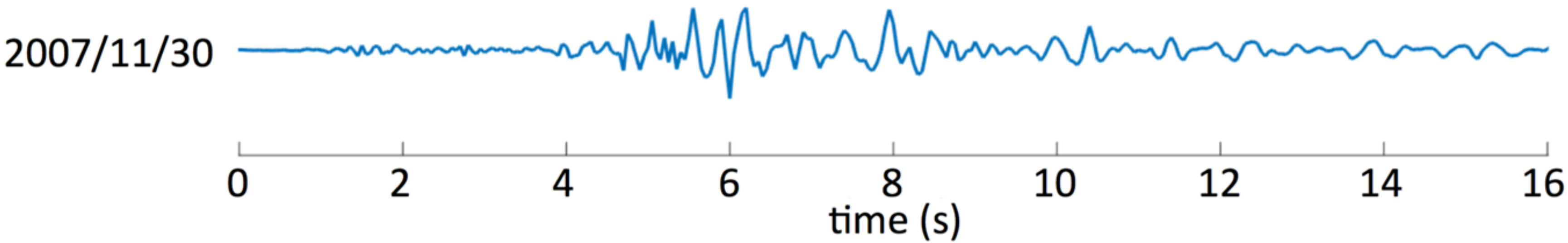
Increasing Frequency

Common Goal:

- **detect**
- **locate**
- **characterize**

earthquakes as completely and accurately as possible.

Our data are seismograms



Digital recordings of ground motion (typically velocity)

Ideally vector quantity = 3 channels

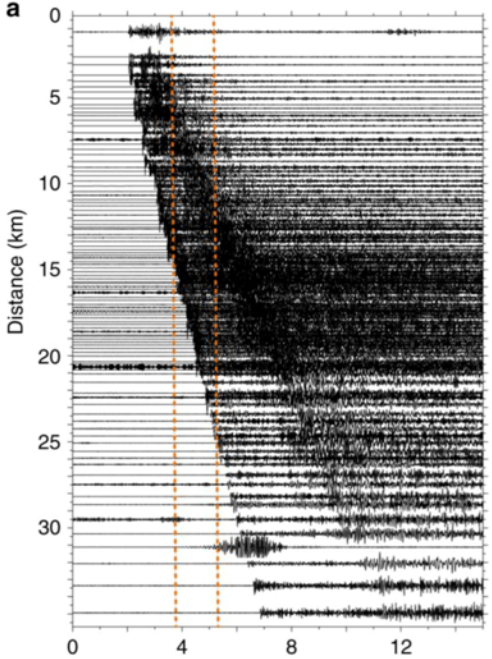
Typically 100 Hz sampling rate

Continuous monitoring 24 x 365

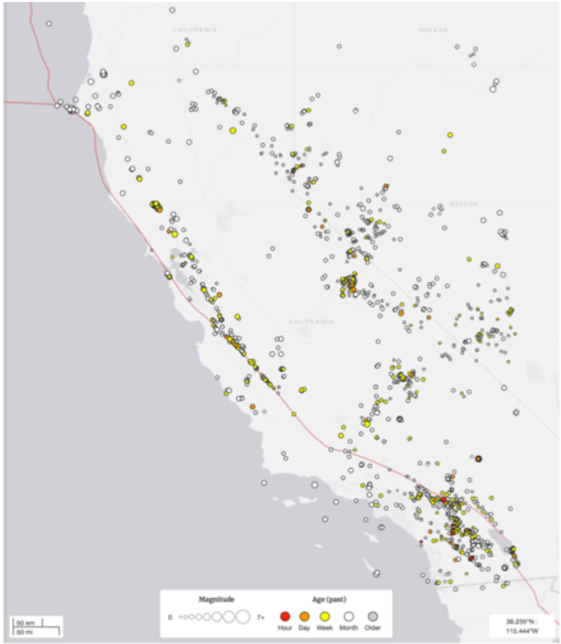
Stations/Networks/Arrays of sensors (1/10-100/100-10,000+)

Earthquake Monitoring Workflow

Input: continuous time series of ground motion from a seismic network

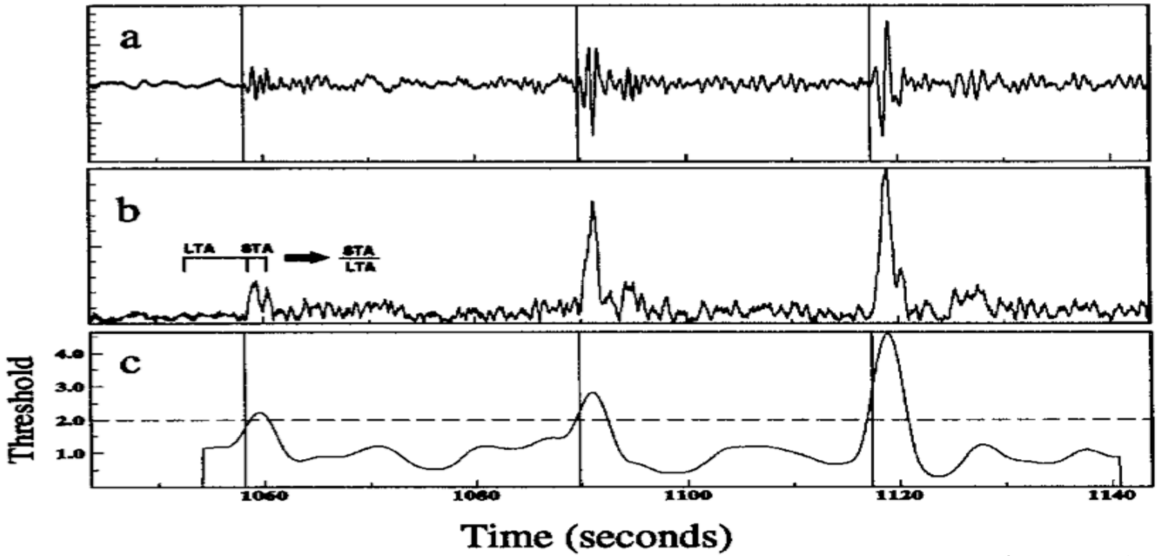


Pre-processing → Event Detection → Phase Picking → Association → Location → Characterization



Output: comprehensive catalog of discrete seismic events

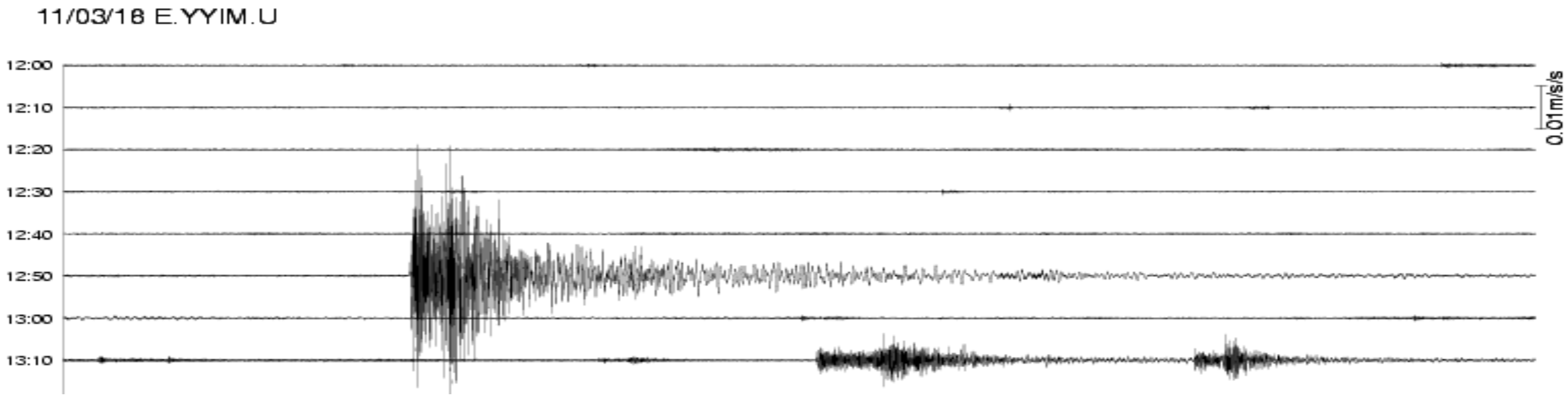
Standard Approach



Earle and Shearer [1994]

(1) Detection (STA/LTA)
(2) Association
(3) Location
(4) Characterization

Standard approach works well when…

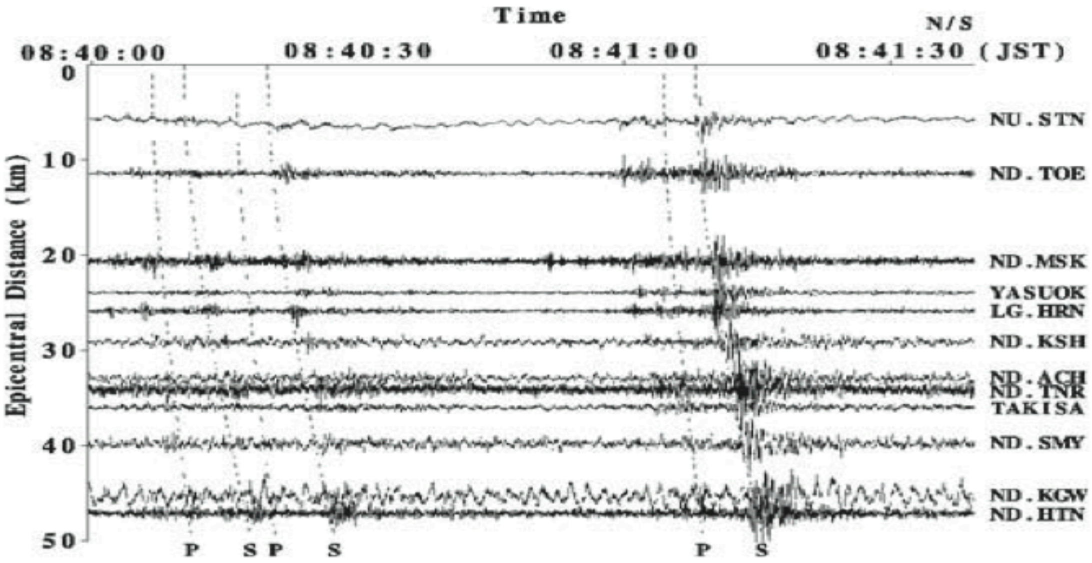


Events are recorded at > 3 stations

Events are impulsive

Events don't overlap

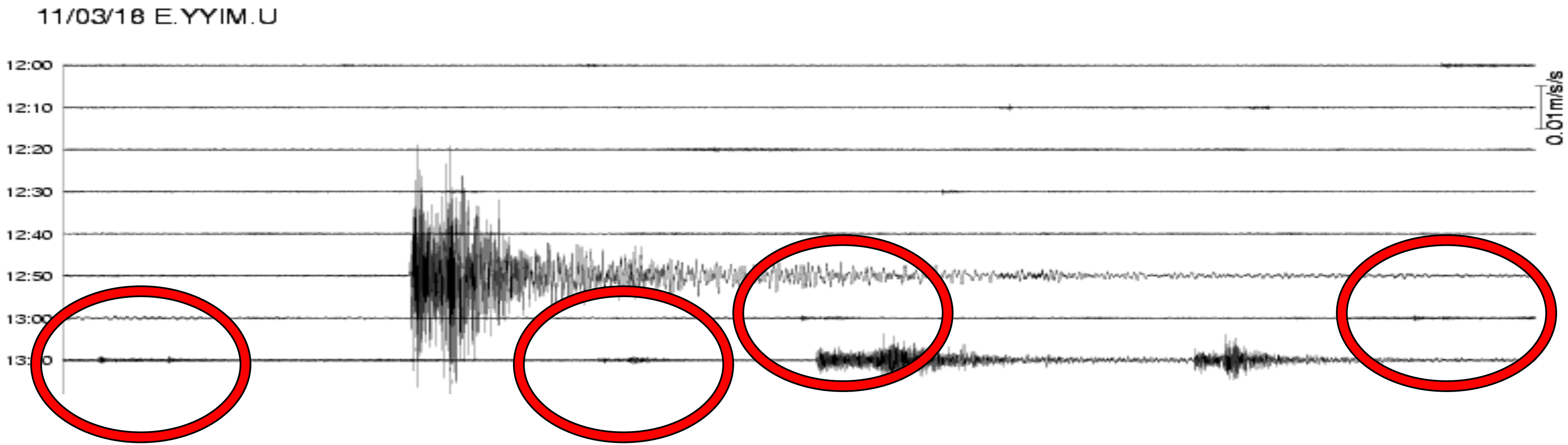
Standard approach works less well for …



…weak events with emergent arrivals (like LFEs)

Katsumata and Kamaya [2003]

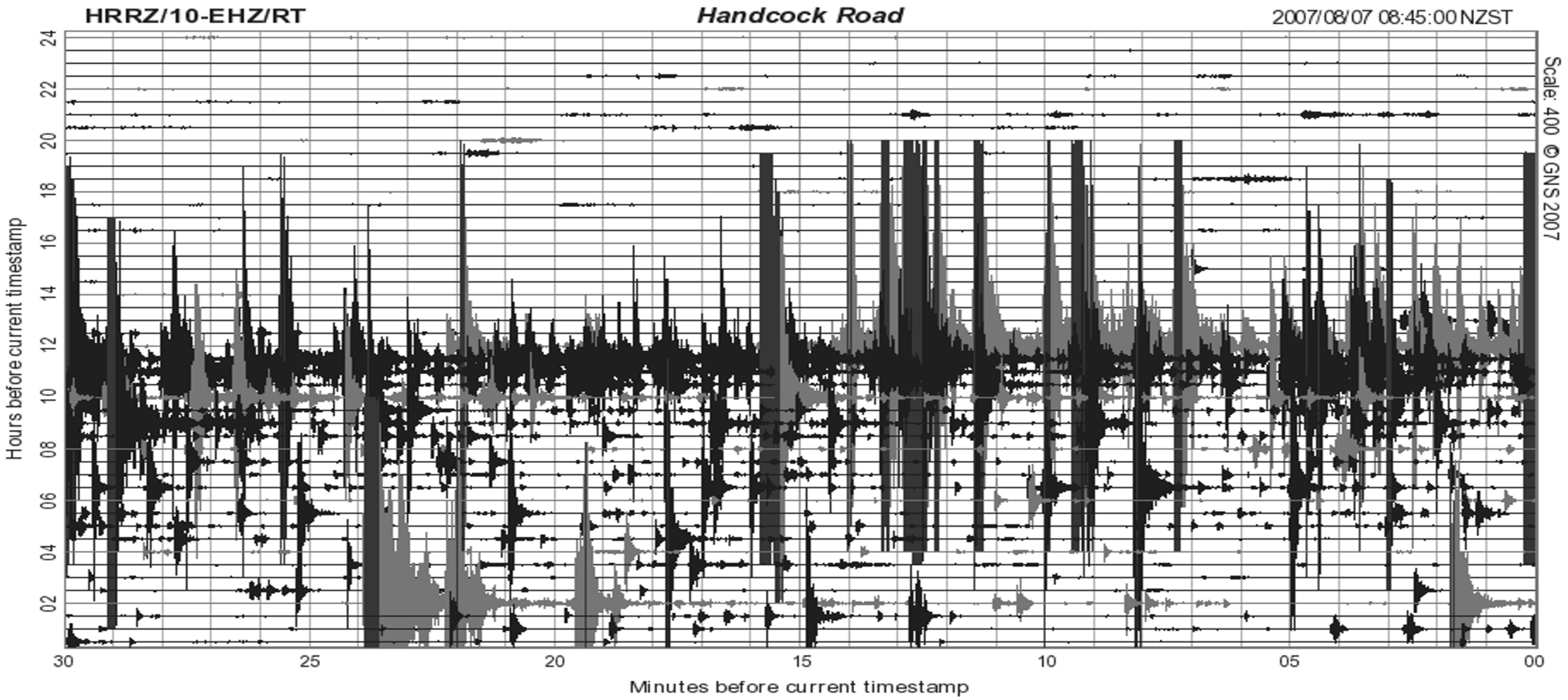
Standard approach works less well for …



…small events with too few arrivals to locate

(recall Gutenberg-Richter statistics)

Standard approach works less well for …



…Overlapping Events during intense activity

Different Approaches

Template Matching: correlation-based search

Subspace Projection: higher order correlation-based search

PageRank: uninformed clustering of multiplicity of waveforms

FAST: uninformed data mining for repeating signals

Data mining using waveform similarity

CRED: informed deep learning for earthquake detection

DeepDenoiser: deep learning for denoising

PhaseNet: deep learning for phase picking

Supervised machine learning

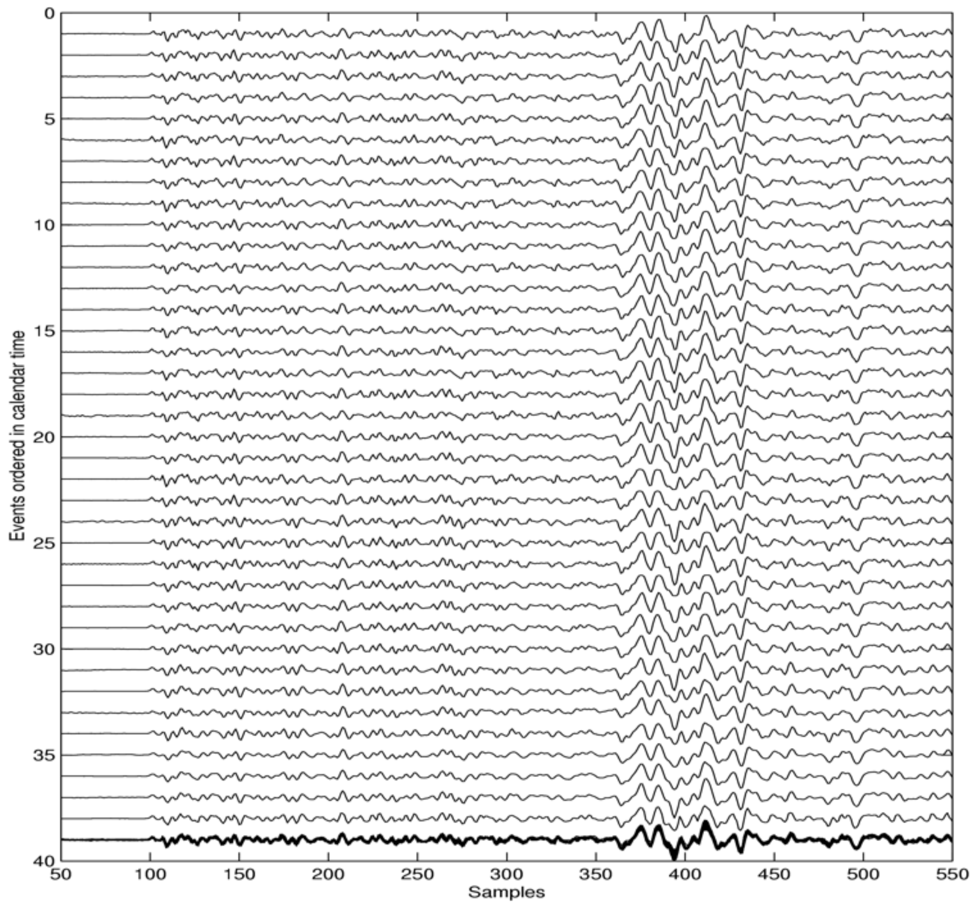
Deep Autoencoder: deep learning with a small data set to discriminate local from teleseismic earthquakes.

Self-supervised machine learning

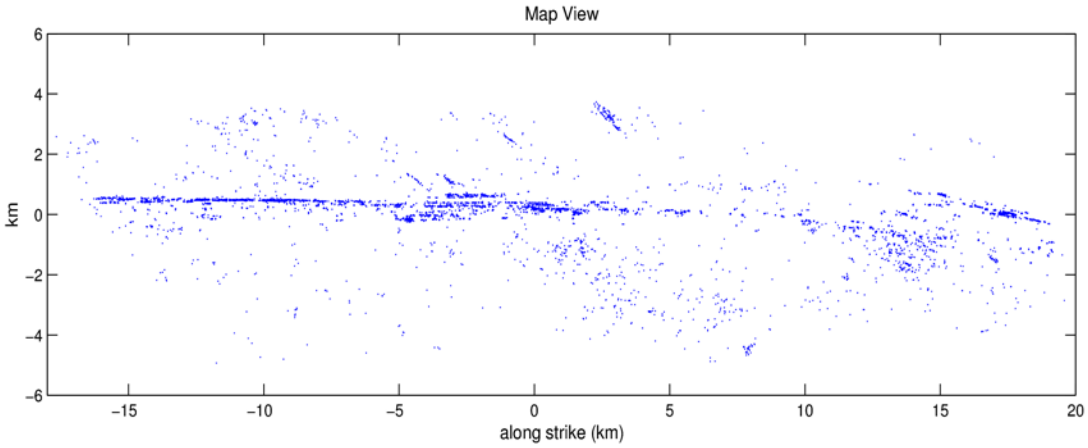
38 Repeats of an earthquake on the Calaveras Fault

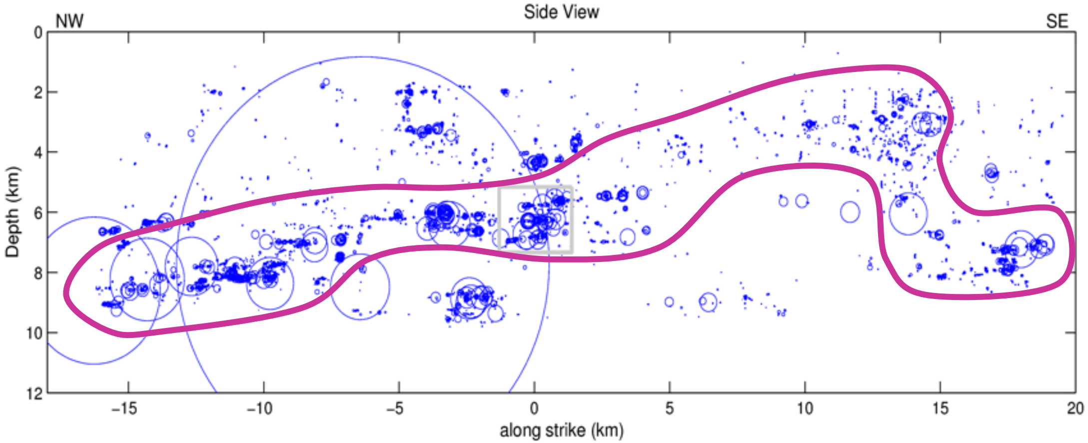
Slip occurring at different times but in the same place generates identical seismograms.

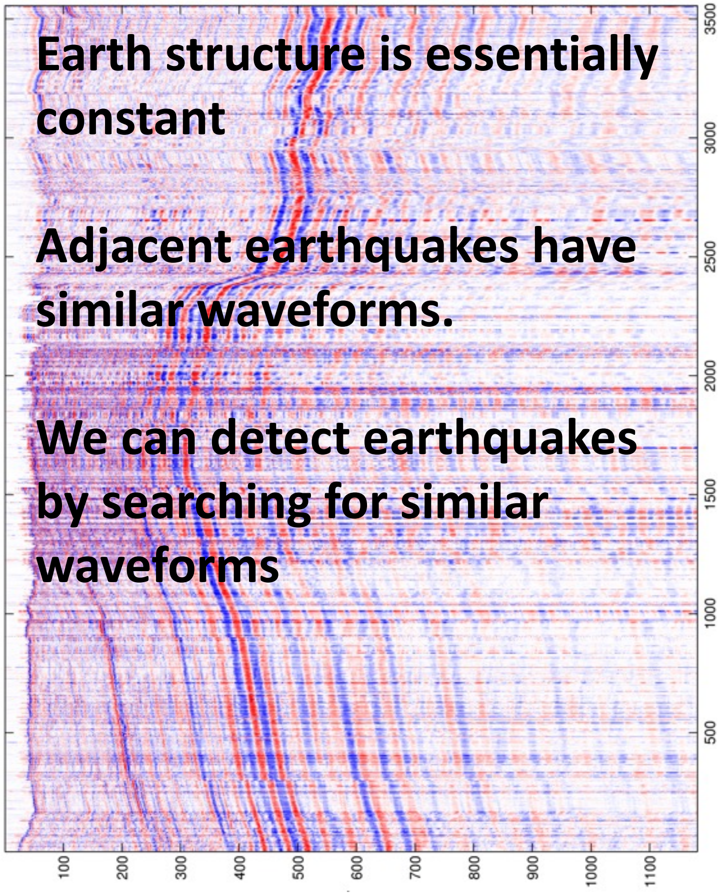
We can look for a repeating signal from a repeating source, but most sources don't exactly repeat.



Adjacent Earthquakes on the Calaveras Fault





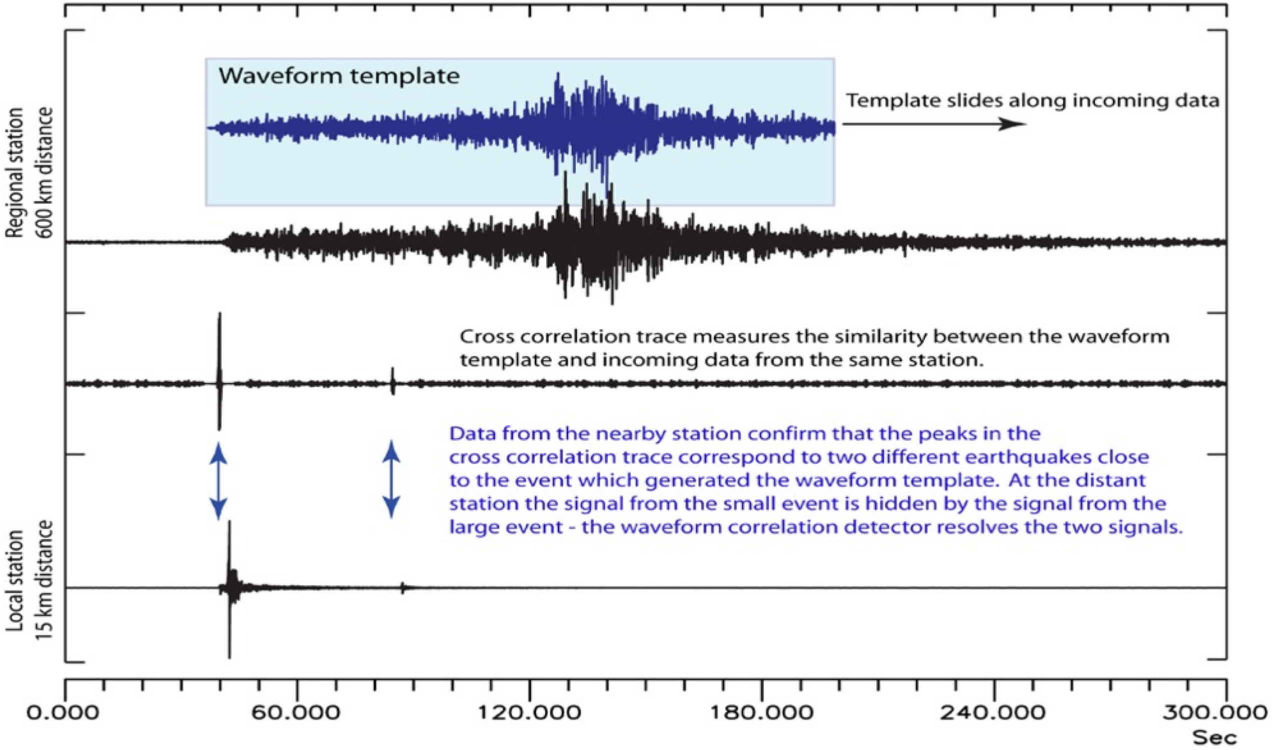


Earth structure is essentially constant

Adjacent earthquakes have similar waveforms.

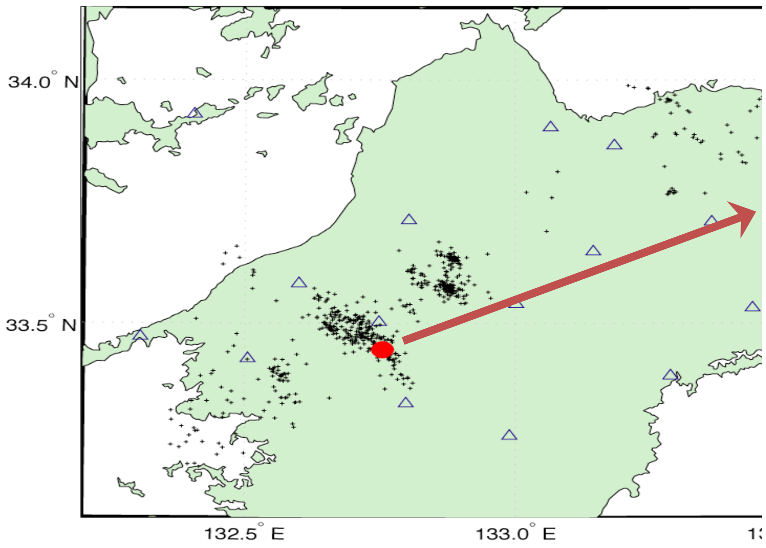
We can detect earthquakes by searching for similar waveforms

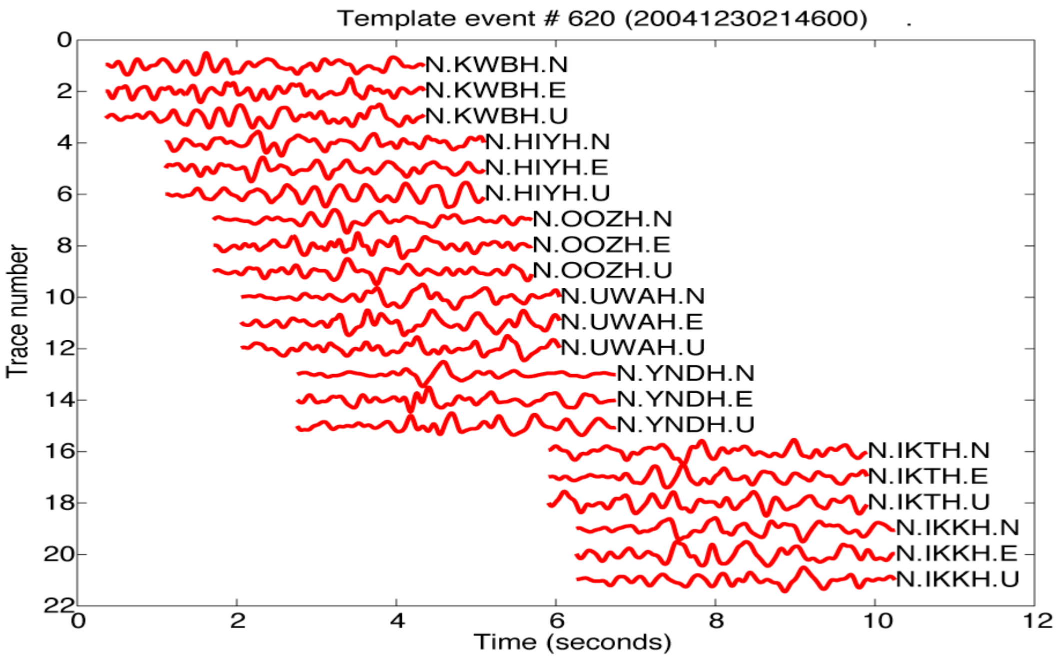
Template Matching



Gibbons and Ringdal [2006]

Template Matching

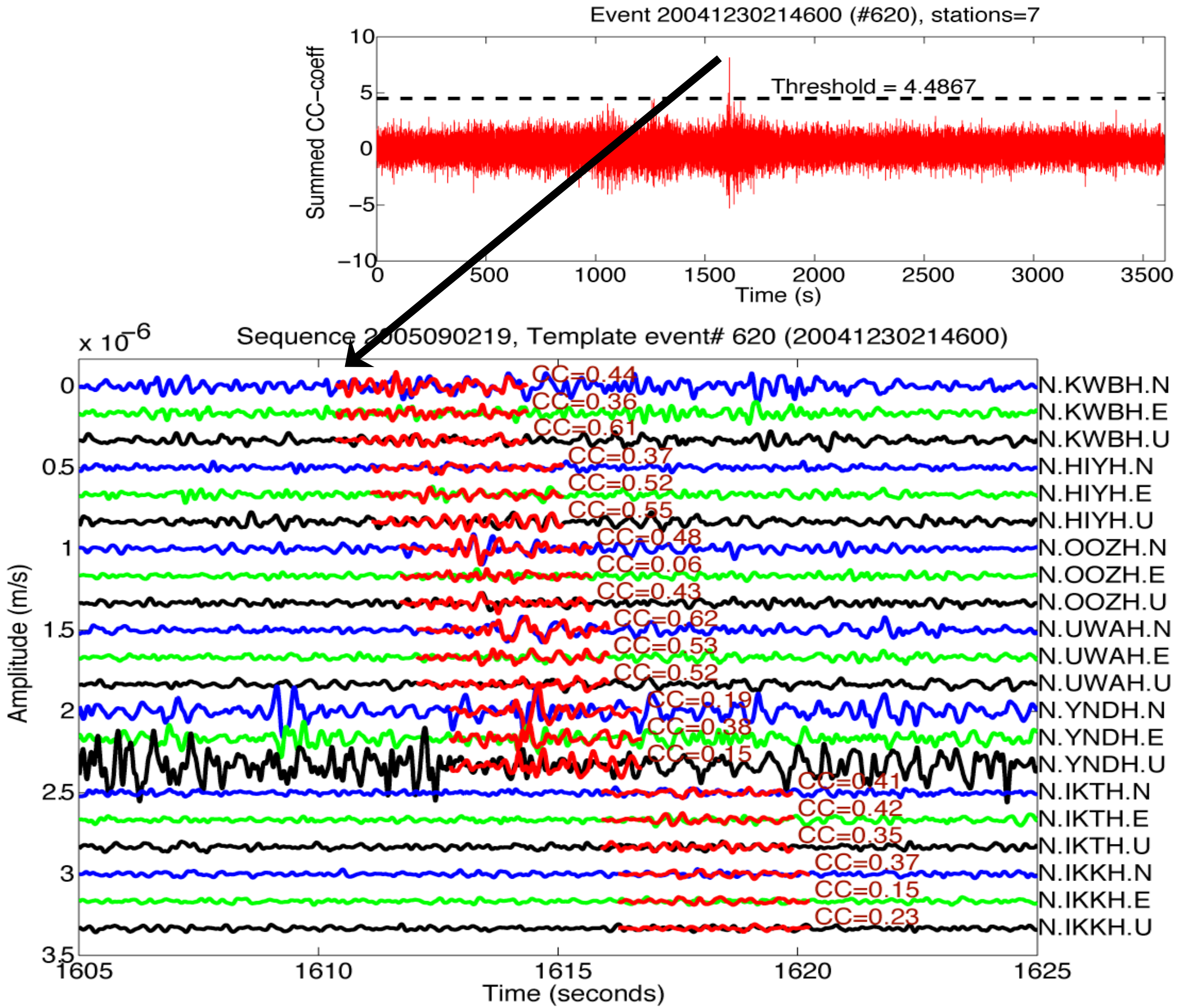




Used to show tectonic tremor is a swarm of LFEs

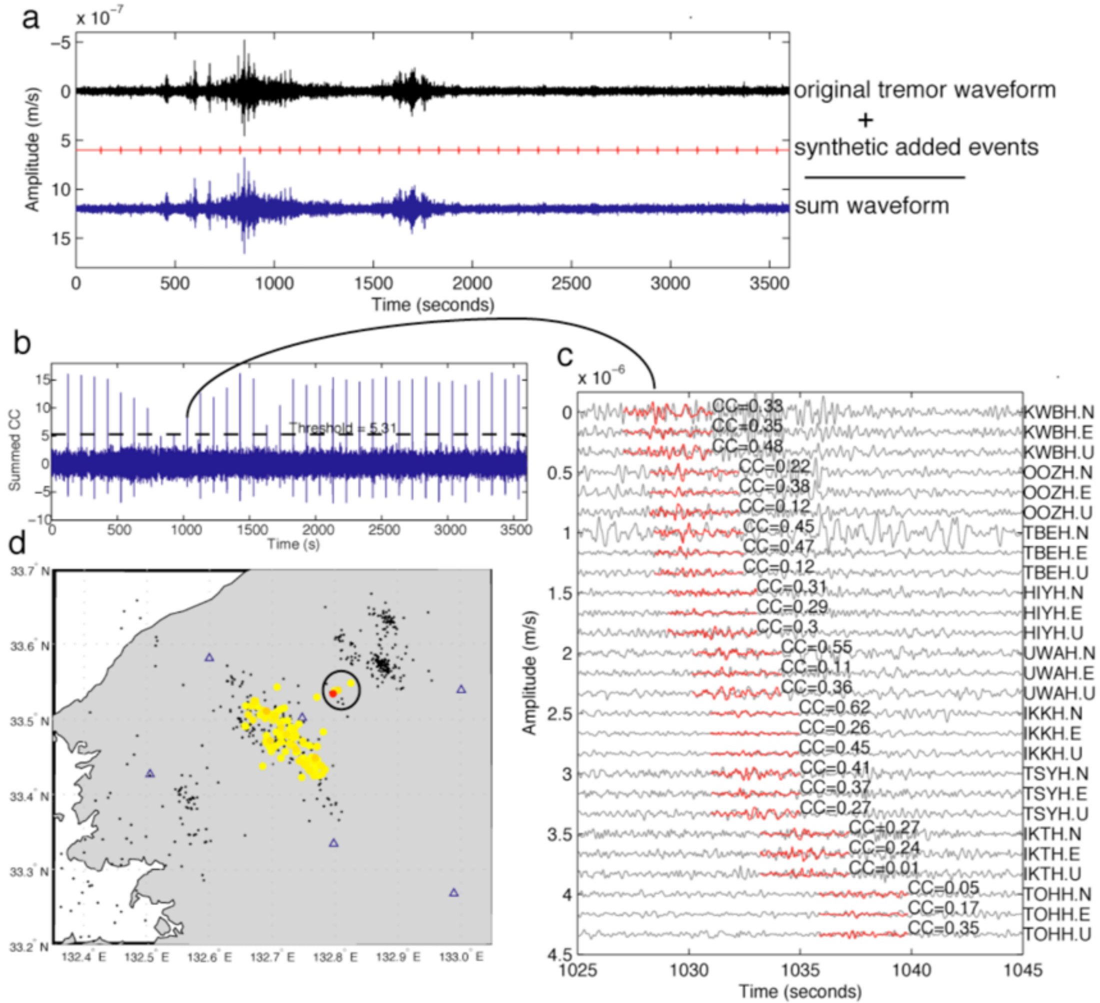
Shelly et al. [2007]

Low Type I Error Rate (false detection)



Shelly et al. (2007)

Low Type II Error Rate (failure to detect)



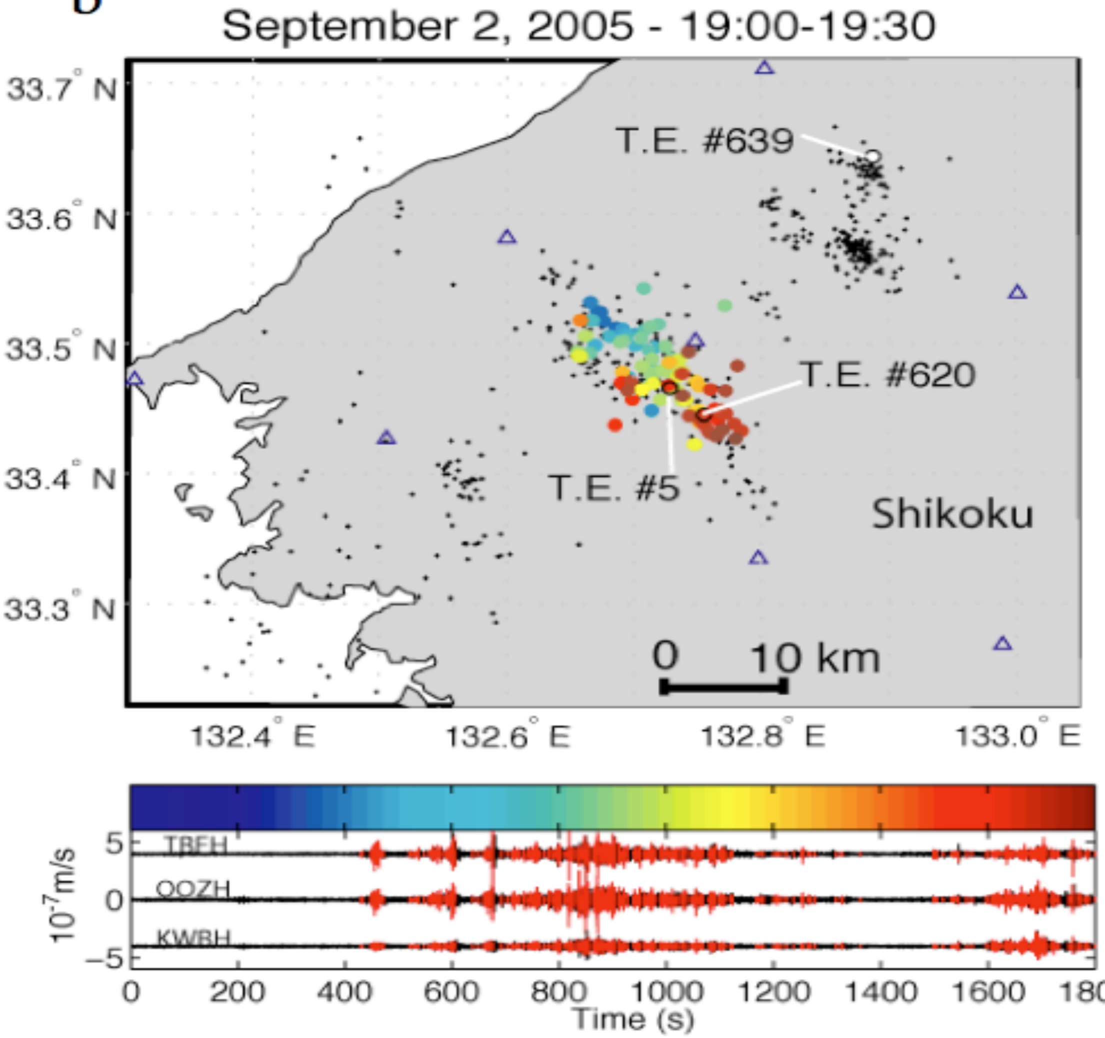
Shelly et al. (2007)

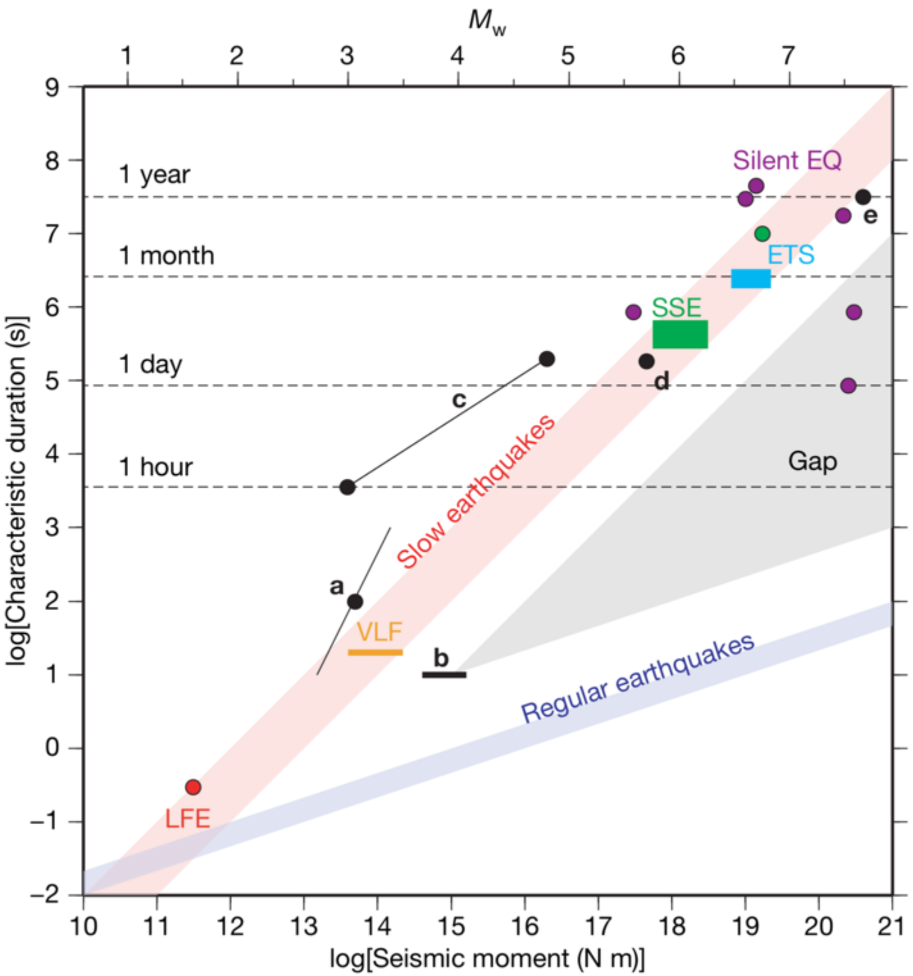
Locates tremor precisely

Tremor matches LFEs (red)

Tremor is a swarm of LFEs, each of which is M ~ 1 and ~1 s

b

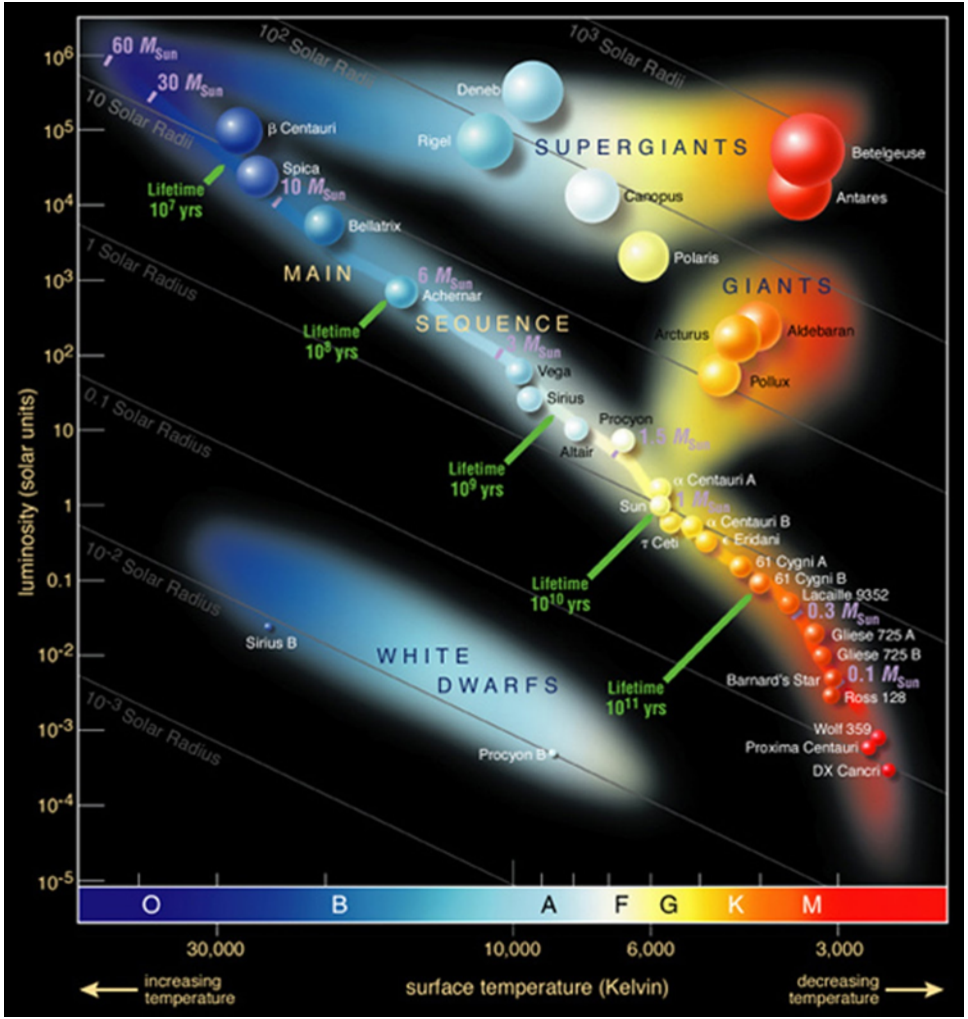




Slow vs. Fast Earthquakes

- **Presented in 2007**
- **Empirical - physical understanding is a work in progress.**

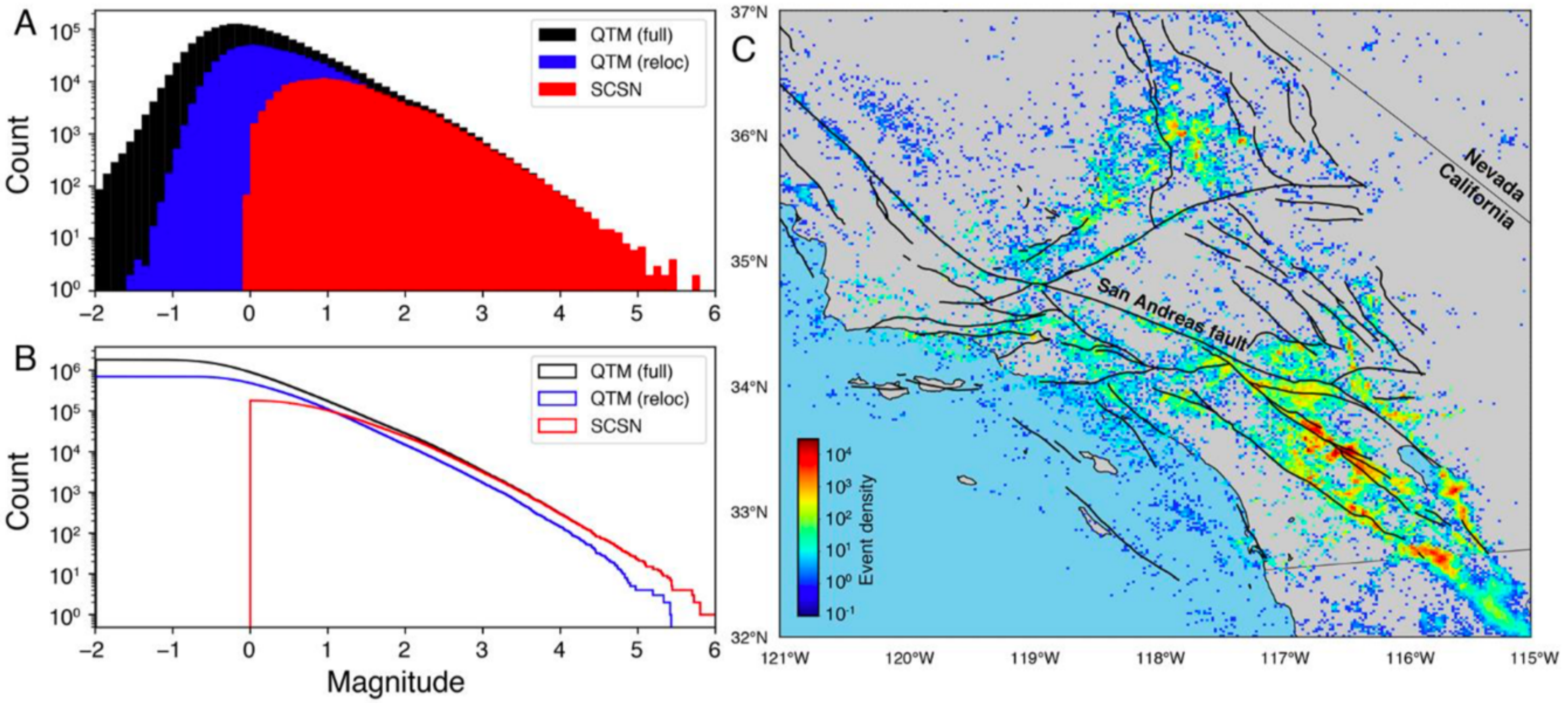
Ide et al. [2007]



Hertzsprung-Russel Diagram

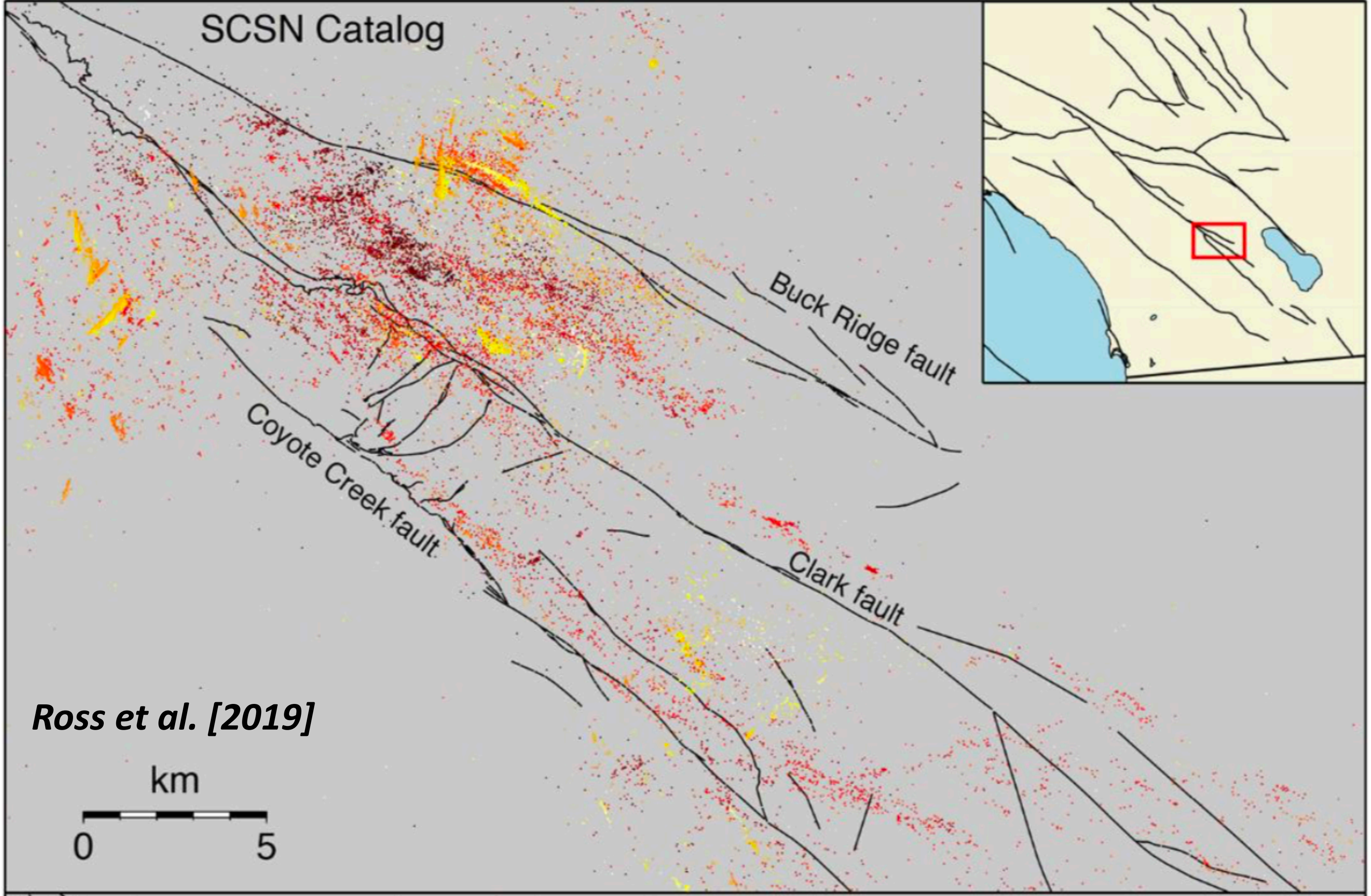
- **Presented in 1912**
- **Empirical - physical understanding was not yet available.**
- **Understood to result from stellar evolution under fusion (~25 years later)**

Comprehensive Template Matching
(all available templates against all continuous data)

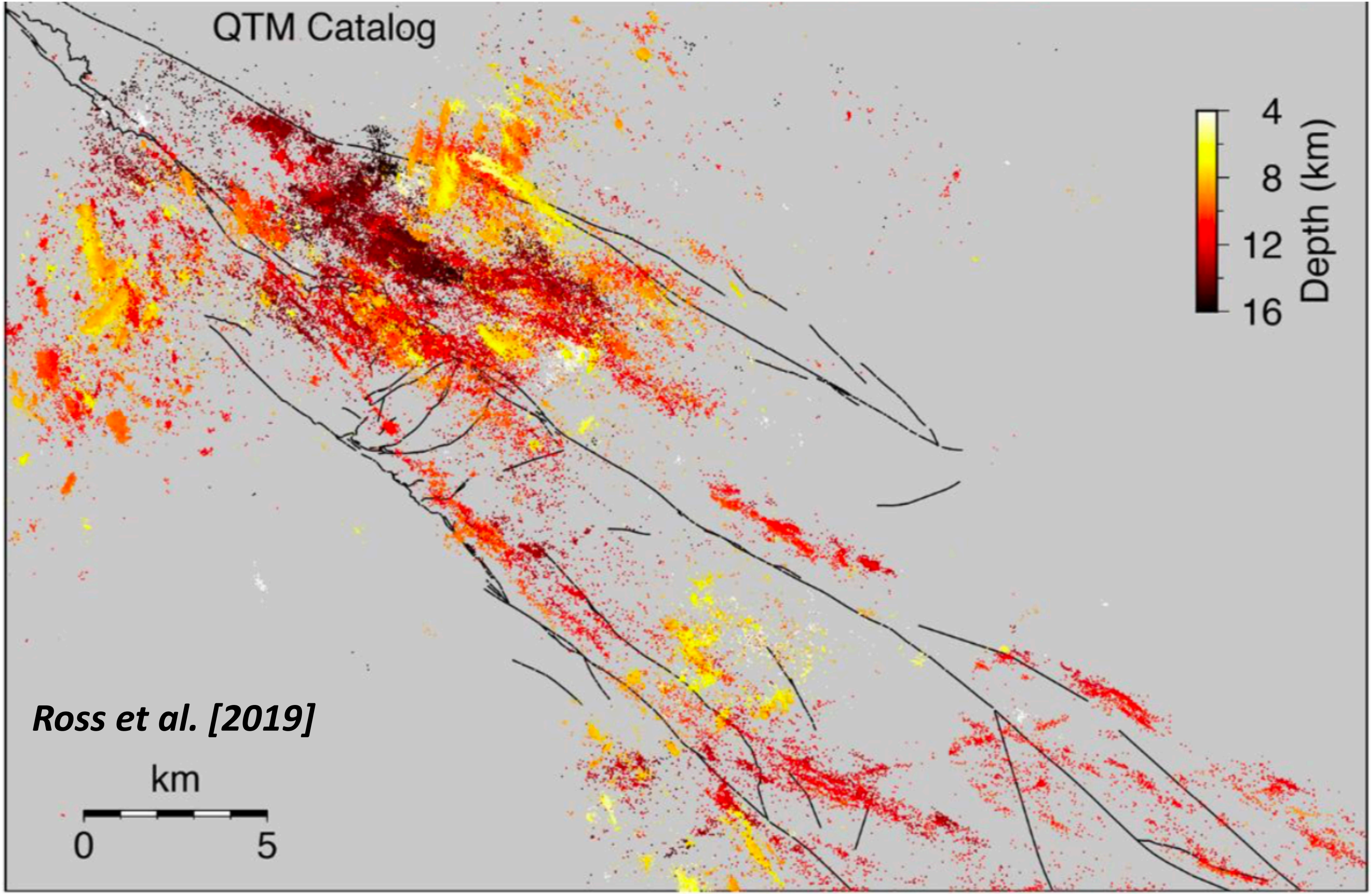


10x more earthquakes

Ross et al. [2019]



Ross et al. [2019]

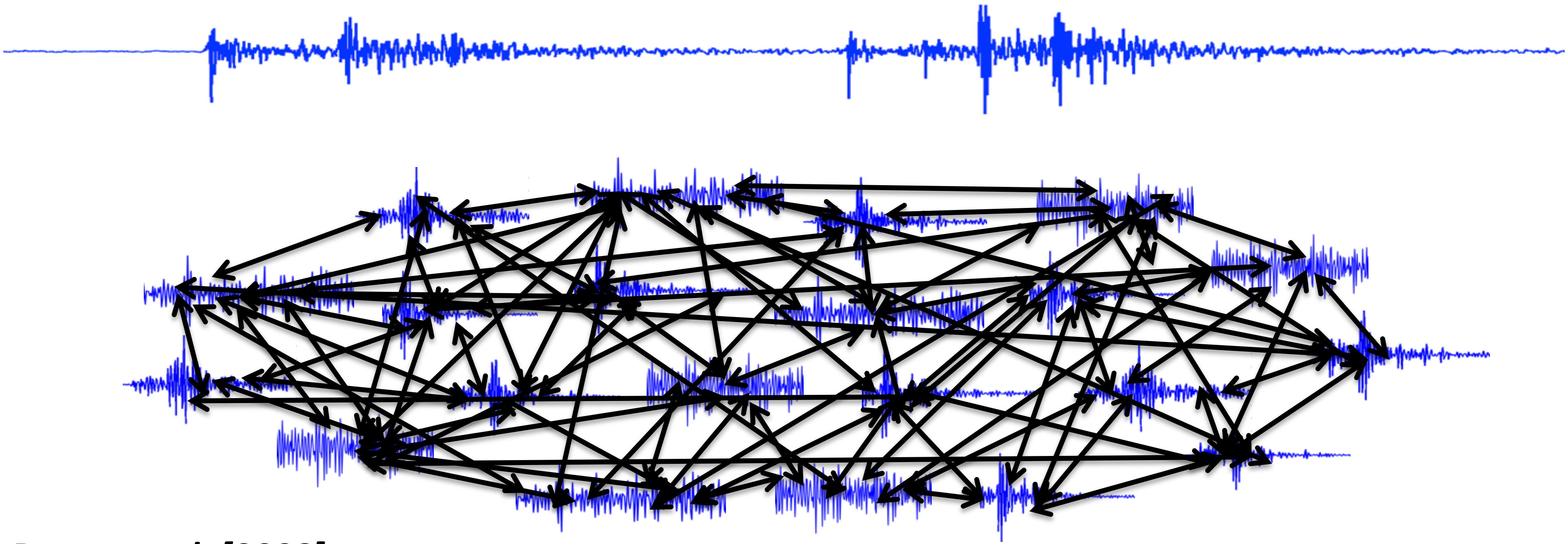


Ross et al. [2019]

What if we don’t have templates?

Naïve Search for Similar Waveforms

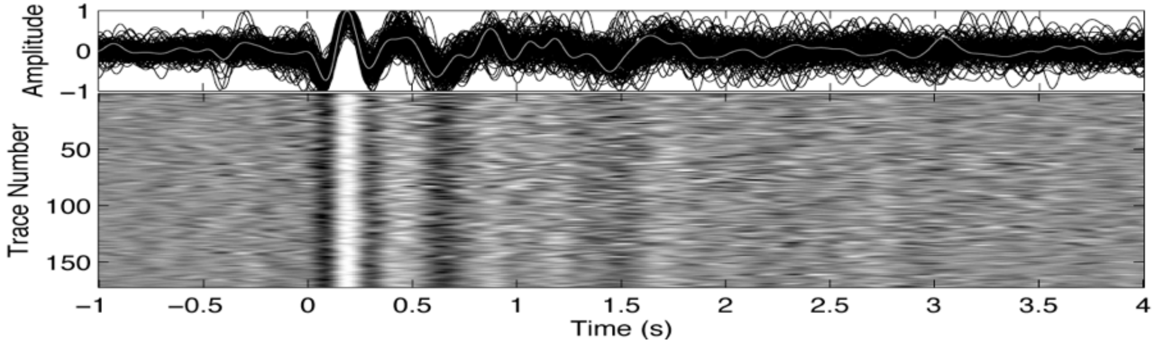
Uninformed search for similar signal – detect events by cross correlating all window pairs (all against all search).



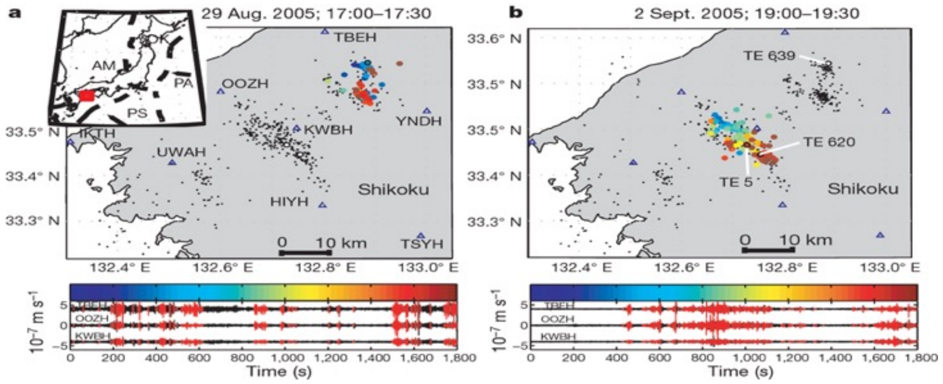
Brown et al. [2008]

Uninformed Detection by Clustered Correlation

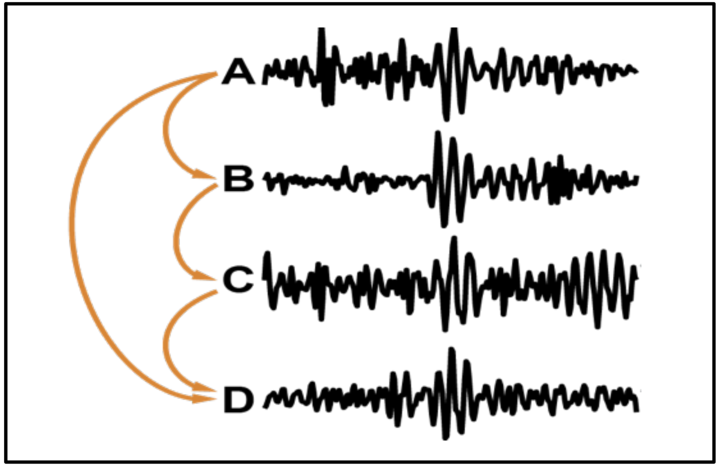
Exploits signals that repeat many times.



All-Against-All Search for Unknown Signals.



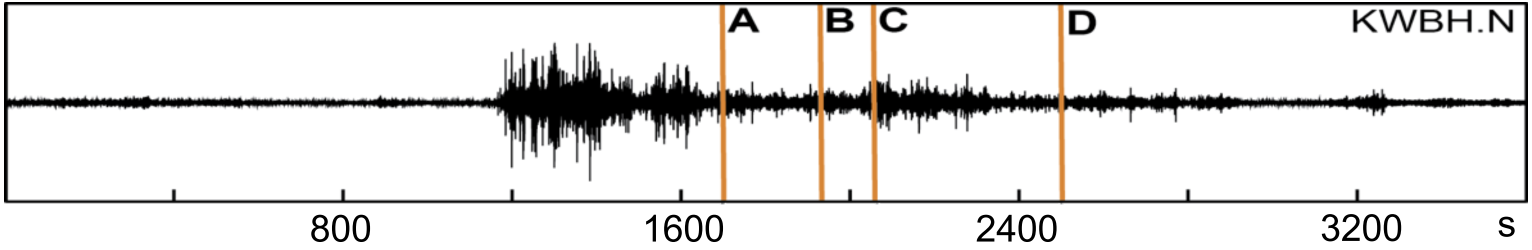
PageRank



One of Google's early ranking methods (*Page and Brin*,1988)

Uses hierarchical relationships

Shows how waveforms relate to each other



Naïve all-against-all methods are impractical on large data sets because computational effort scales quadratically with time considered.

Big Data Technologies for Approximate Similarity Search

Shazam – identify songs from a sample of recording.

Soundhound – identify songs from singing (seriously).

TinEye – Search the web for the source of a known image.

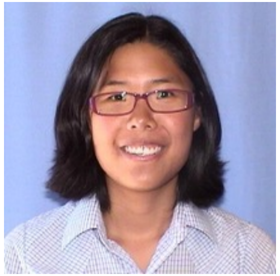
YouTube – detect copyright infringement

CopyLeaks – detect plagiarism

Altavista – remove duplicate web pages from search results

Fingerprinting and Similarity Thresholding (FAST)

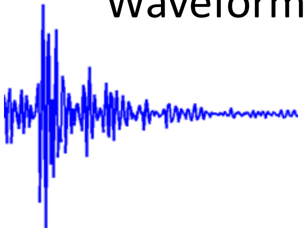
Clara Yoon



Data Compression

Fingerprint

Waveform

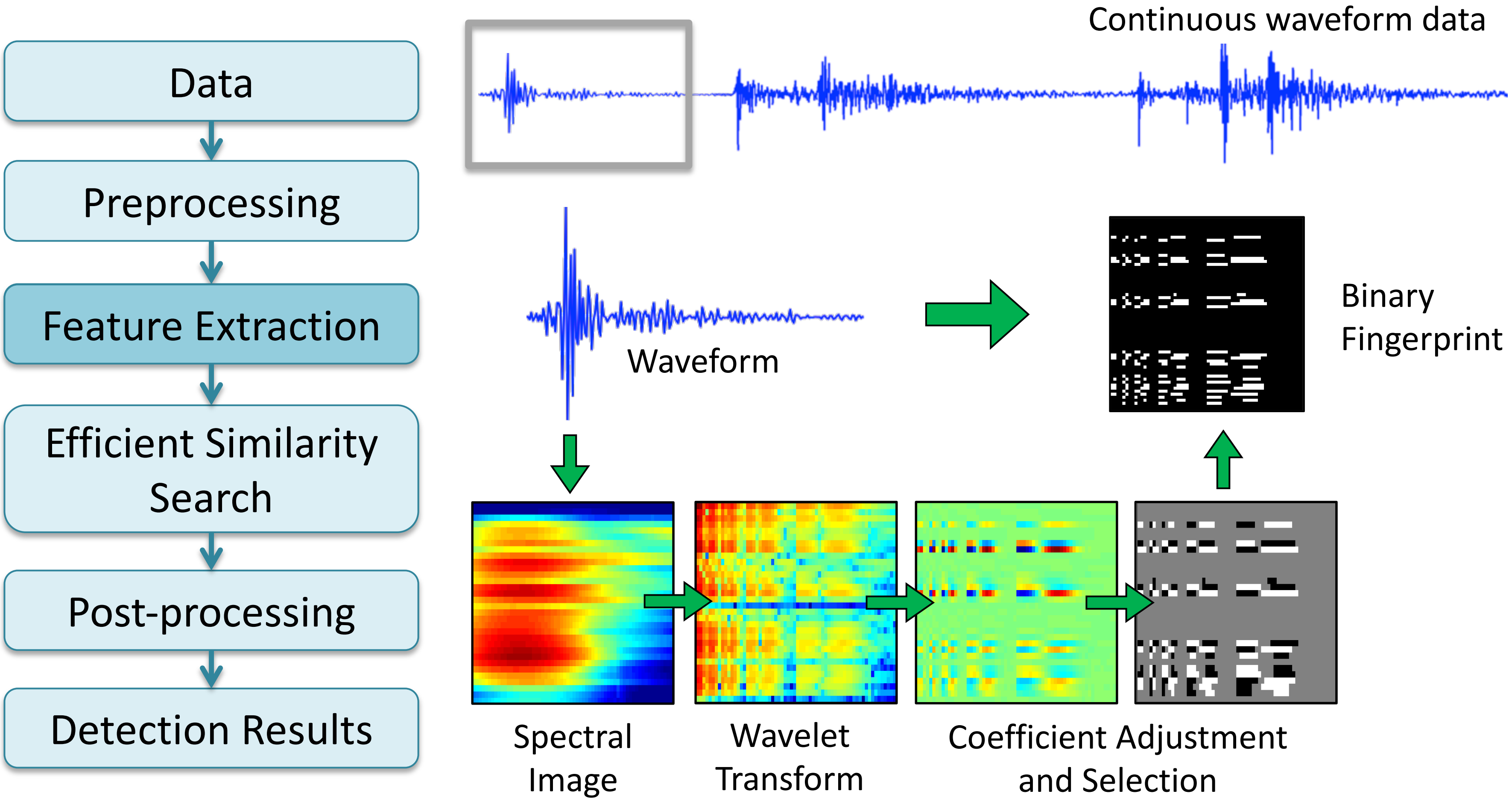


Data Compression

Binary Fingerprint

1. **"Fingerprint" waveform with sparse, diagnostic description**
2. **Store fingerprints in database and search it efficiently**

FAST Detection Pipeline

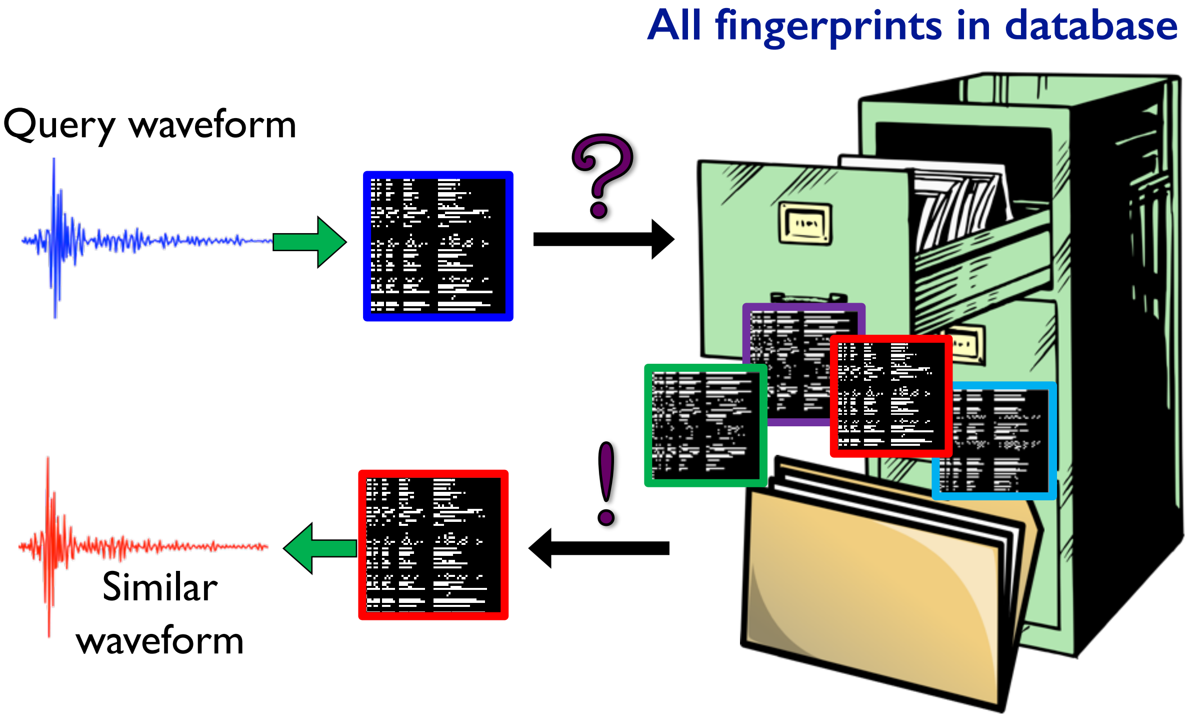


FAST Detection Pipeline

Data
↓
Preprocessing
↓
Feature Extraction
↓
Efficient Similarity Search
↓
Post-processing
↓
Detection Results

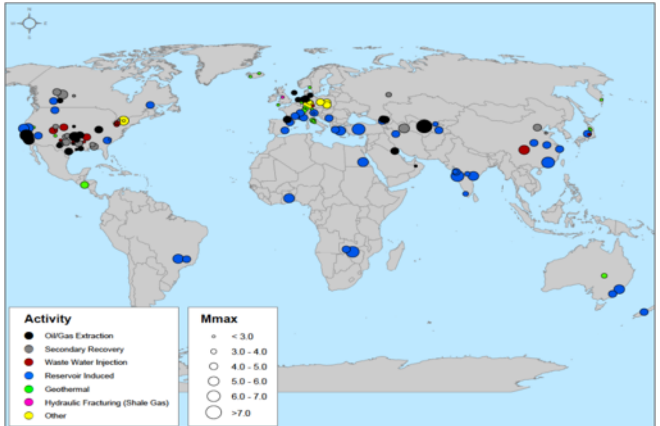
Fast approximate similarity search

- **MinHash** and **Locality Sensitive Hashing (LSH)**

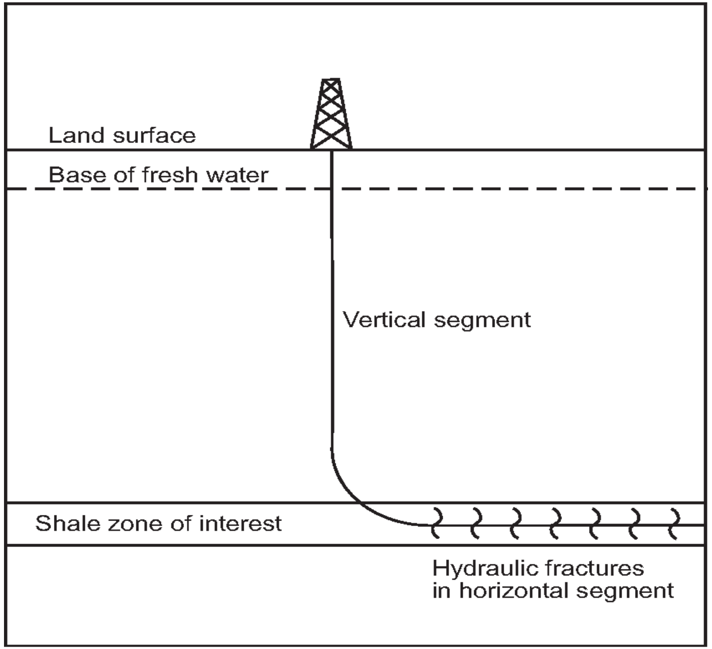


Application: Human-Induced Earthquakes

- Cogdell, Texas (injection for Enhanced Oil Recovery)
- Basel, Switzerland (injection for Enhanced Geothermal System)
- Lorca, Spain (groundwater withdrawal)
- Offshore Valencia, Spain (gas storage)
- Groningen, The Netherlands (gas production)
- Northern Germany & Southwestern Poland (mining)
- The Geysers, California (geothermal production)
- Koyna, India (reservoir induced seismicity)
- Horn River, British Columbia, Canada (hydraulic fracturing)
- Decatur, Illinois (CO2 sequestration)
- Youngstown, Ohio (wastewater disposal)

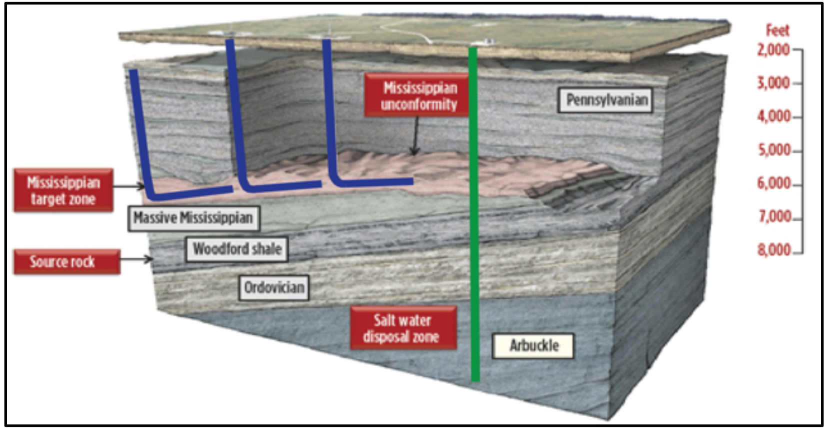


Committee on Induced Seismicity Potential in Energy Technologies, *Induced Seismicity Potential in Energy Technologies* [National Research Council] (2012).

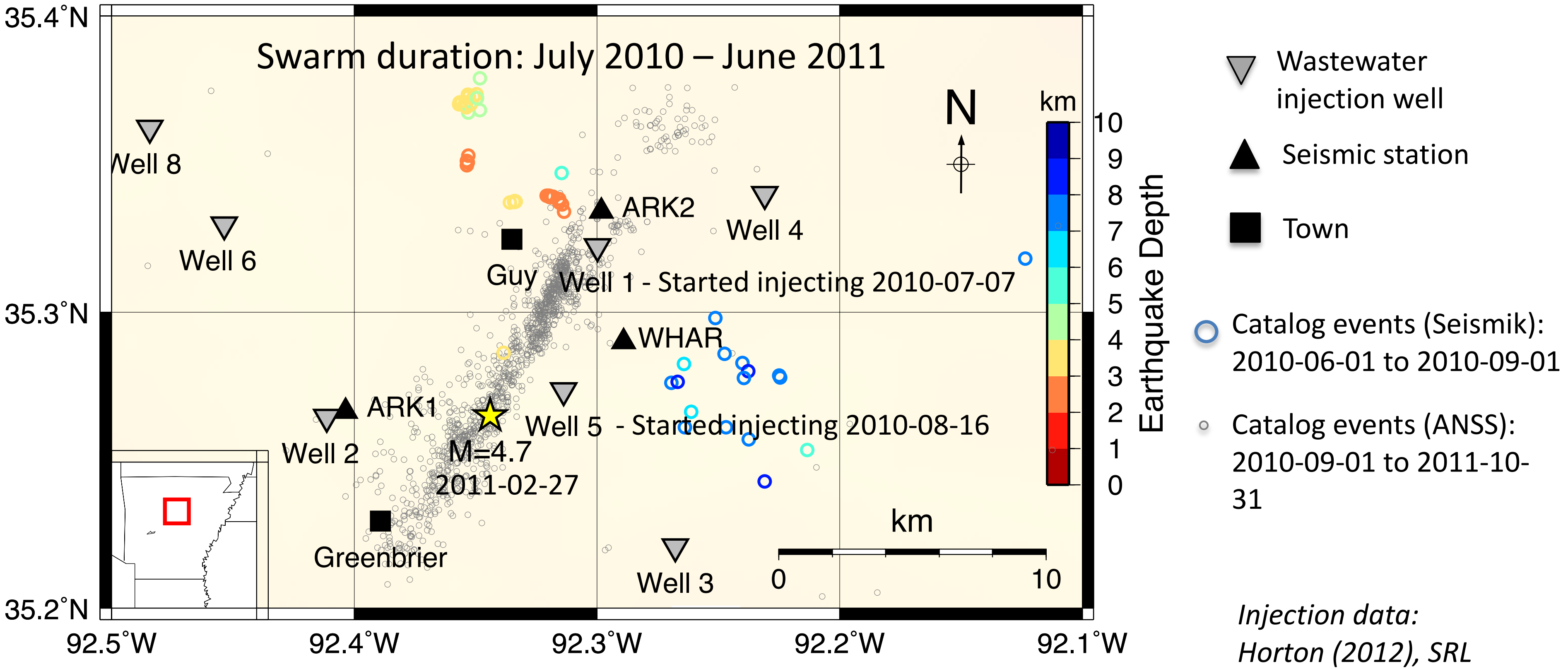


I. Hydraulic Stimulation (Fracking) uses staged injection of fluid to increase permeability and access hydrocarbons.

II. Deep disposal wells inject produced water, or fracking flowback water, to dispose of it.

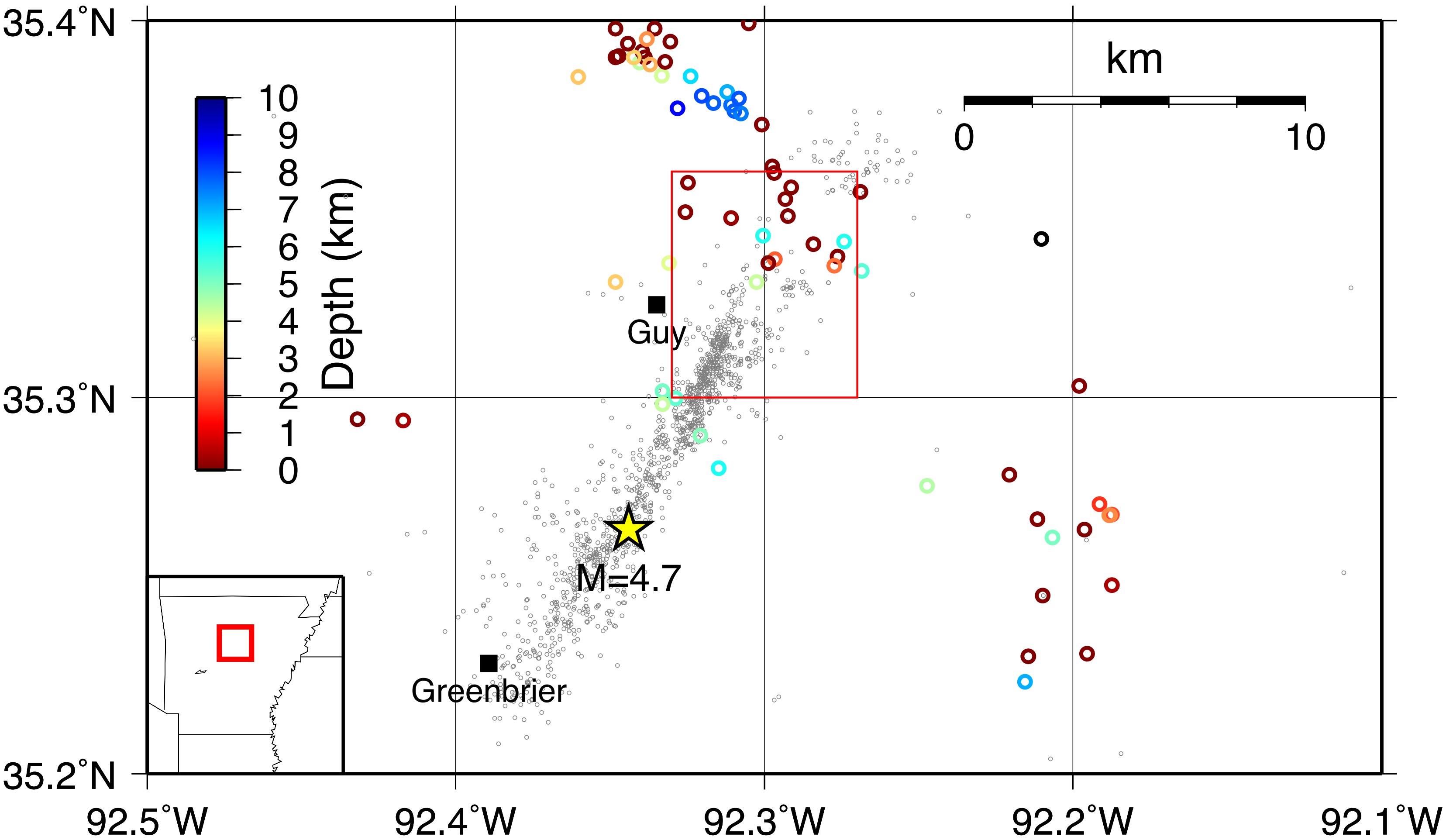


Guy-Greenbrier, Arkansas Sequence

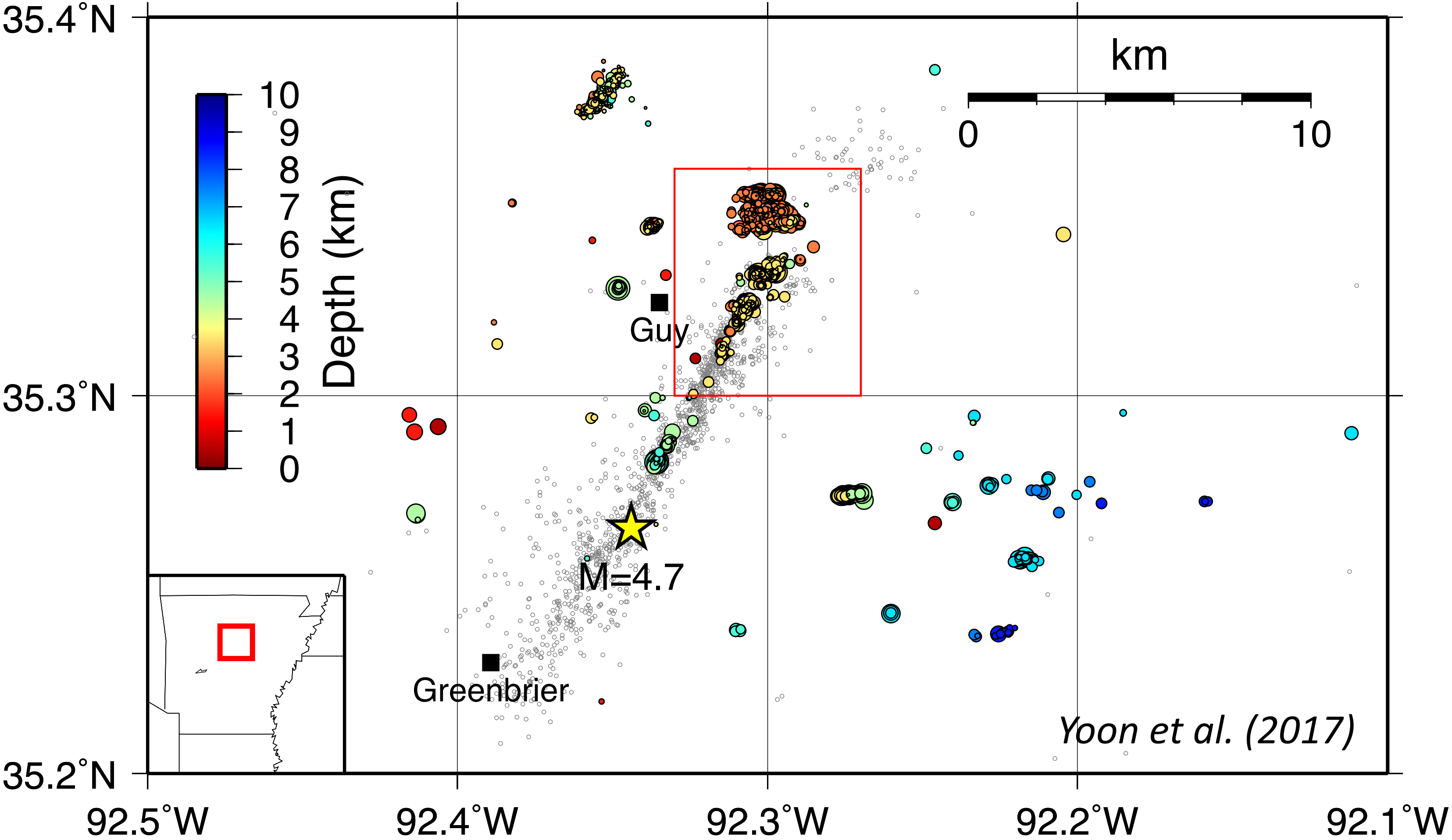


Injection data: Horton (2012), SRL

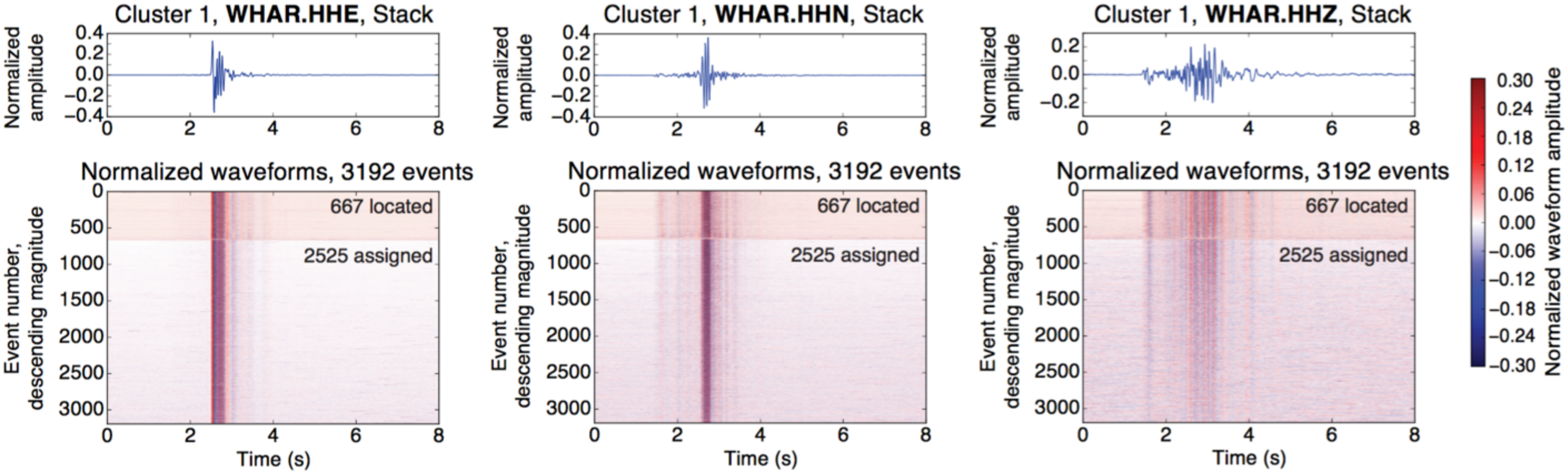
Before: 75 catalog earthquakes, $1.2 < M_L < 2.9$



After: 14,604 detected earthquakes, $-1.5 < M_L < 2.9$



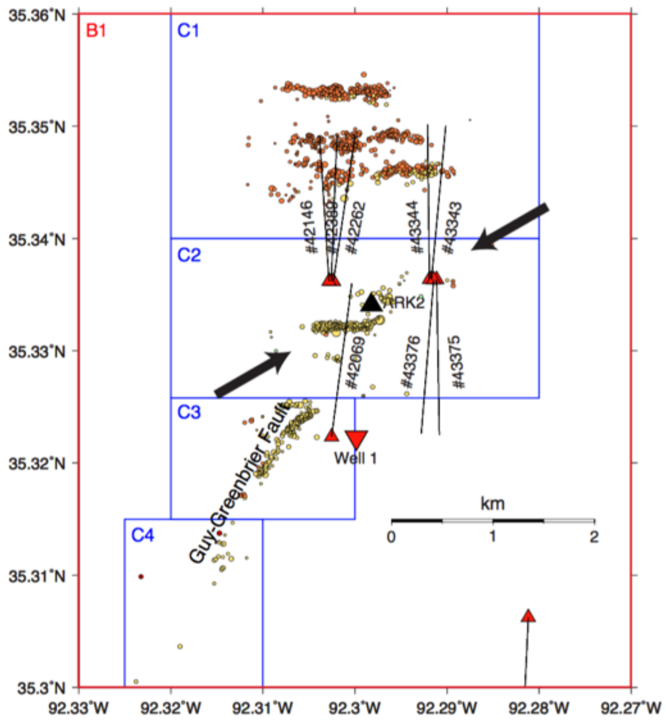
Cluster #1: 3143 events (667 located + 2525 assigned)

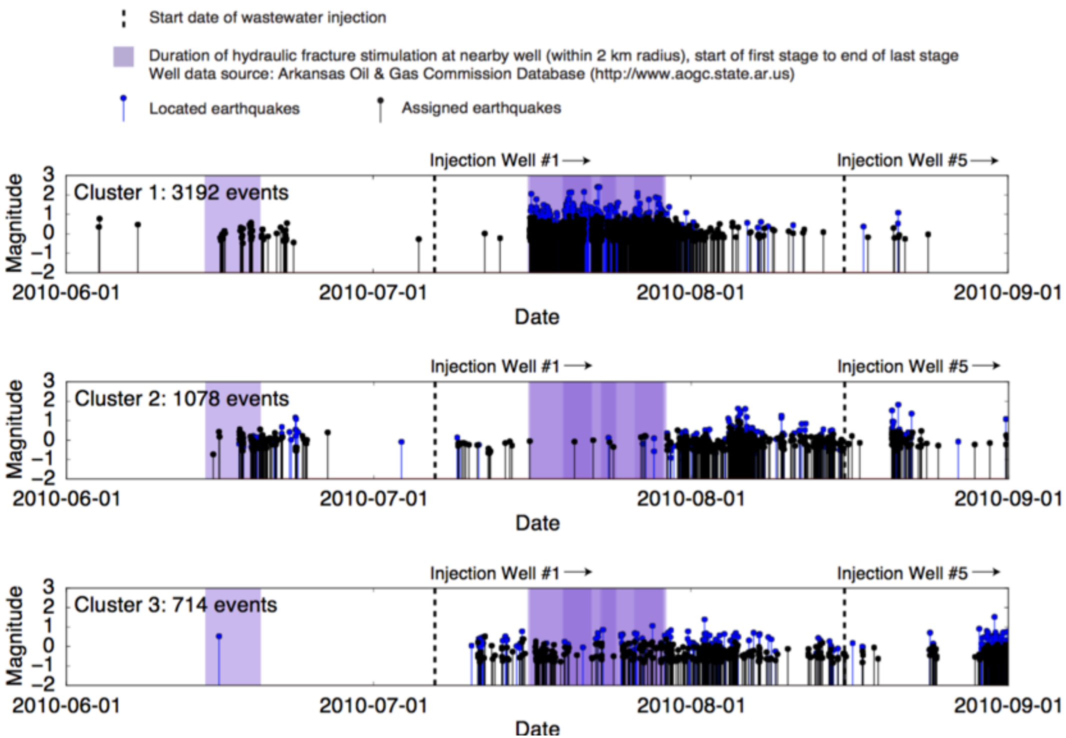


Can only locate 667 events (~20%)

2500+ associated events can’t be located, but provide insight into process

Guy-Greenbrier, Arkansas: Results





Yoon et al. (2017)

Evolution of similarity Search

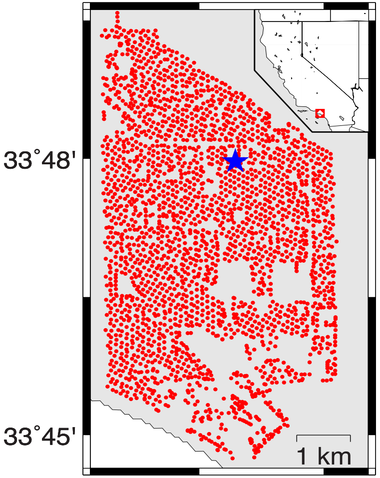
Informed Search: Template matching or subspace projection of known event waveforms.

Uninformed Search: Discovery of templates through naïve correlation, Pagerank clustering, or approximate search by LSH.

Generalized Similarity Search: Generalization of strict similarity search to more permissive similarity in characteristics using machine learning.

Recent developments → New opportunities

- Massive seismic data sets
- New ML algorithms and models
- Improvements in computing technology



Nakata et al. (2015)

Long Duration (Large-T)

>10 years continuous waveform data

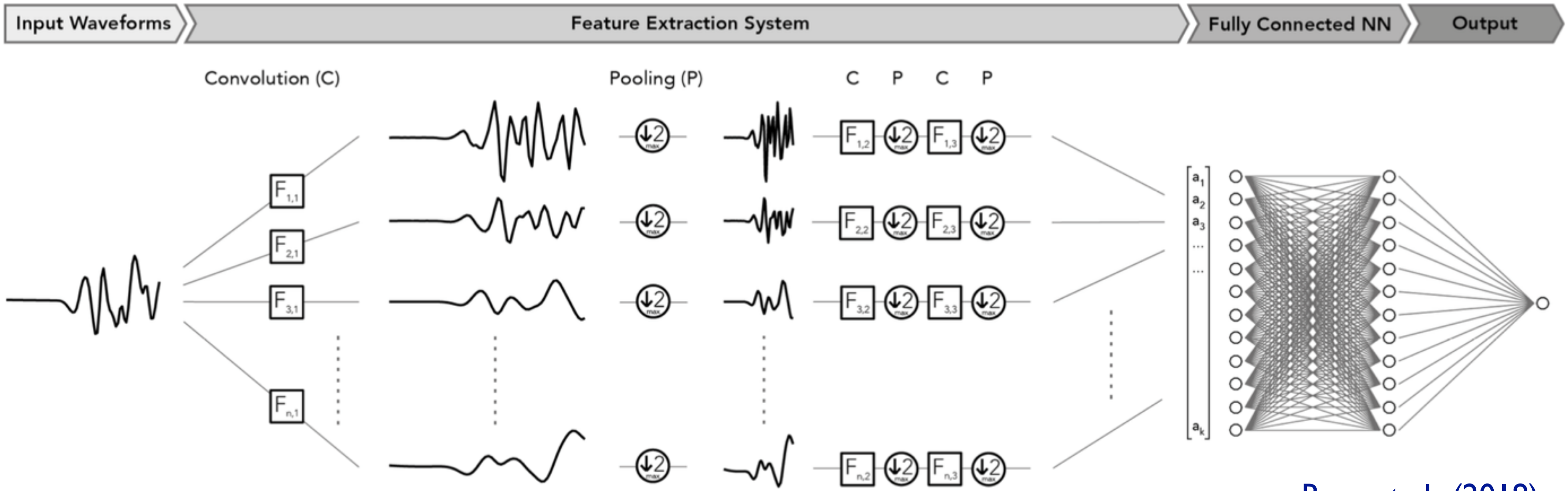
Big Networks (Large-N)

1000's of sensors

New Data Sources

Recent developments → New opportunities

- Massive seismic data sets
- New ML algorithms and models
- Improvements in computing technology



Ross et al., (2018)

Recent developments → New opportunities

- Massive seismic data sets
- New ML algorithms and models
- Improvements in computing technology

GPU Computing

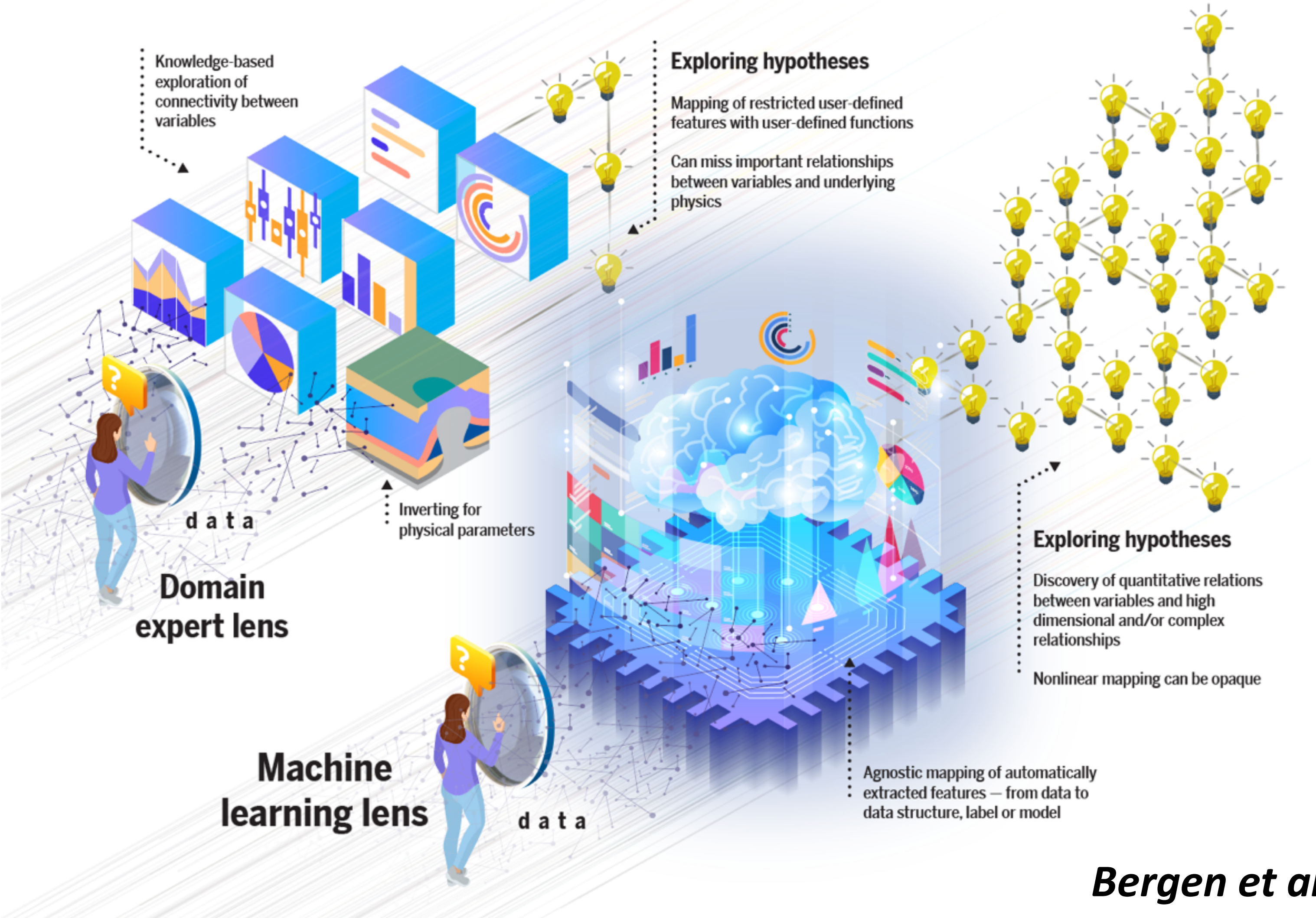
Open Source Tools

TensorFlow

Caffe2

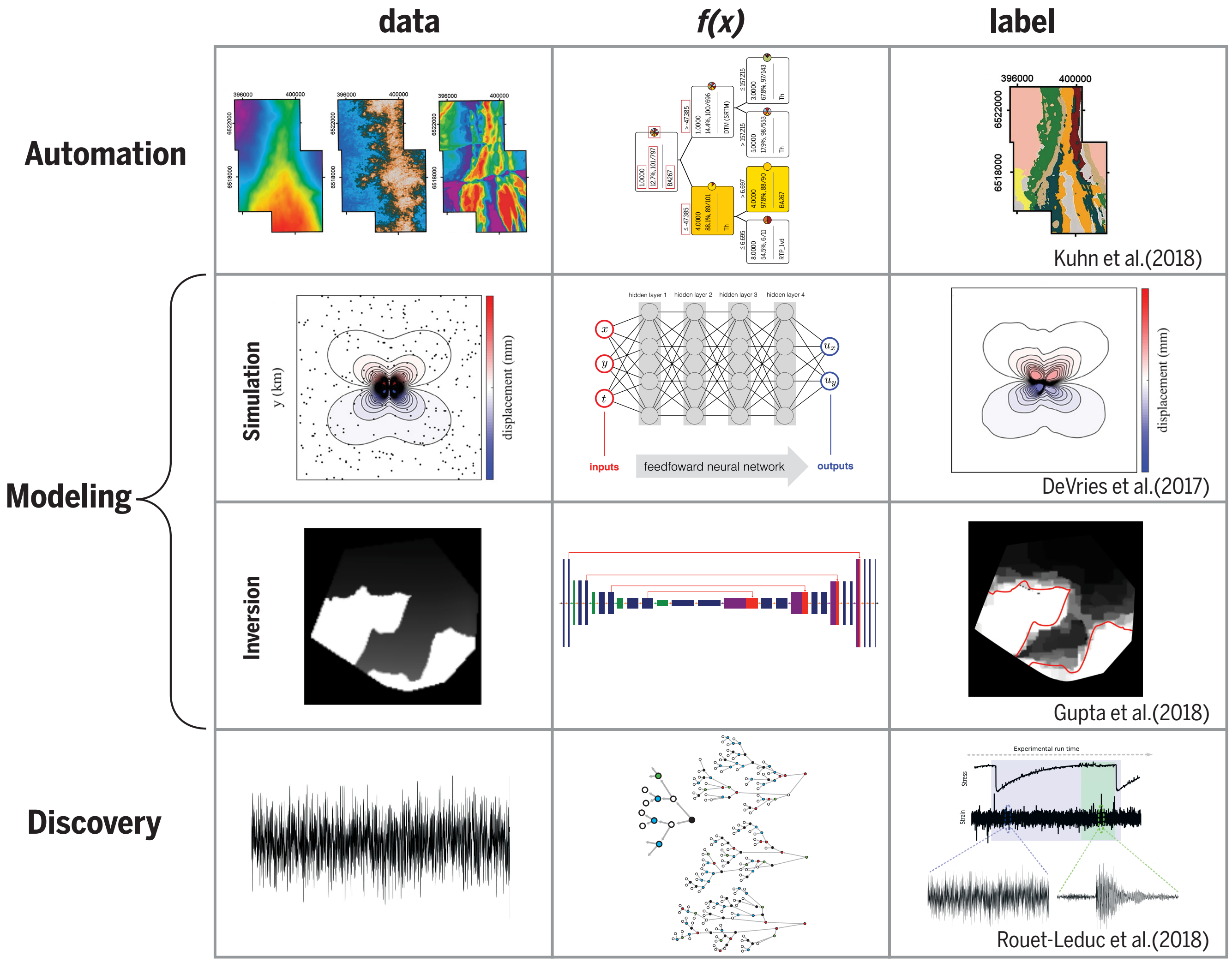
Keras

PYTORCH



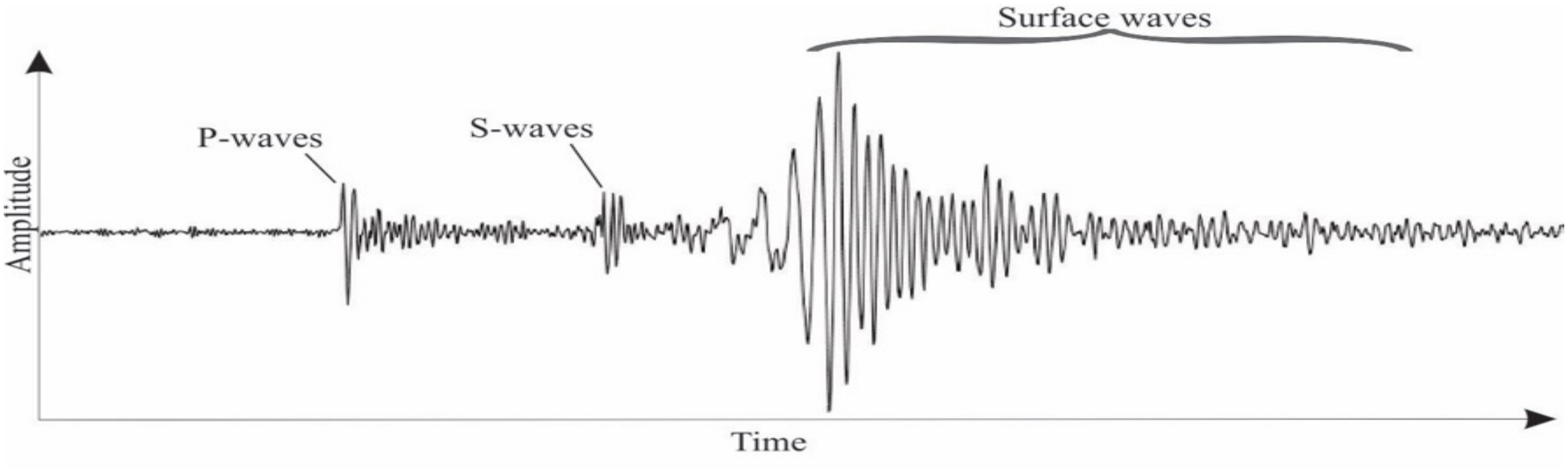
Bergen et al. (2019)

Modes of AI



***Bergen et al.* (2019)**

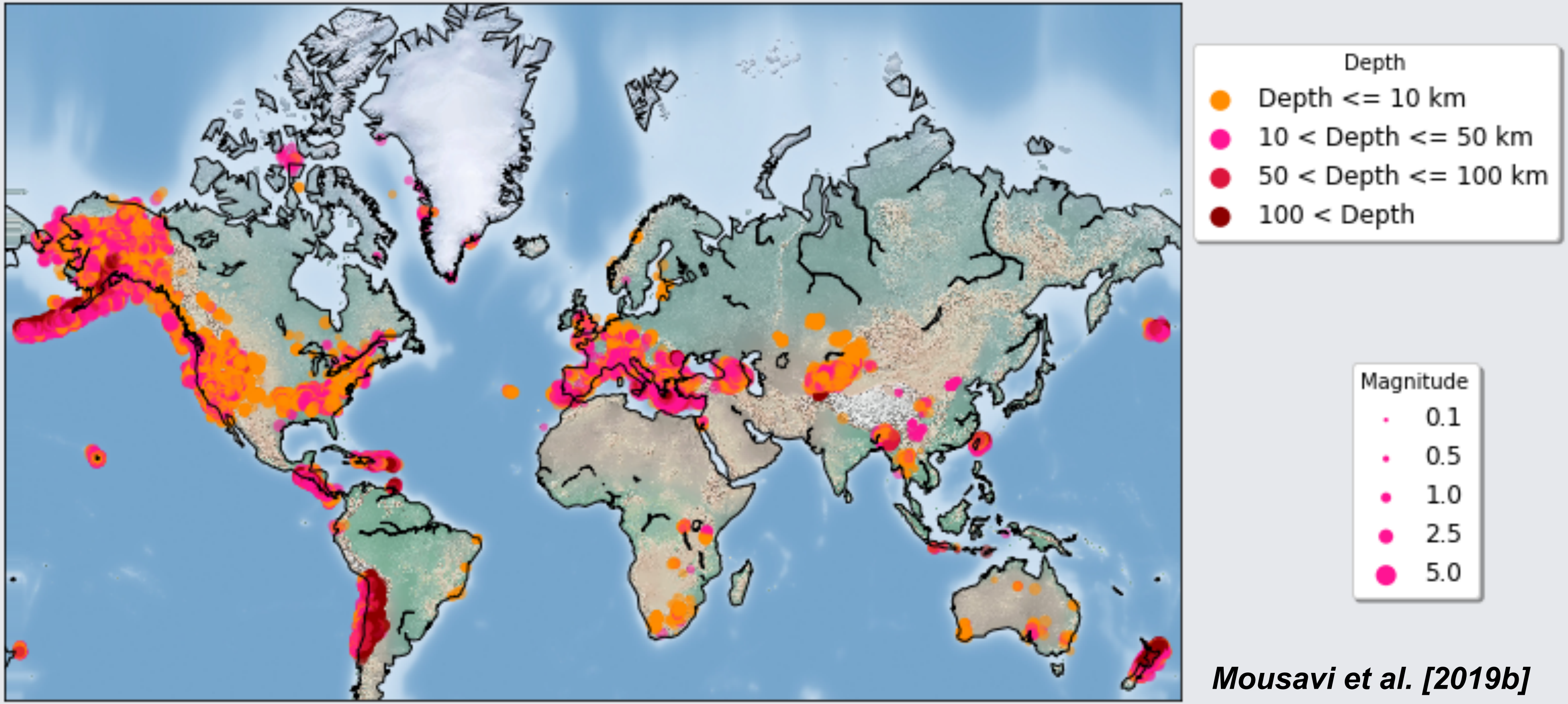
Seismology has Large Labeled Datasets, but…



…the labels are far from perfect.

Curated Data Set

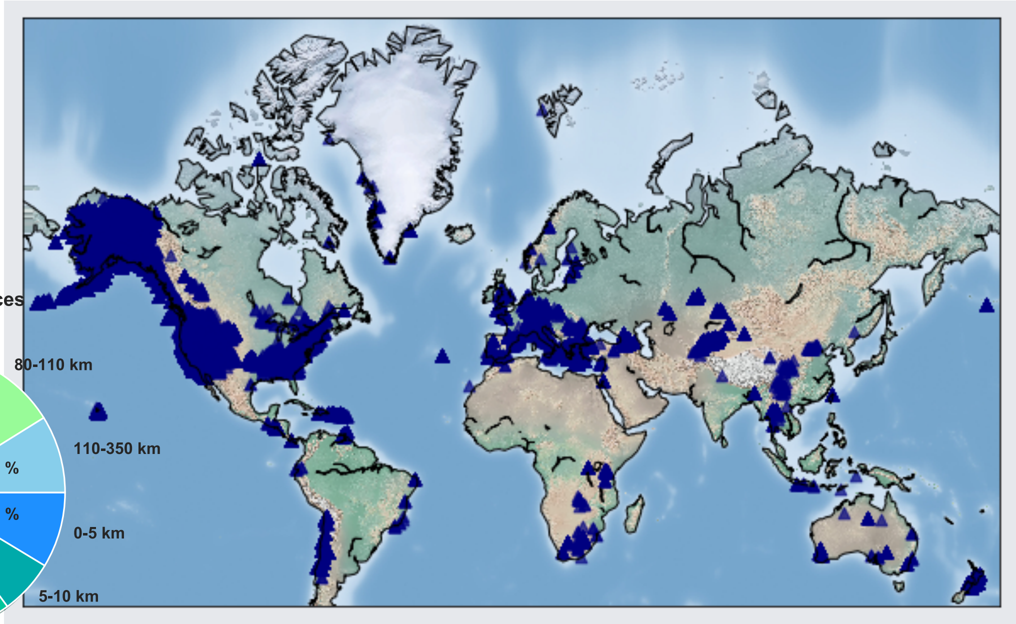
1.2 M seismograms. 500k earthquakes. *STanford Earthquake Dataset (STEAD)*

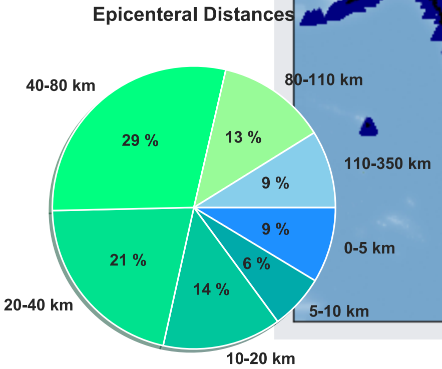


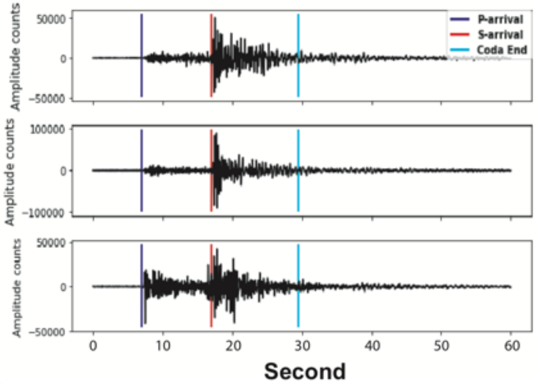
Mousavi et al. [2019b]

STEAD

2,650 seismometers. **Local distances only**







b)

{'back_azimuth_deg': 141.5,
'category': 'earthquake',
'coda_end_sample': 2951,
'earthquake_depth_km': 13.9,
'earthquake_distance_deg': 0.7026,
'earthquake_distance_km': 77.99,
'earthquake_id': 'ci37917624',
'earthquake_lat': 34.001,
'earthquake_lon': -116.9018,
'earthquake_magnitude': 3.5,
'earthquake_magnitude_type': 'ml',
'earthquake_origin_time': '2017-06-25 13:53:24.760000',
'file_name': 'ADO.CI_20170625135330_EV',
'instrument_type': 'HH',
'network_code': 'CI',
'p_arrival_time': '2017-06-25 13:53:37.650',
'p_status': 'manual',
'p_travel_sec': 12.89,
'p_weight': 0.46,
's_arrival_time': '2017-06-25 13:53:47.650',
's_status': 'manual',
's_weight': 0.5125,
'snr_db': [39.3, 39.5, 36.5],
'station_elevation_m': 908.0,
'station_lat': 34.55046,
'station_lon': -117.43391,
'station_name': 'ADO',
'trace_start_time': '2017-06-25 13:53:30.650000'}

Both signals and noise

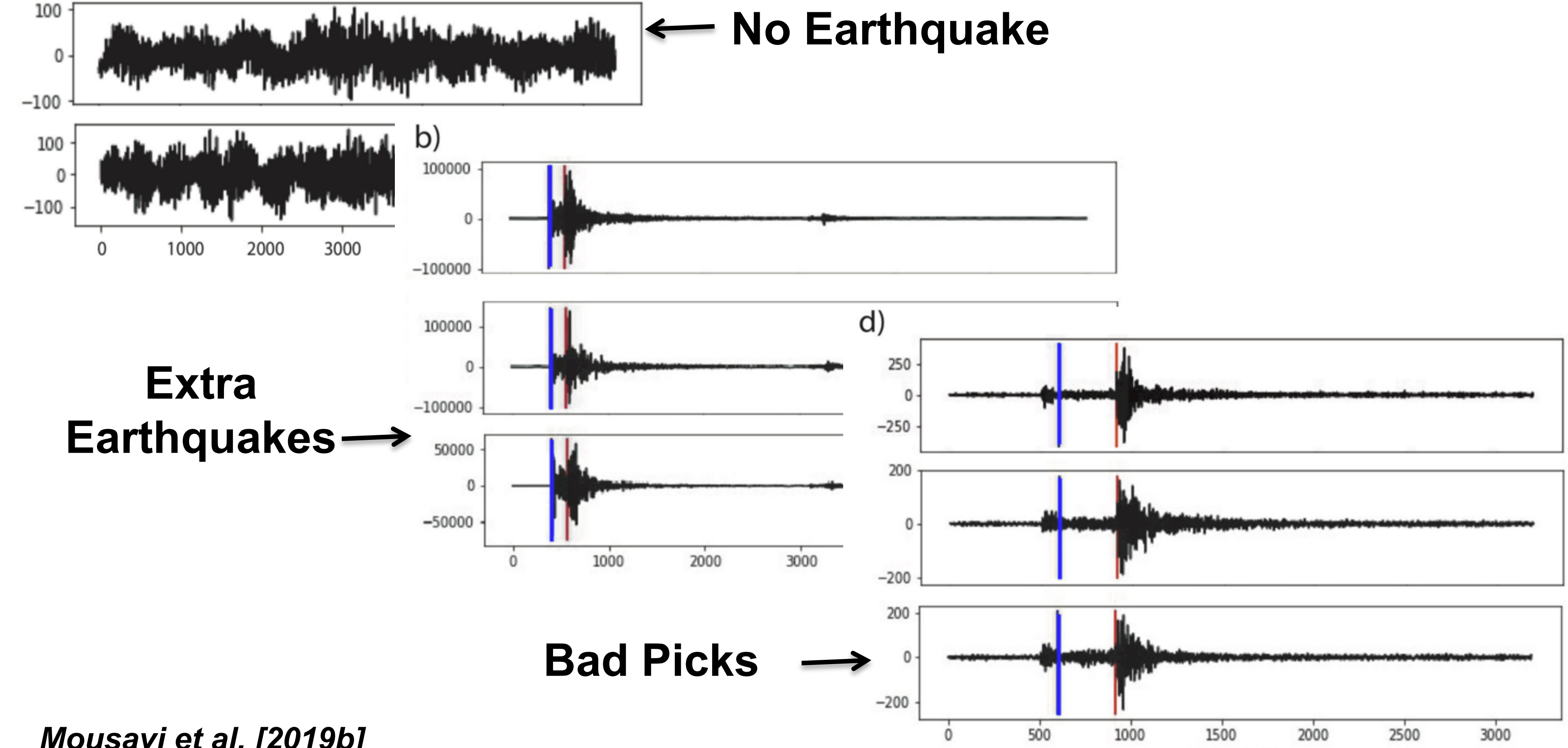
Extensive quality control

Additional labels

HDF5 format for easy I/O

Mousavi et al. [2019b]

Some Data Issues



Mousavi et al. [2019b]

STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI

S. MOSTAFA MOUSAVI[1], YIXIAO SHENG[1], WEIQIANG ZHU[1], and GREGORY C. BEROZA[1]
[1]Geophysics Department, Stanford University, 397 Panama Mall, Stanford, 94305-2215, CA, United States (e-mail: mmousavi@stanford.edu)

Published in *IEEE Access*

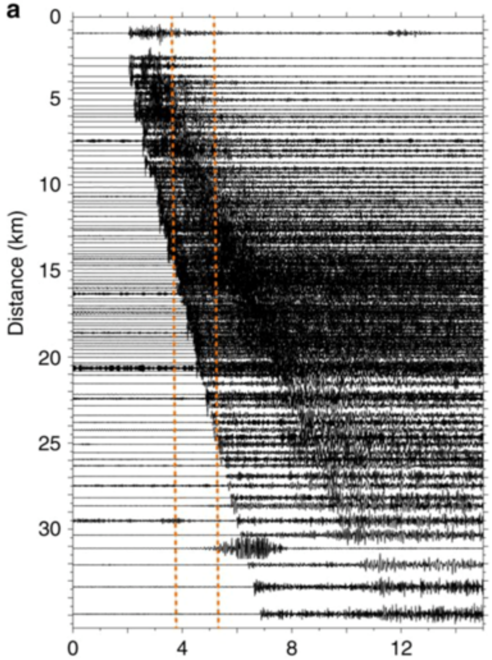
Data-science-friendly data format

Seismology 101 (get data scientists interested in our research)

First release has only local earthquakes (Δ < 350 km)

Earthquake Monitoring Workflow

Input: continuous time series of ground motion from a seismic network



Pre-processing

↓

Event Detection ← **CRED**

↓

Phase Picking

↓

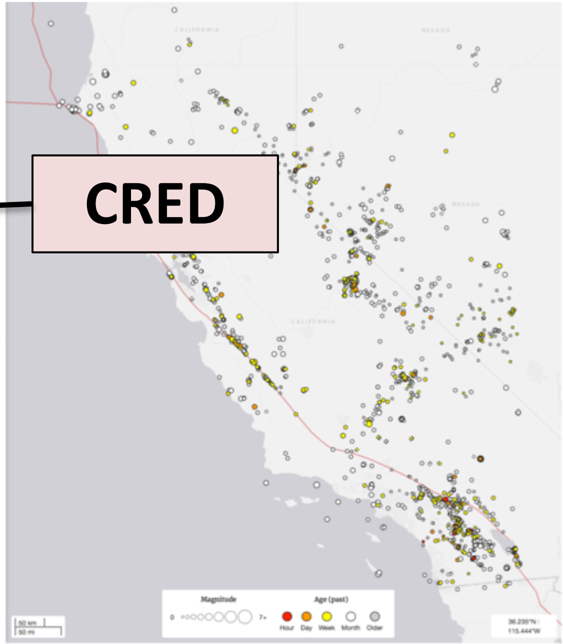
Association

↓

Location

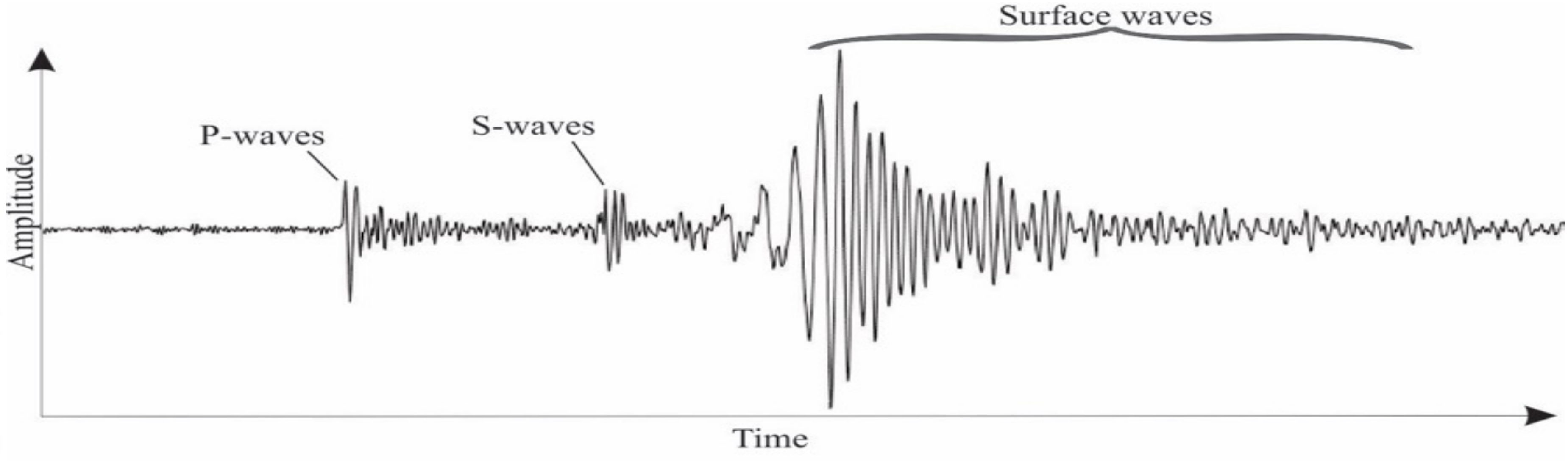
↓

Characterization



Output: comprehensive catalog of discrete seismic events

Convolutional Recurrent Earthquake Detector (CRED)

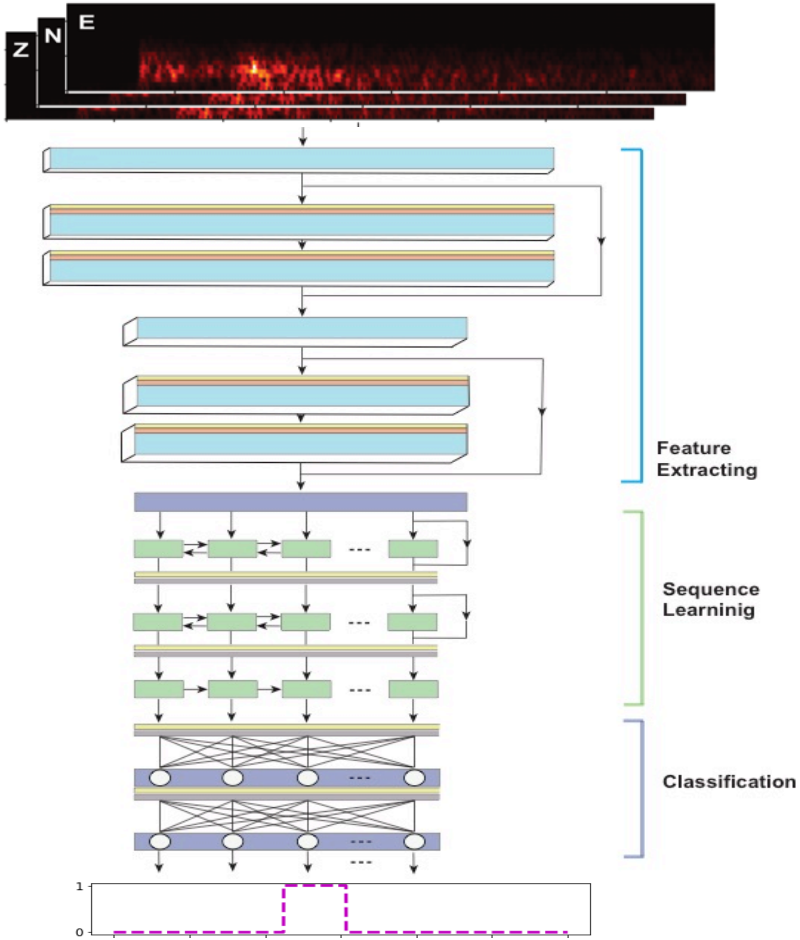


Convolutional layers to capture features: motion from different kinds of waves

Recurrent layers to capture sequence: P wave then S wave then Rayleigh wave

***Mousavi et al.* (2019)**

Convolutional-Recurrent Earthquake Detector



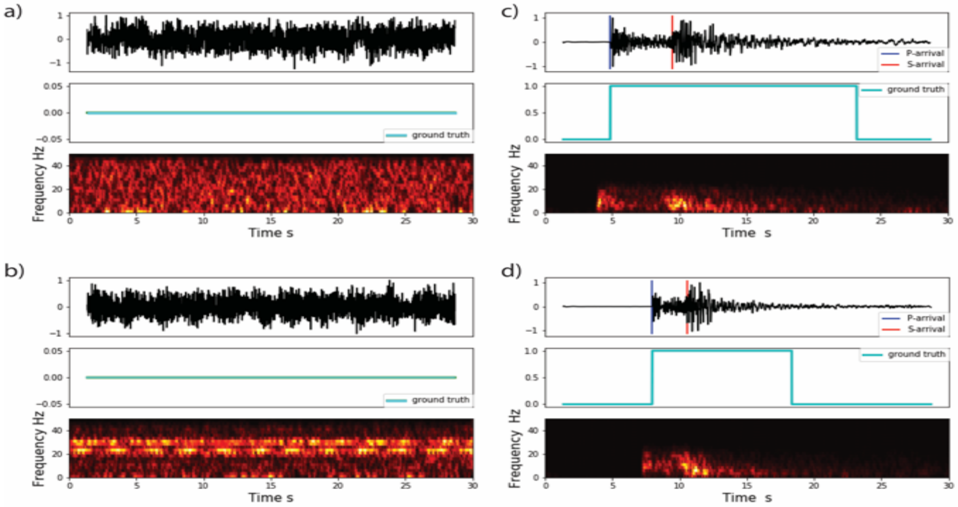
Mousavi et al. (2019a)

12 layers; 256,000 trainable parameters

550,000 3-component 30 s NCSN seismograms

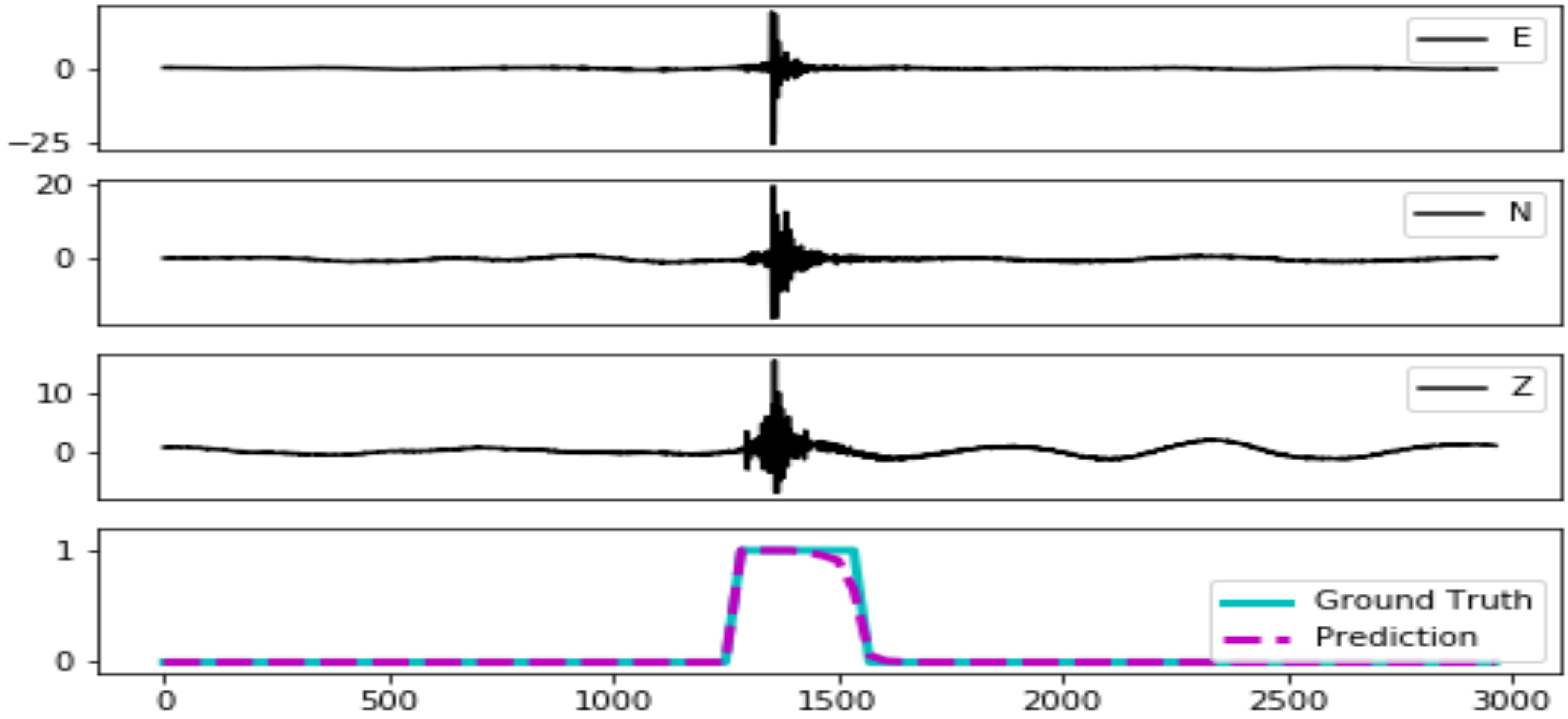
50% are earthquakes and 50% are noise

500,000 for training & validation + 50,000 for testing



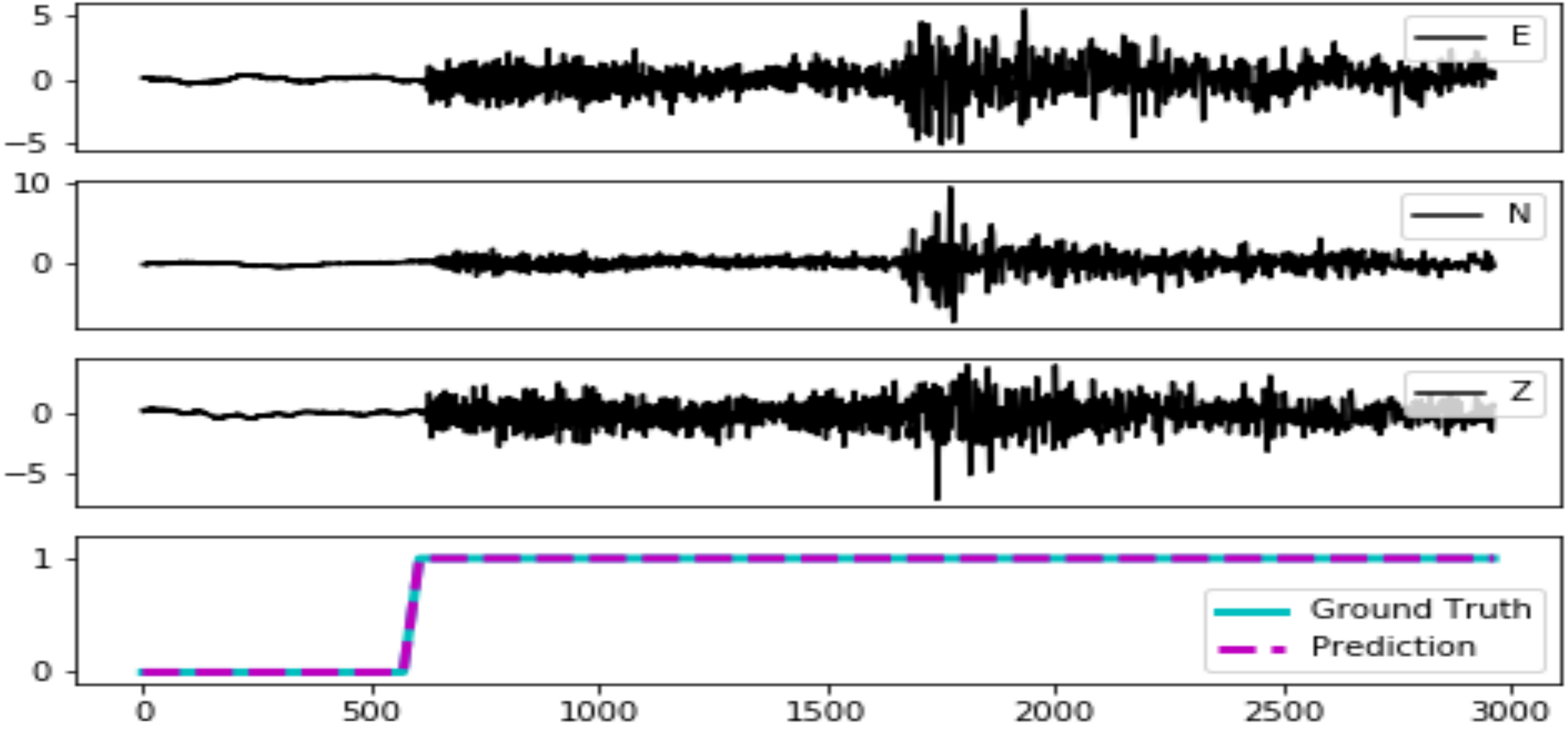
P + 3(S-P)

Works for local earthquakes



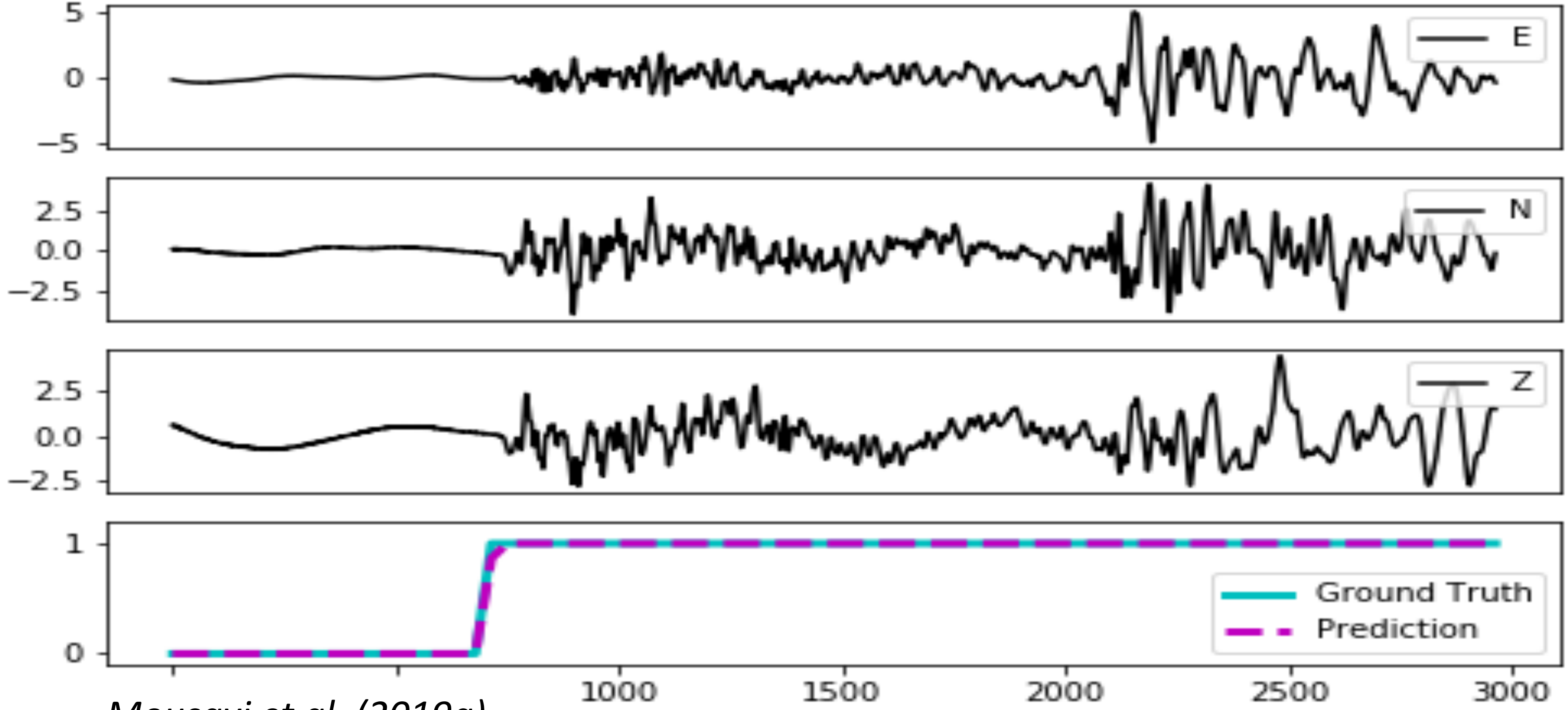
Mousavi et al. (2019a)

Works for regional earthquakes



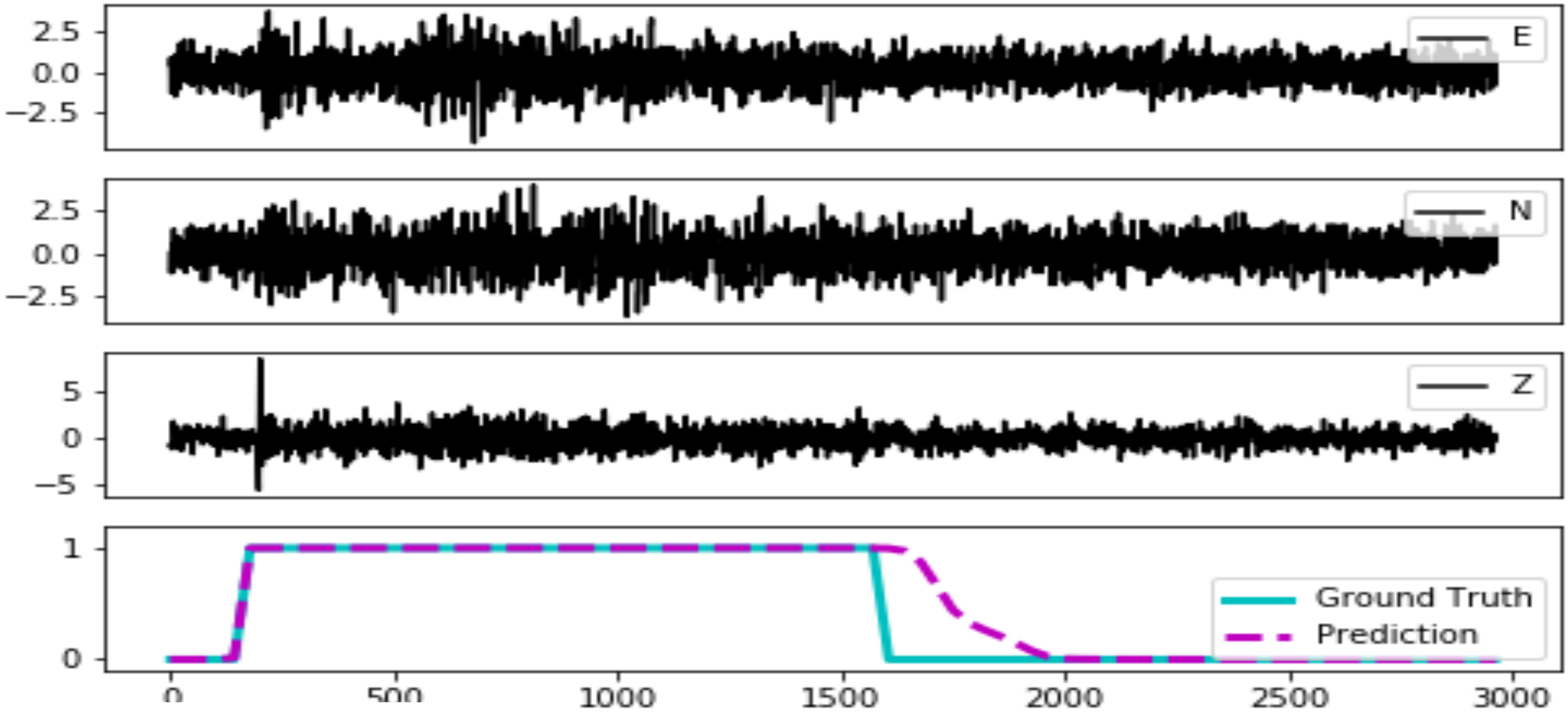
Mousavi et al. (2019a)

Works for teleseisms



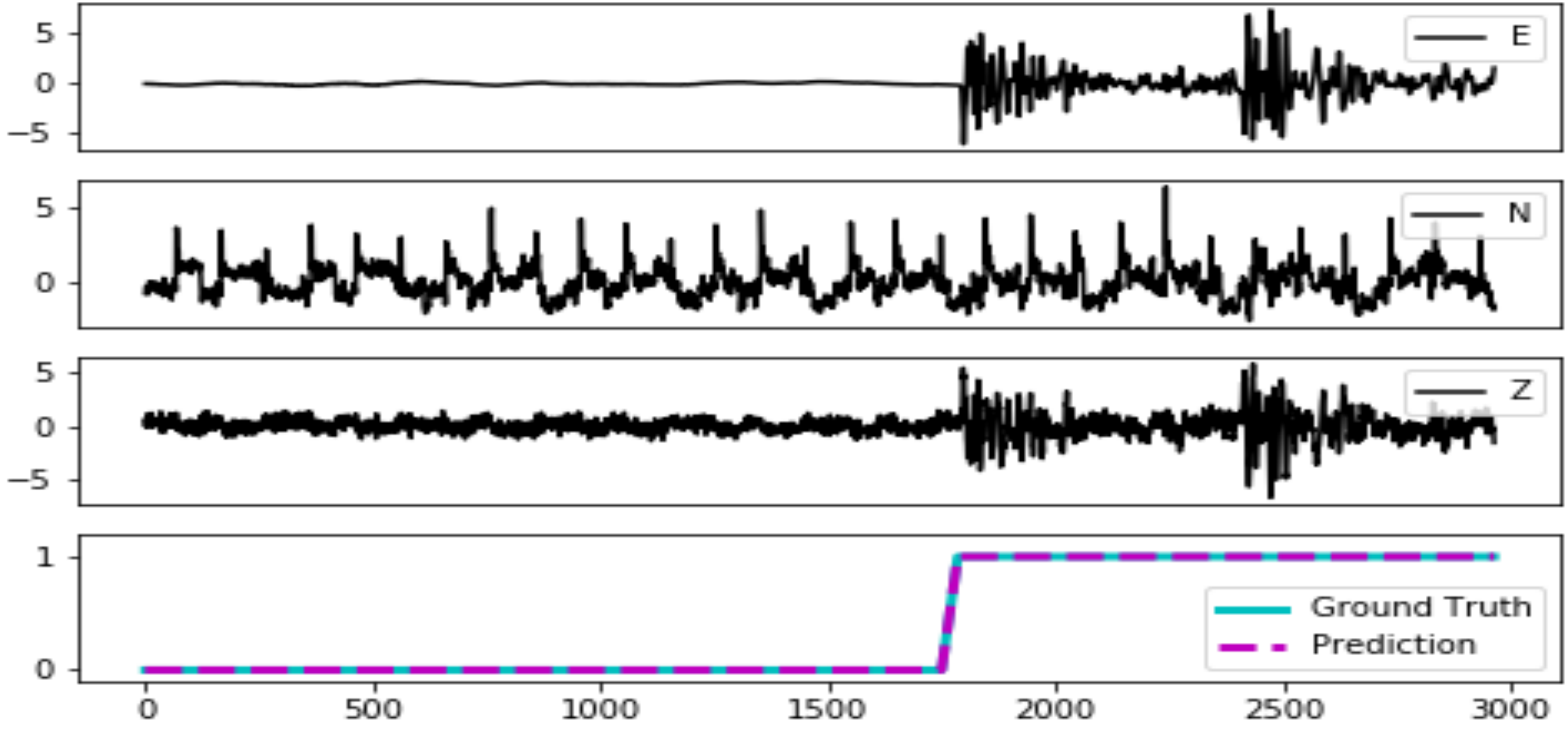
Mousavi et al. (2019a)

Works at low snr



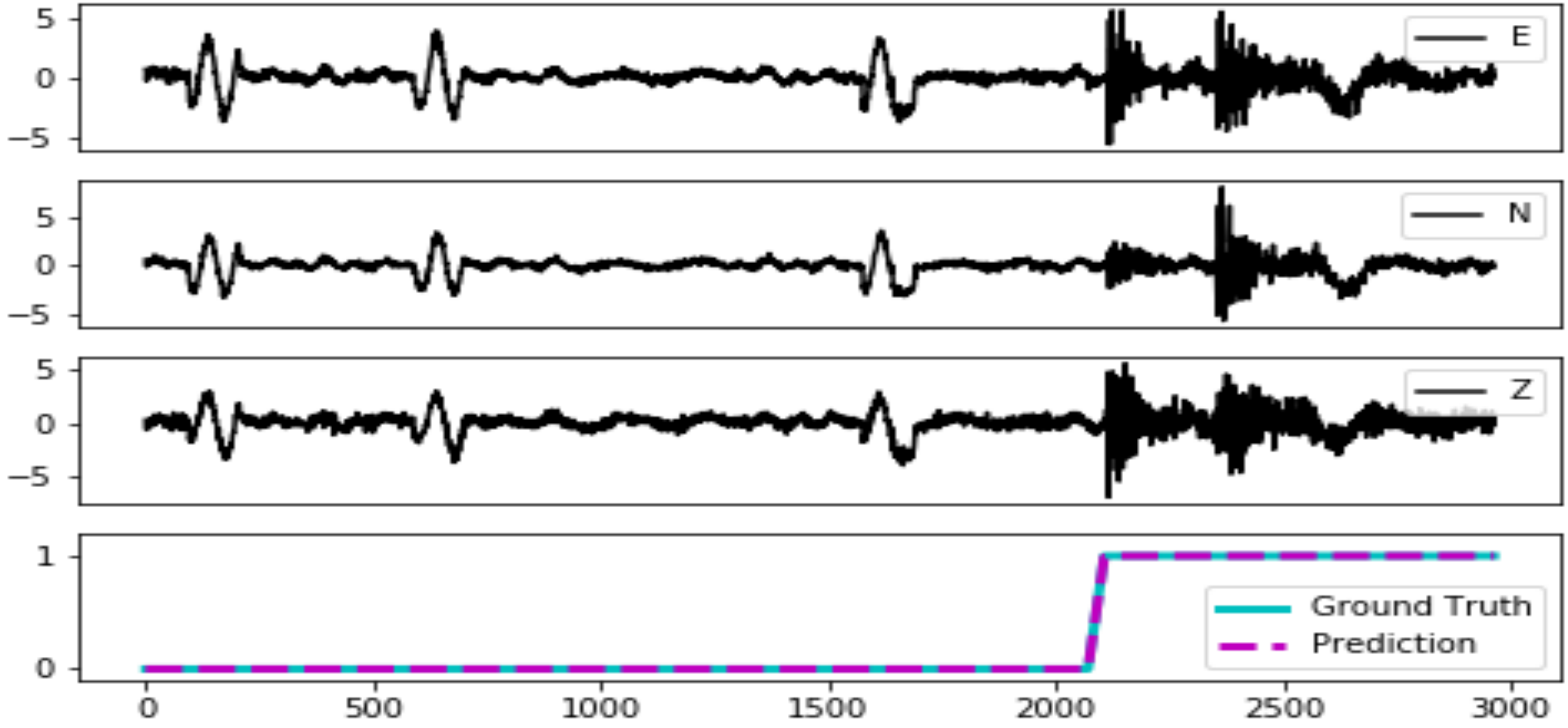
Mousavi et al. (2019a)

Works with a bad channel



Mousavi et al. (2019a)

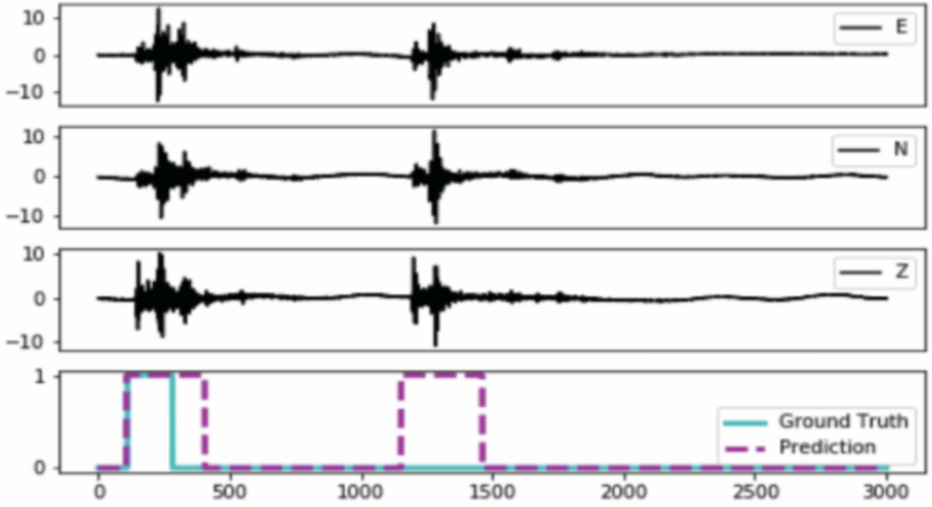
Works in presence of noise



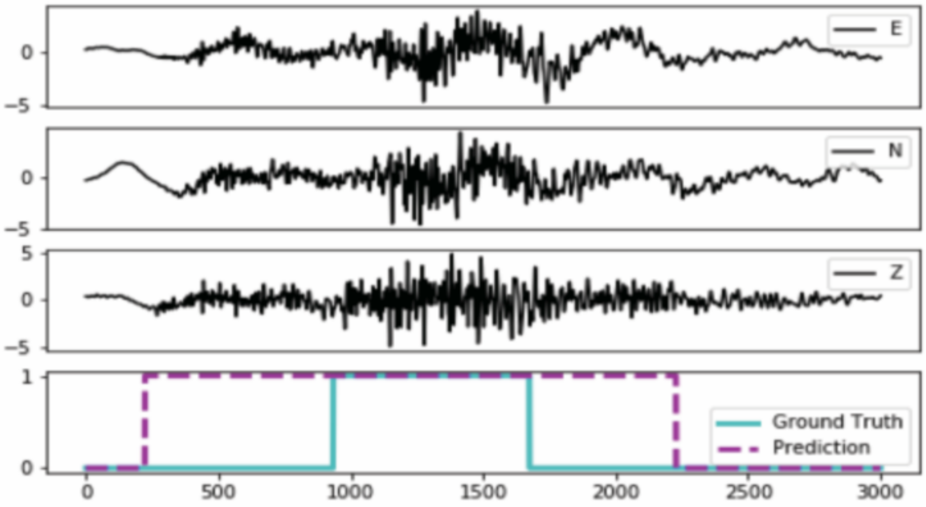
Mousavi et al. (2019a)

Improves on “ground truth”

Multiple Events



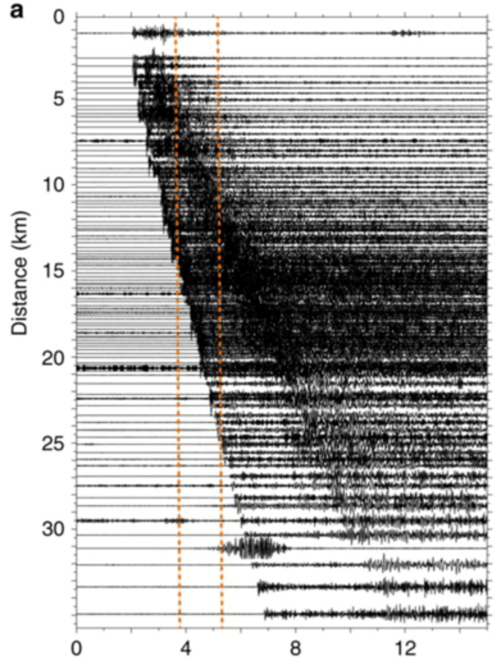
Improved Duration



Mousavi et al. (2019a)

Earthquake Monitoring Workflow

<u>Input:</u> continuous time series of ground motion from a seismic network



Pre-processing

Event Detection

Phase Picking

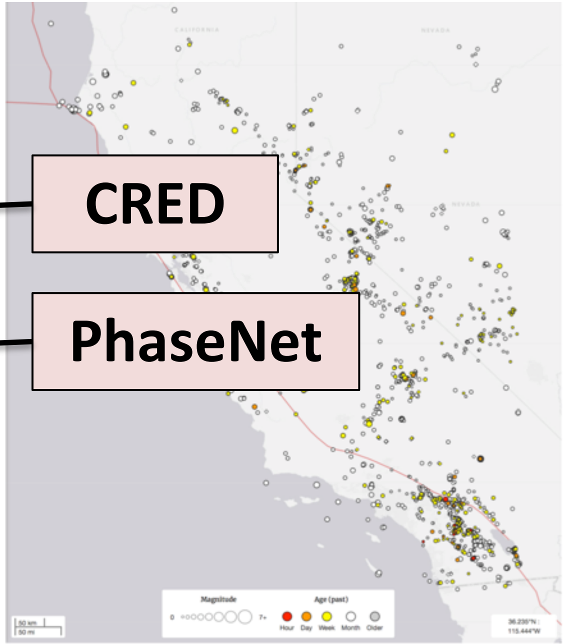
Association

Location

Characterization

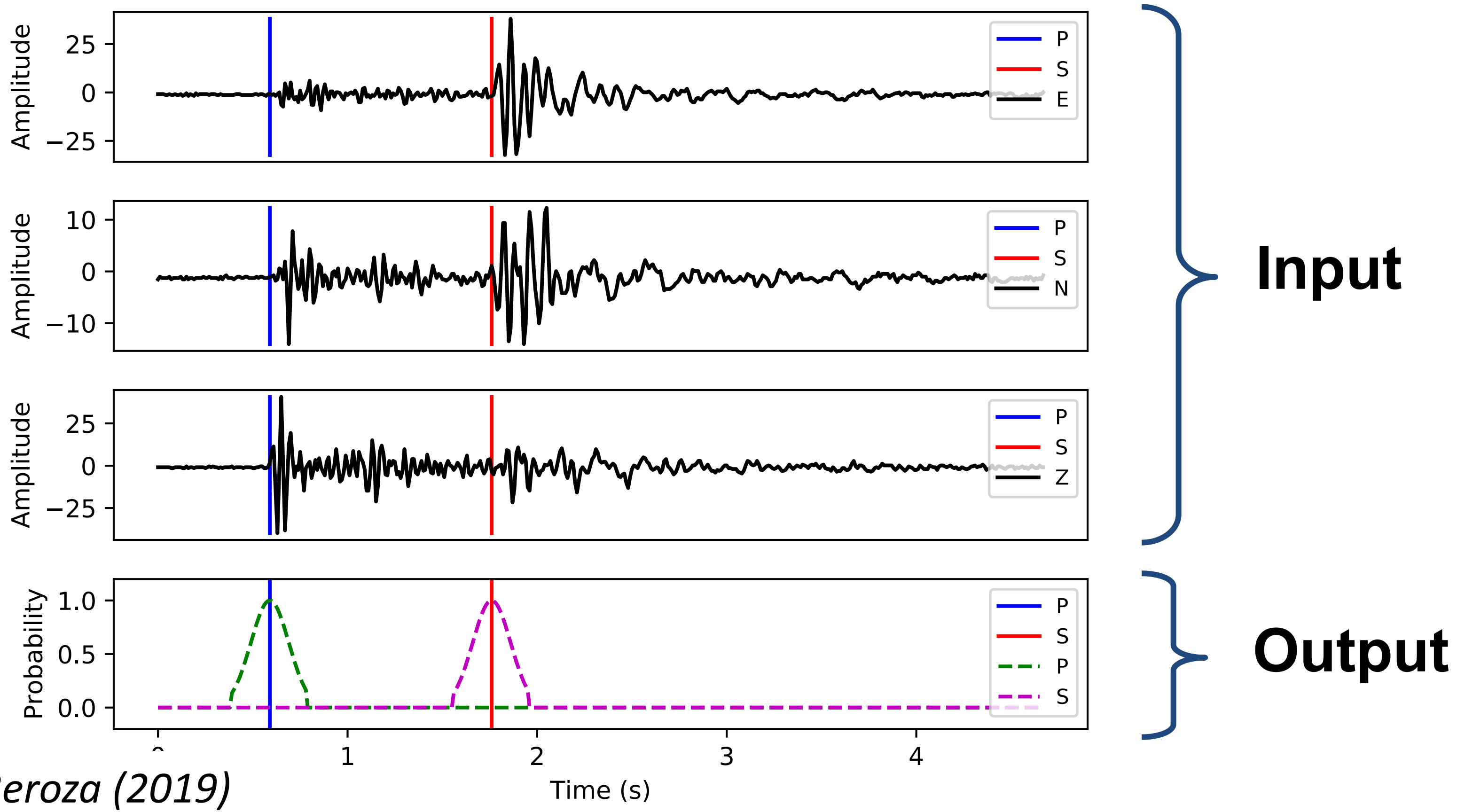
CRED

PhaseNet



<u>Output:</u> comprehensive catalog of discrete seismic events

PhaseNet for arrival time picking



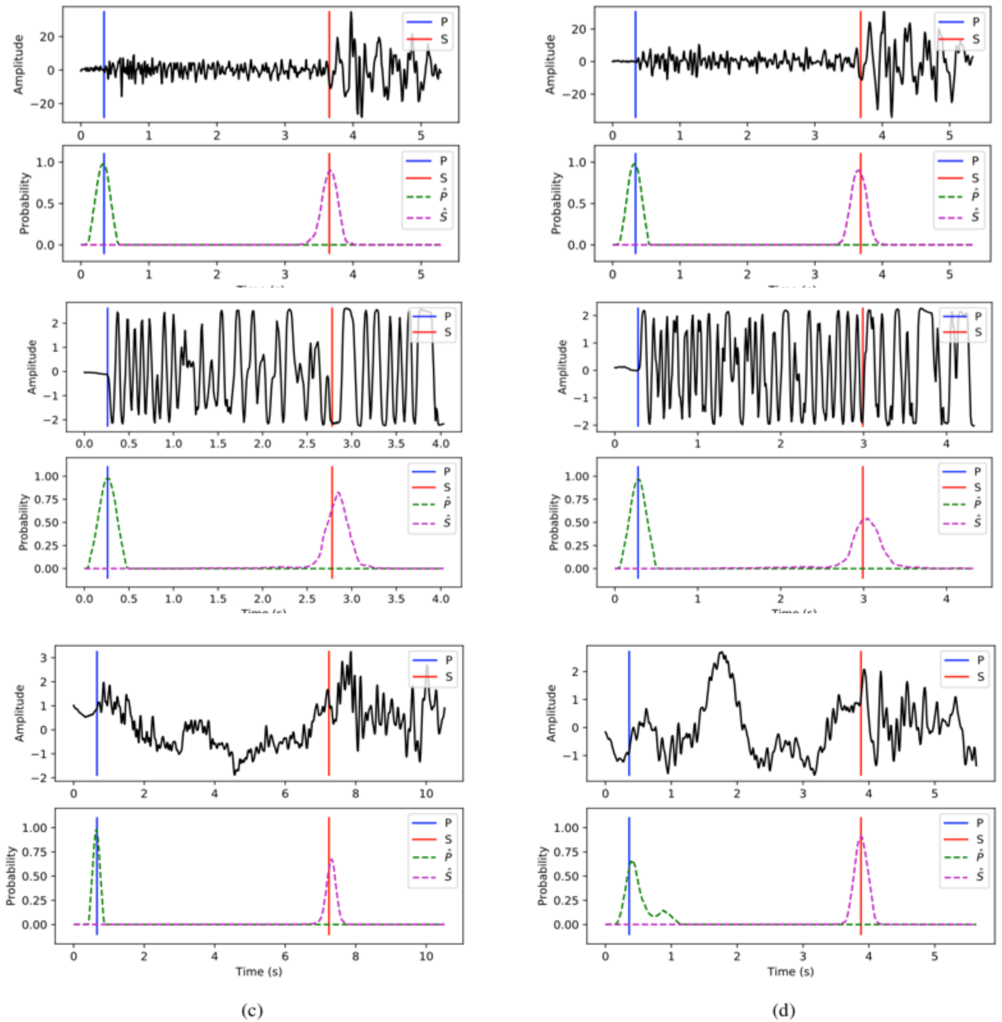
Zhu and Beroza (2019)

PhaseNet: deep neural net to classify P and S wave arrivals

Large, labeled data set: ~750,000 NCSN P and S wave picks + corresponding waveforms.

Classify waveforms with P, S, or noise probabilities.

Design so probability peaks at arrival time.

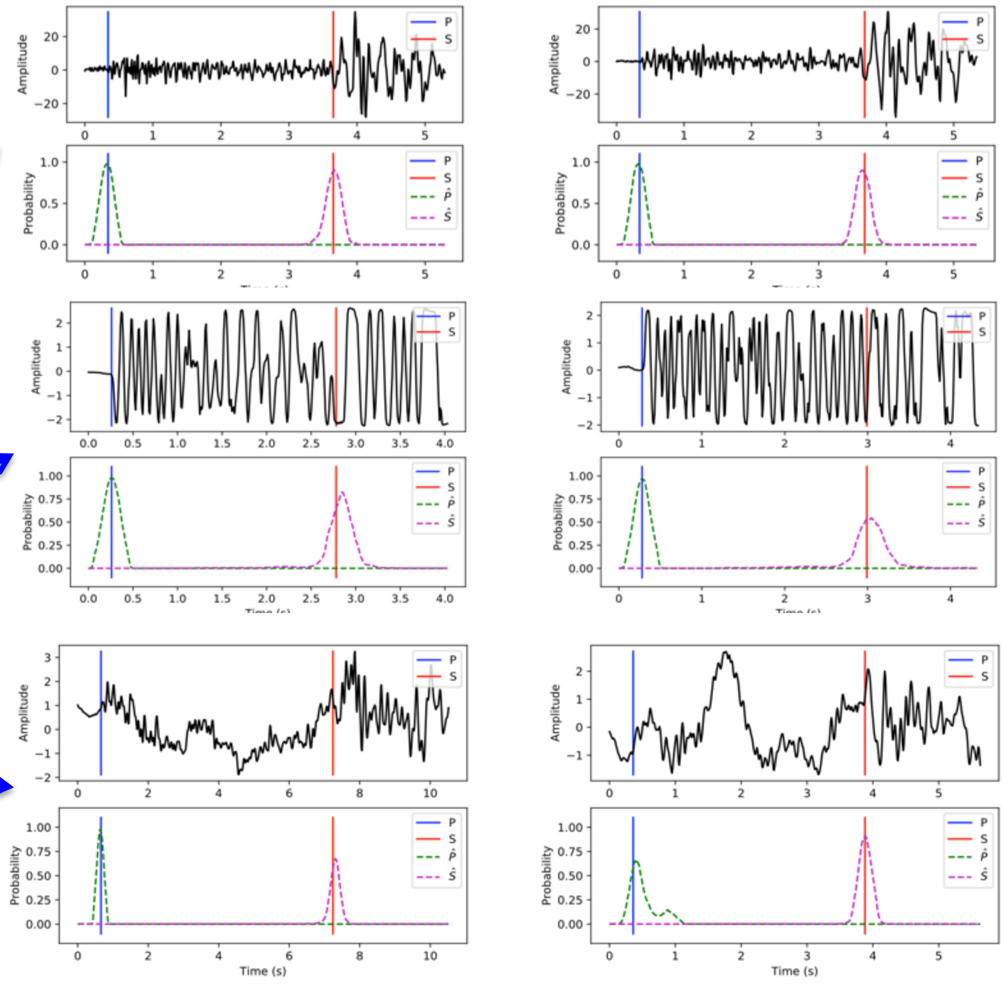


Zhu and Beroza (2019)

Separately classifies P & S

Works well on clipped data

Works well on noisy data



(c)

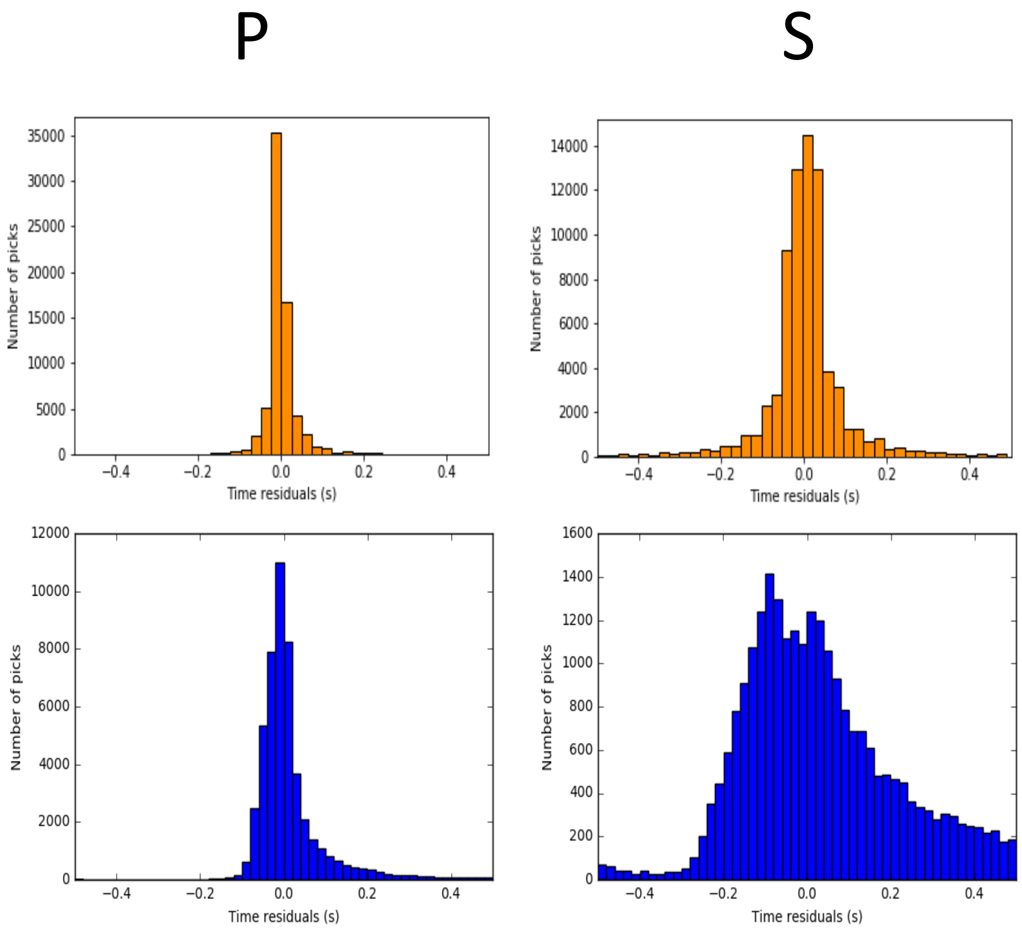
(d)

Zhu and Beroza (2019)

Classification of P, S Waves, and noise is accurate with high recall

PhaseNet

Standard Approach

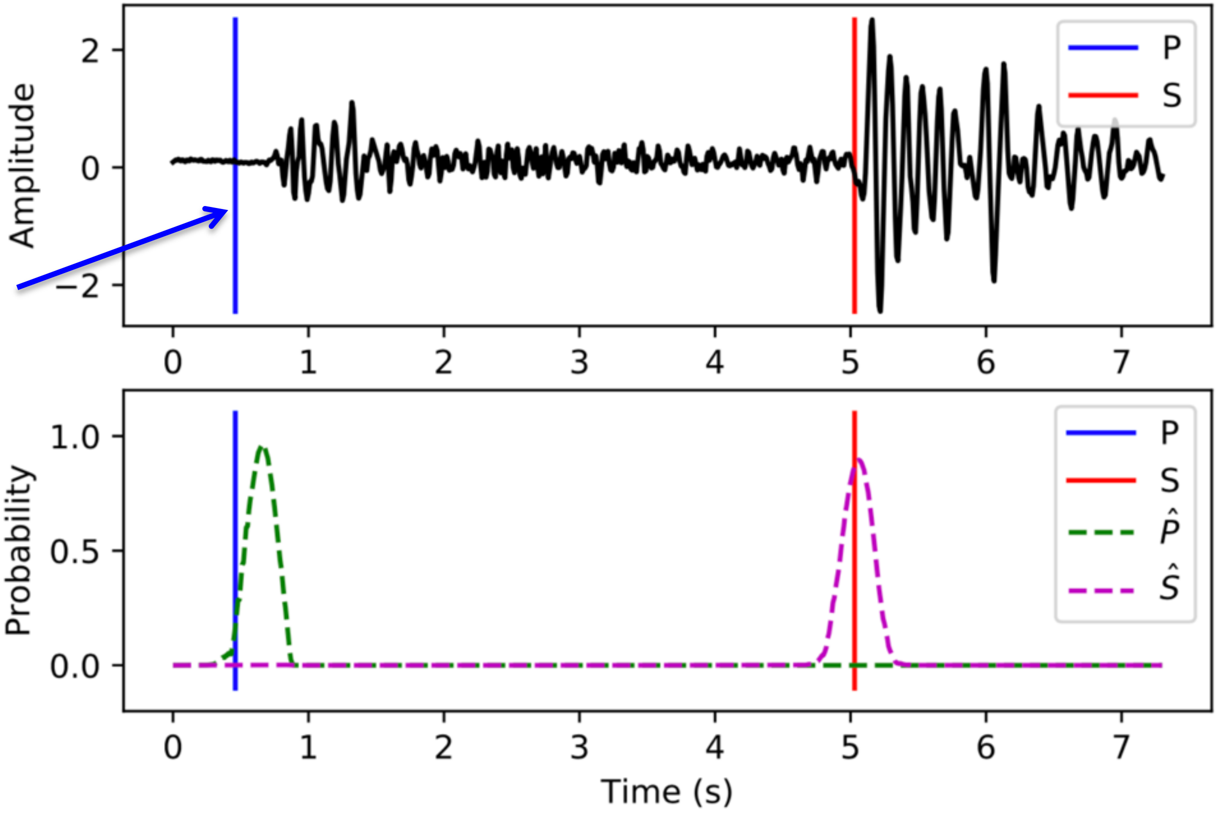


Zhu and Beroza (2019)

Compared with reviewed picks – taken as correct

What's the right answer? Analyst-reviewed picks have errors

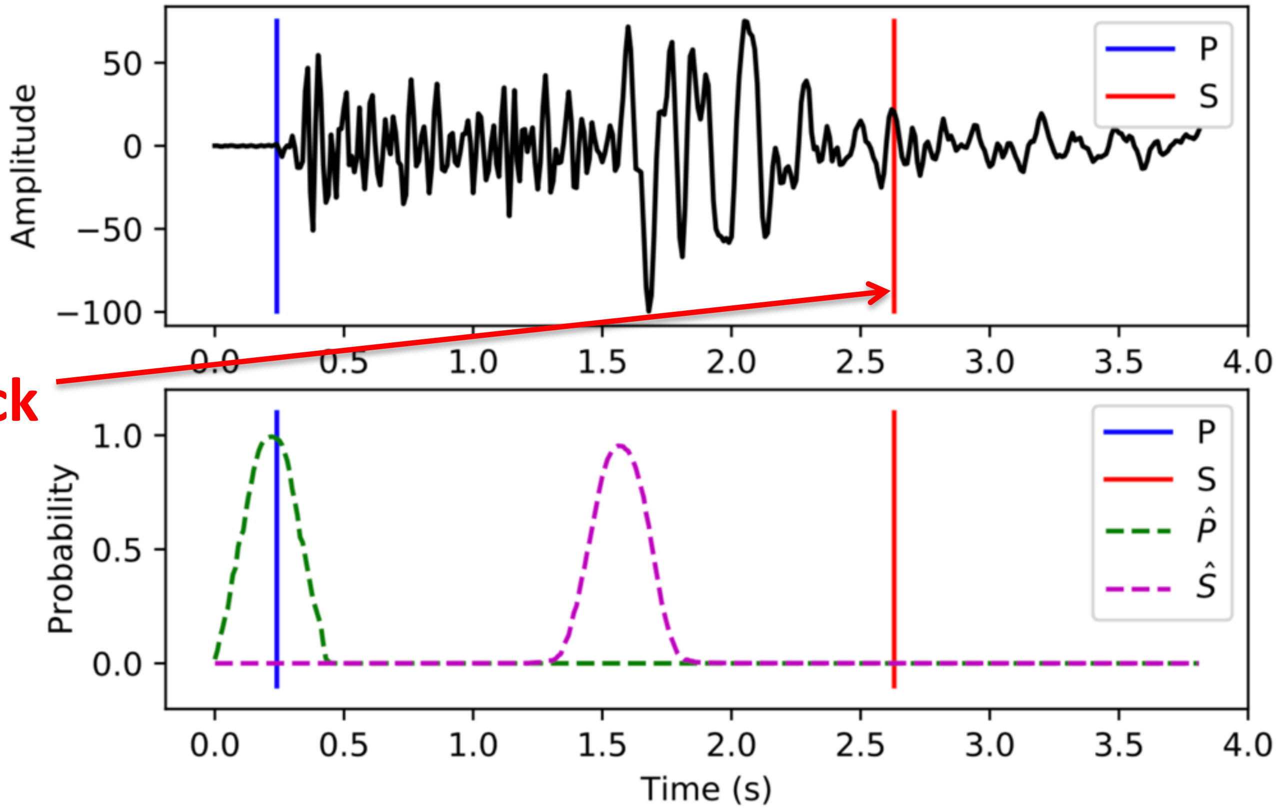
Example of bad P pick



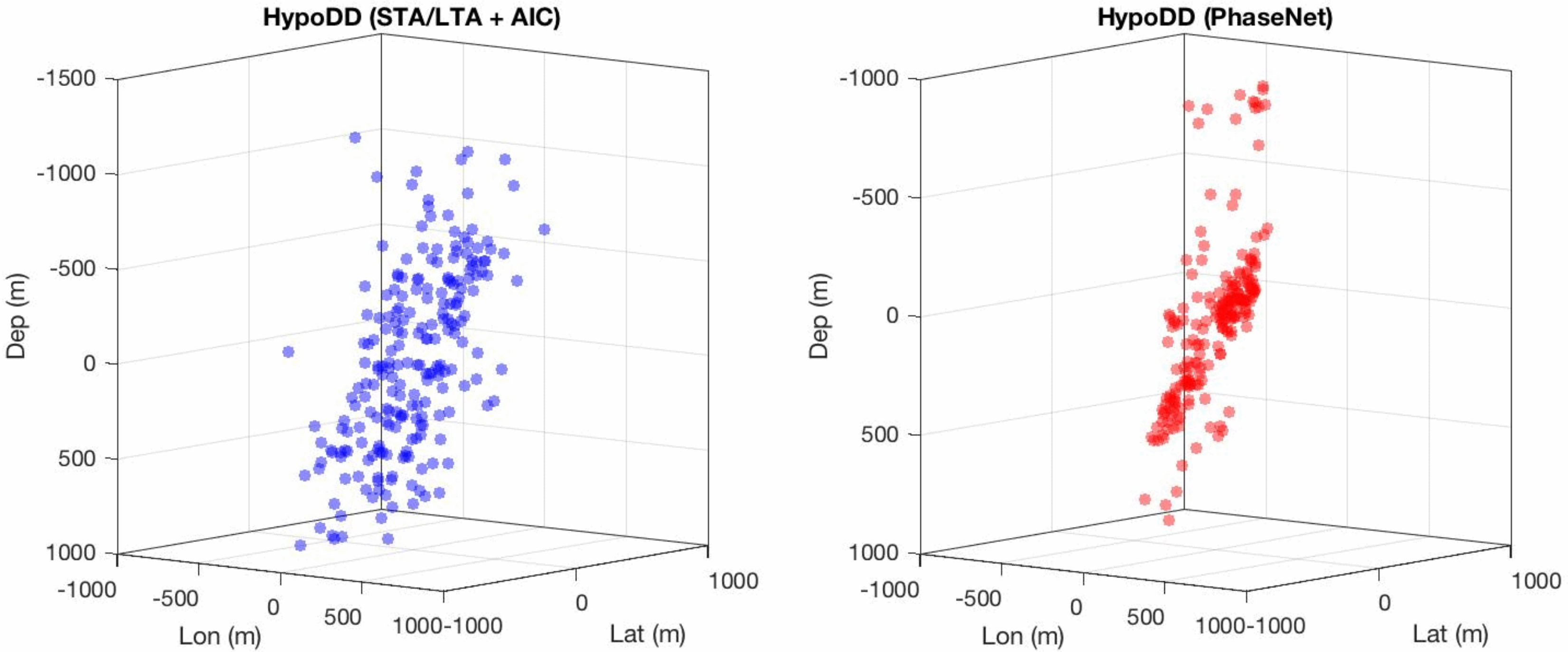
Zhu and Beroza (2019)

What's the right answer? Analyst-reviewed picks have errors

Example of bad S pick



Zhu and Beroza (2019)



HypoDD (STA/LTA + AIC)
-1500
-1000
-500
0
500
1000
Dep (m)
-1000
-500
0
500
1000
Lon (m)
-1000
0
1000
Lat (m)
HypoDD (PhaseNet)
-1000
-500
0
500
1000
Dep (m)
-1000
-500
0
500
1000
Lon (m)
-1000
0
1000
Lat (m)

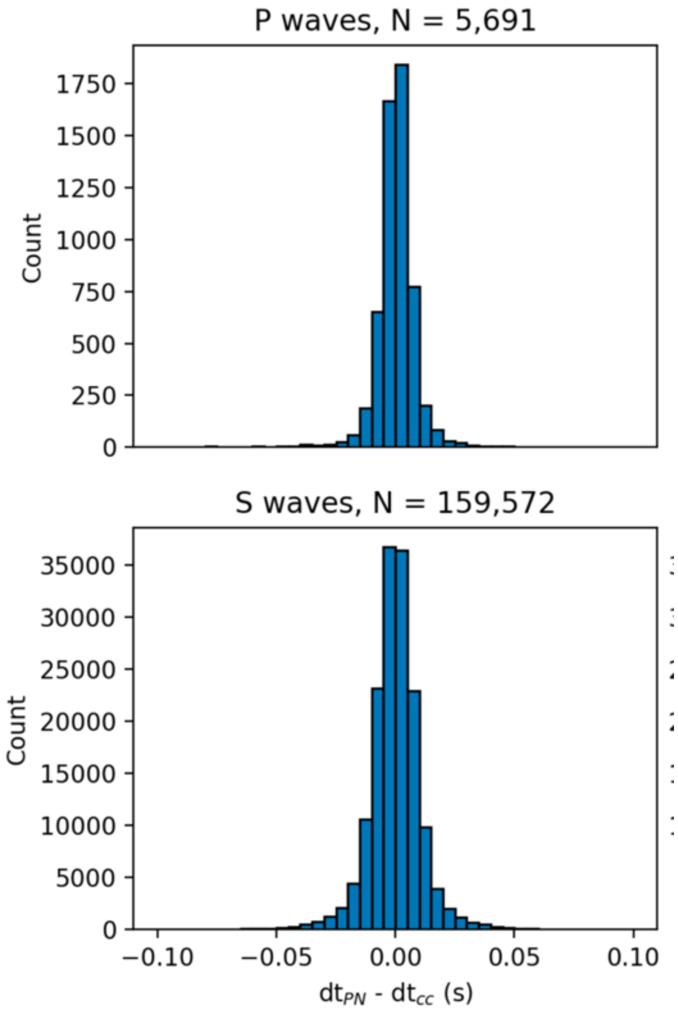
Test on data from Axial Seamount

Histogram of differential arrival times:
PhaseNet vs. cross correlation

Trained on NCSN, but applied to OBS data

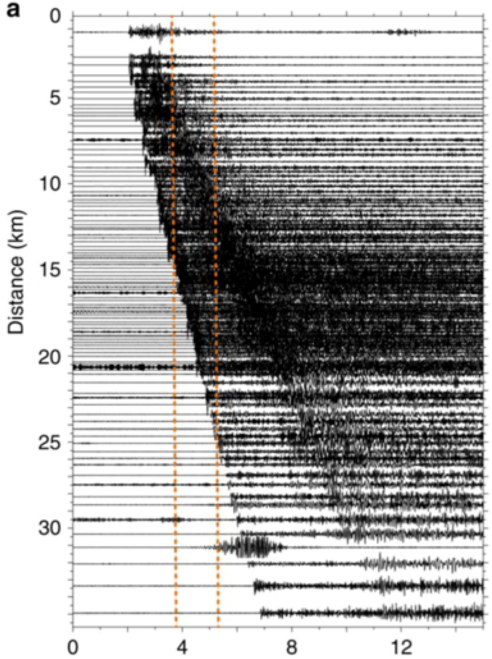
PhaseNet gets differential arrival times within a few samples (better than analyst)

Tan et al. [*2020*]



Earthquake Monitoring Workflow

<u>Input:</u> continuous time series of ground motion from a seismic network



Pre-processing ← **DeepDenoiser**

↓

Event Detection ← **CRED**

↓

Phase Picking ← **PhaseNet**

↓

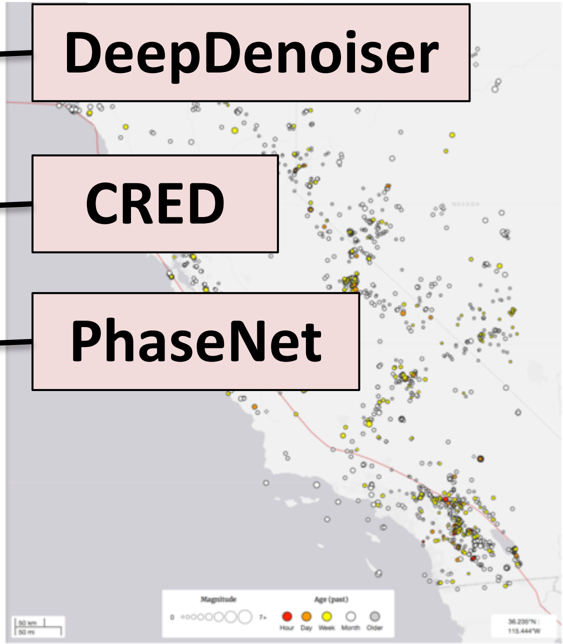
Association

↓

Location

↓

Characterization

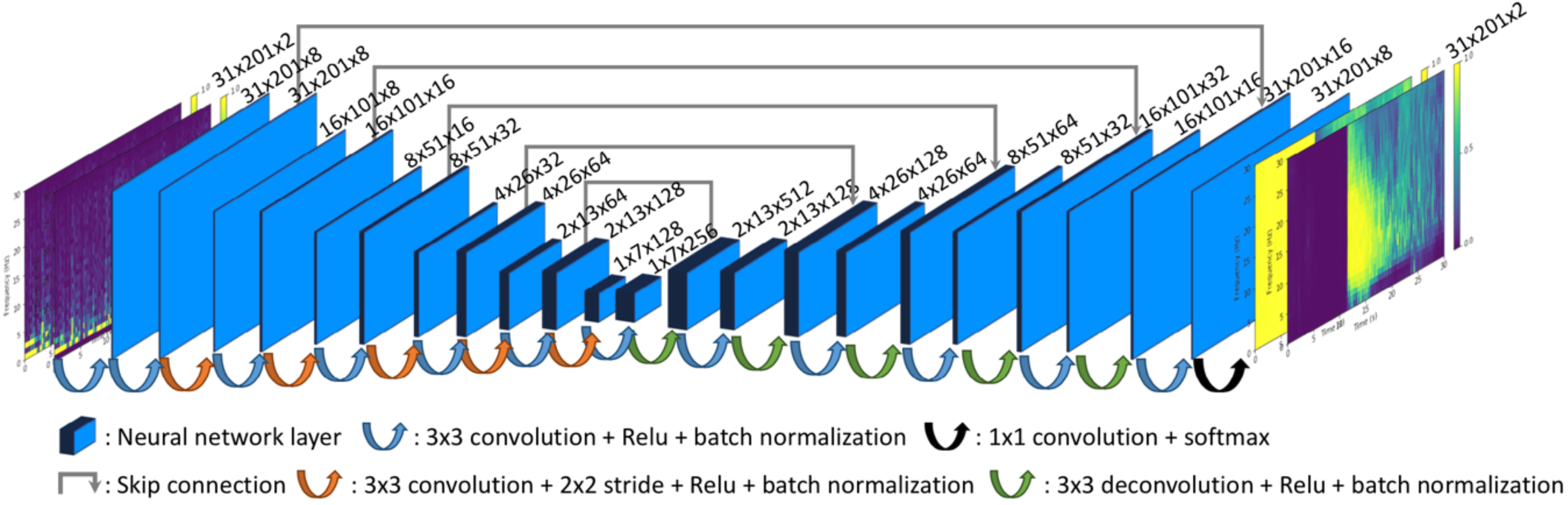


<u>Output:</u> comprehensive catalog of discrete seismic events

Train a deep neural network (DeepDenoiser) to separate signal from noise

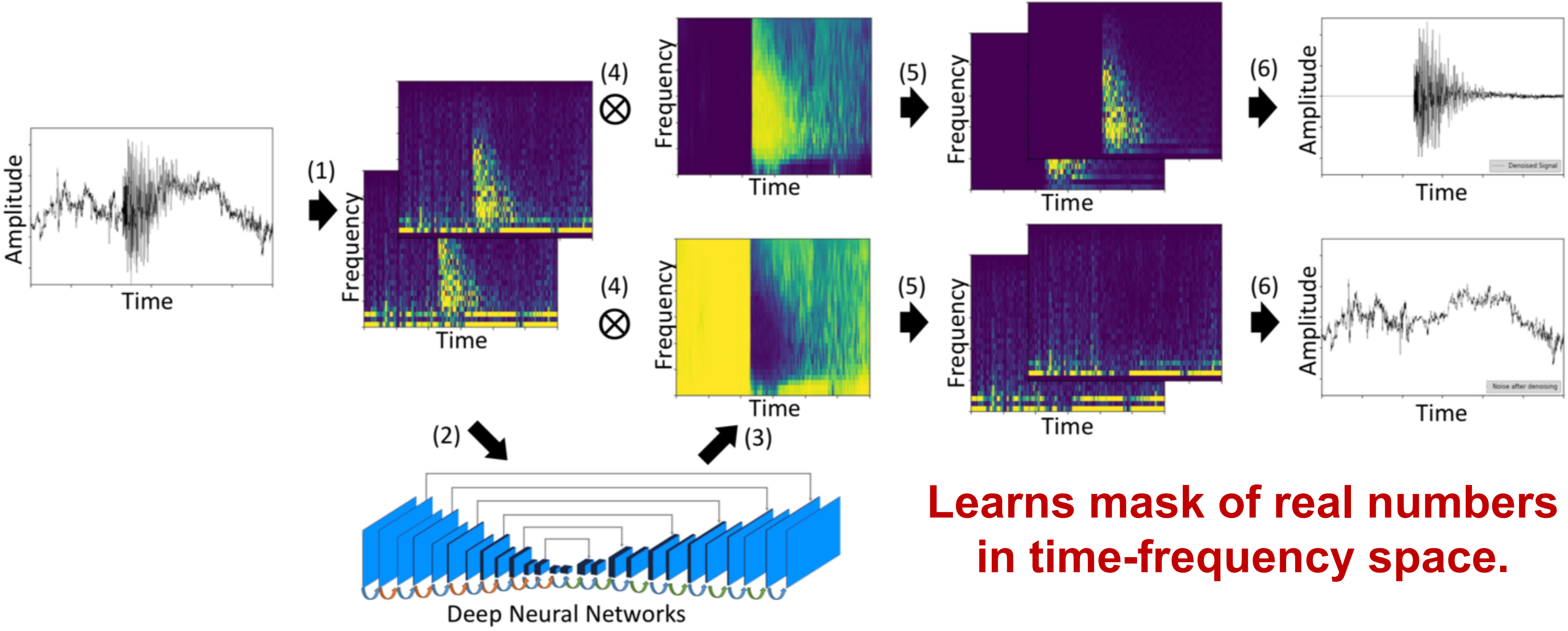
Data are 56,345 high snr earthquake waveforms and 179,233 noise waveforms

Expand training data to ~10^9 by adding signals to noise at varying snr

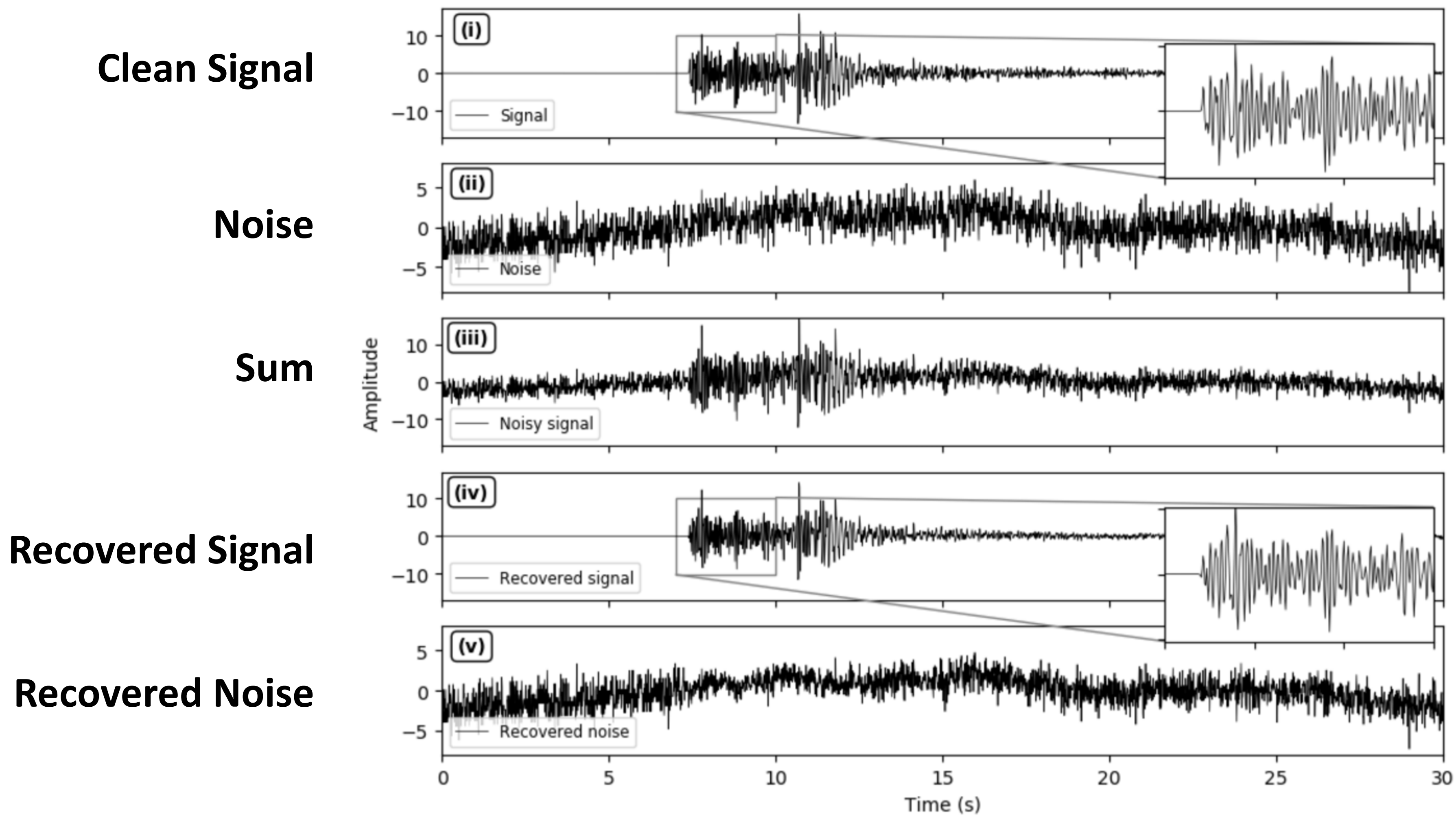


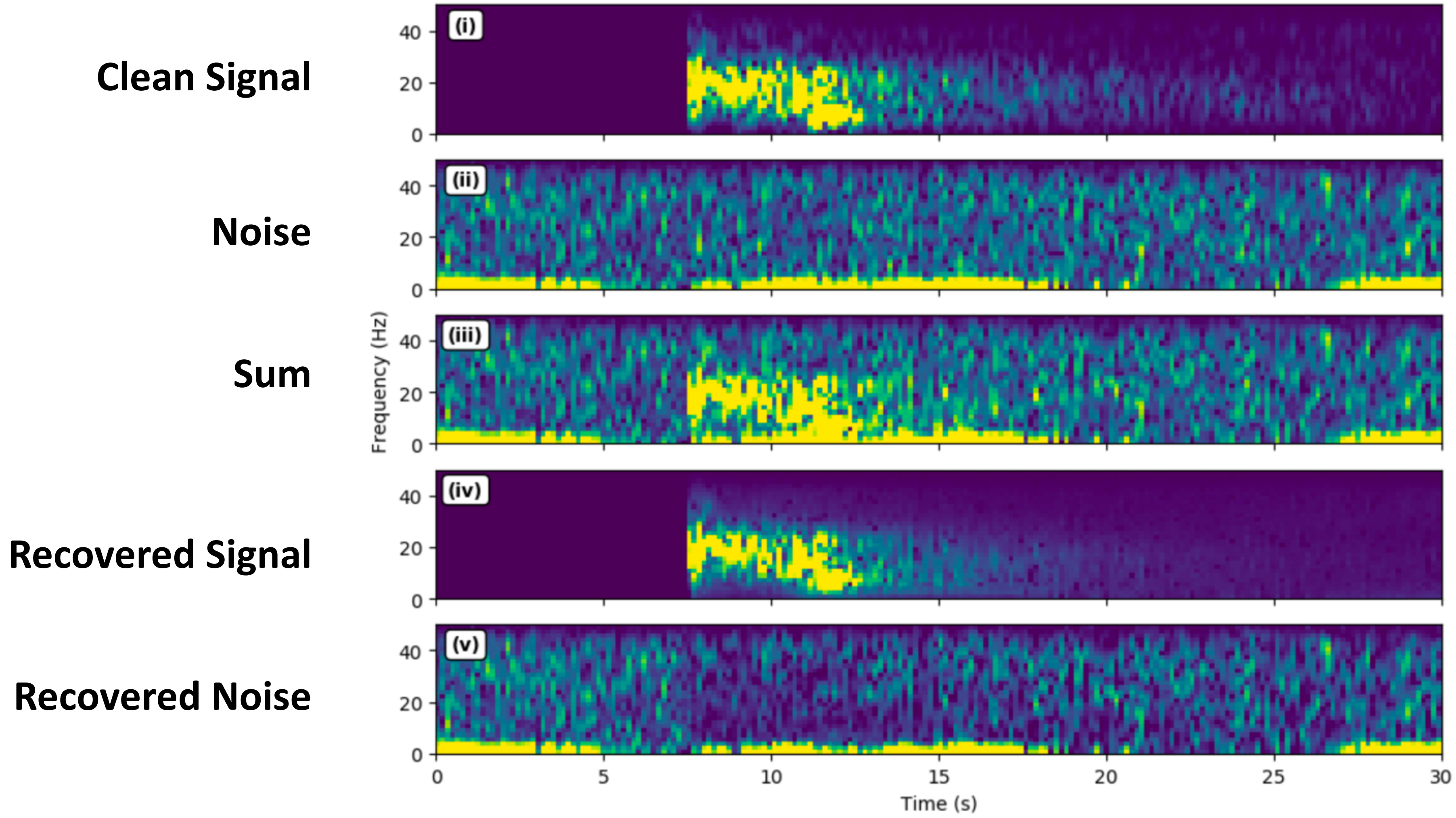
Zhu et al. (2019b)

DeepDenoiser develops masks to isolate signal and noise



Learns mask of real numbers in time-frequency space.


Clean Signal
Noise
Sum
Recovered Signal
Recovered Noise
(i)
Signal
(ii)
Noise
(iii)
Noisy signal
(iv)
Recovered signal
(v)
Recovered noise
Amplitude
Time (s)
10
0
−10
5
−5
0
5
10
15
20
25
30



Clean Signal
Noise
Sum
Recovered Signal
Recovered Noise
(i)
(ii)
(iii)
(iv)
(v)
Frequency (Hz)
Time (s)
0
20
40
0
5
10
15
20
25
30

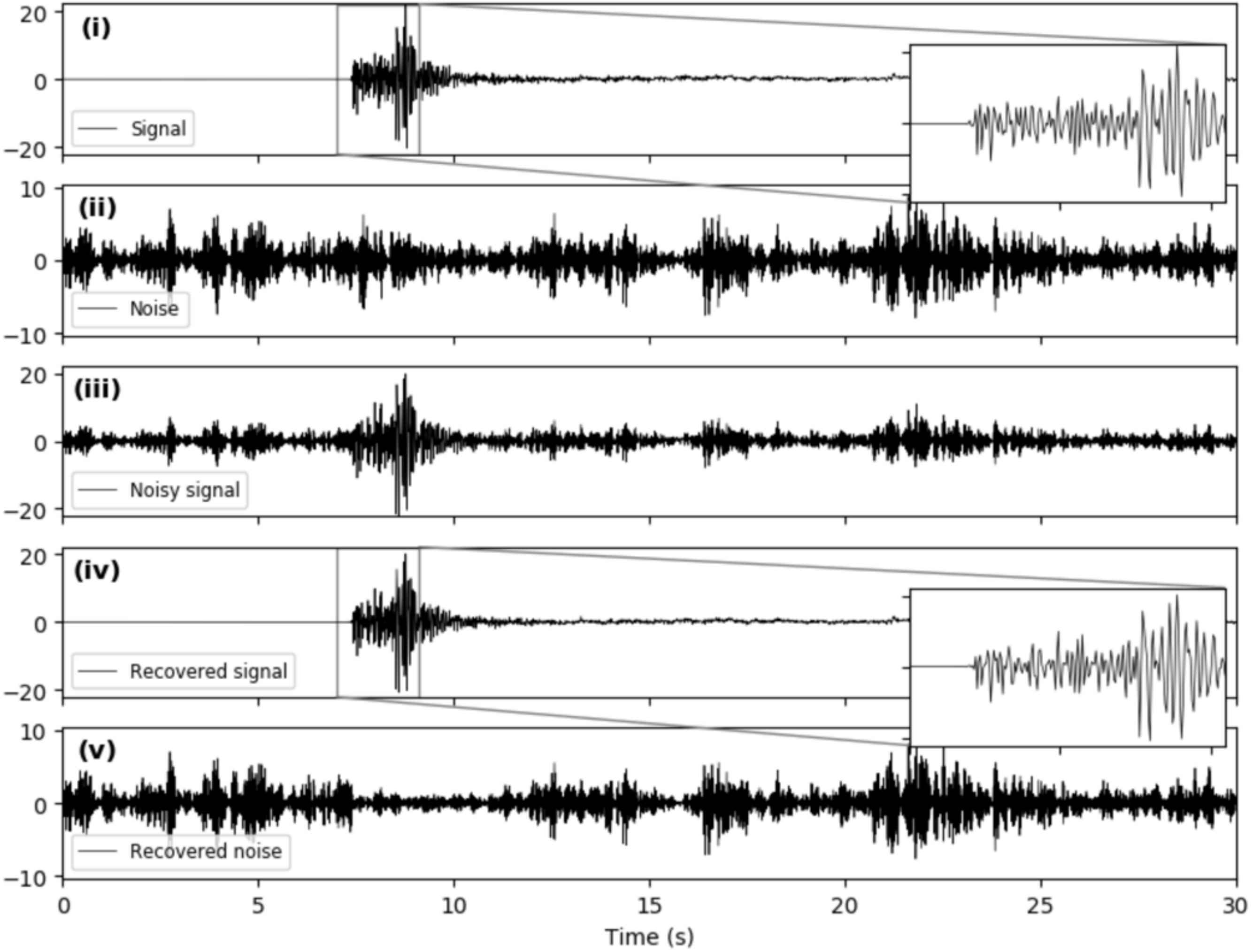
Clean Signal

Noise

Sum

Recovered Signal

Recovered Noise



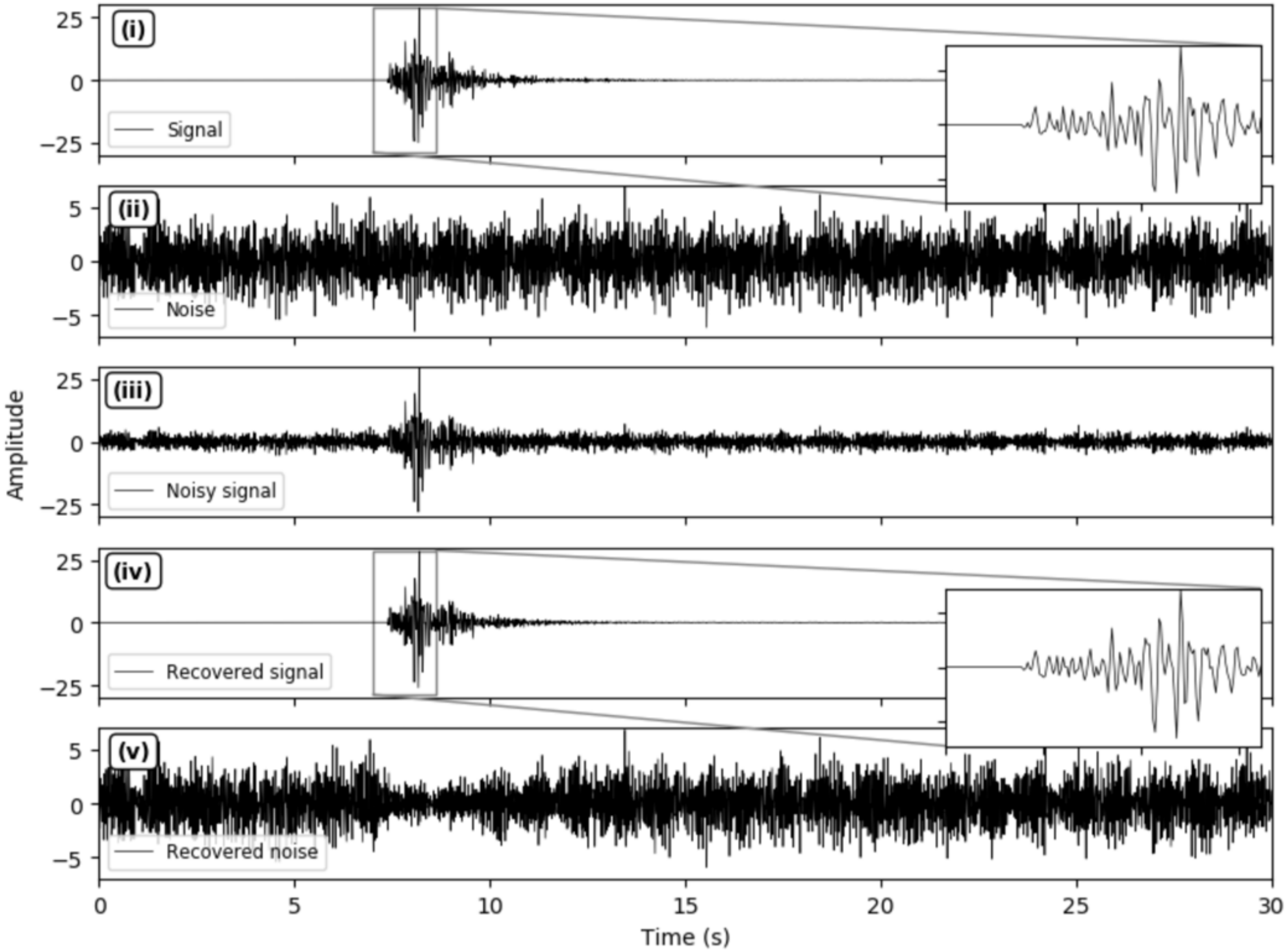
Clean Signal

Noise

Sum

Recovered Signal

Recovered Noise



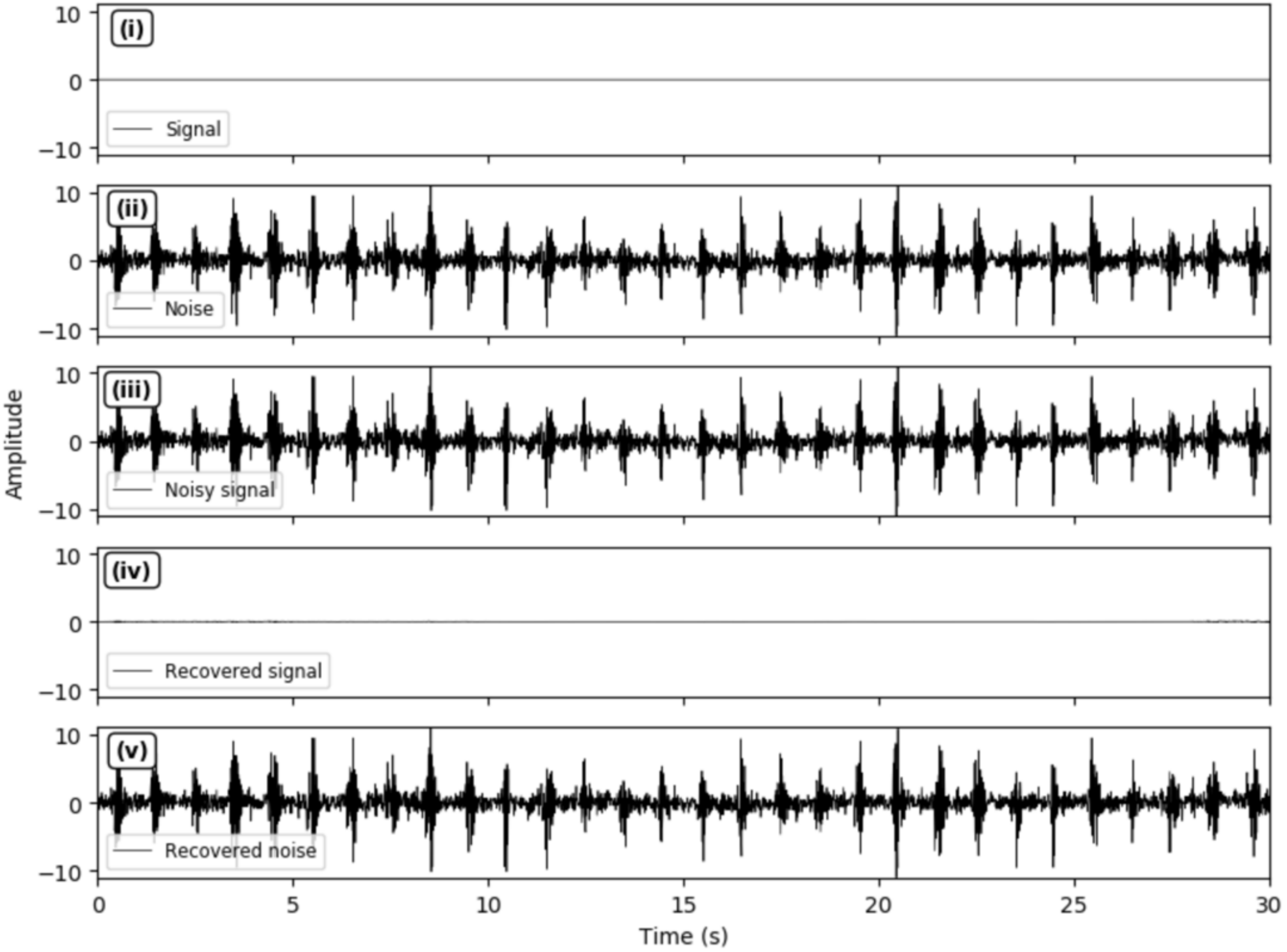
Clean Signal

Noise

Sum

Recovered Signal

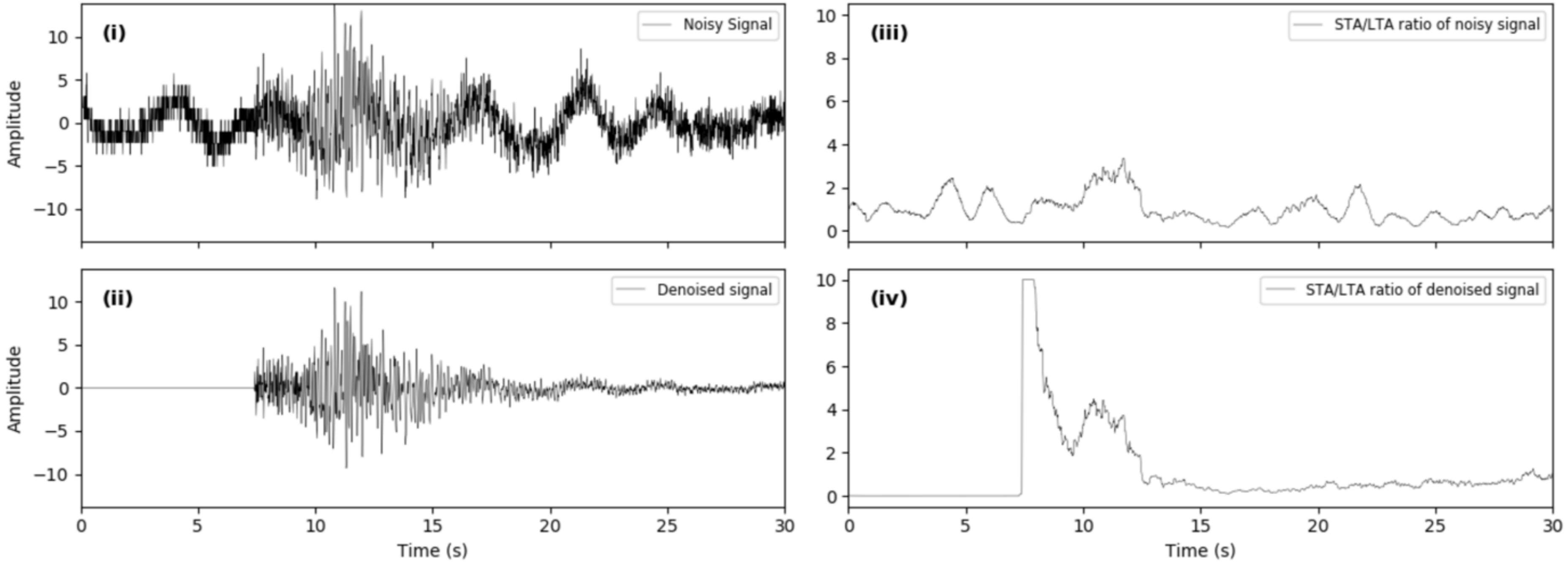
Recovered Noise



DeepDenoiser improves detection sensitivity

Time Series

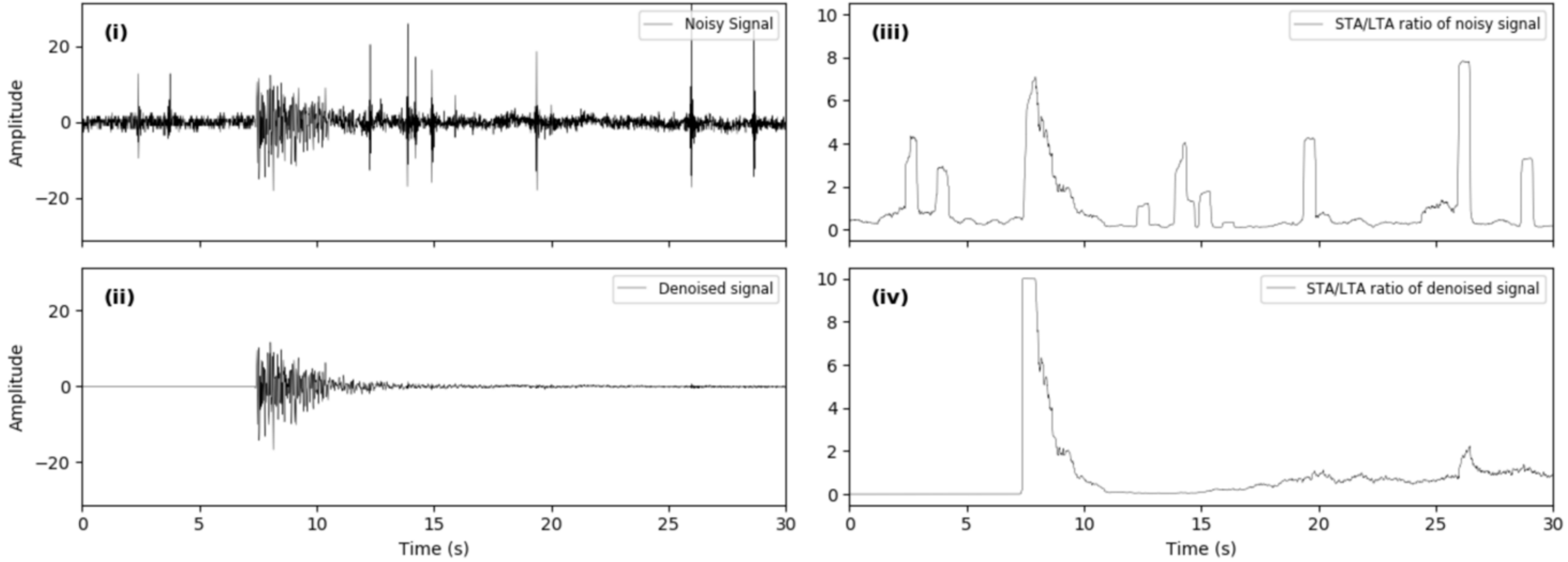
STLA/LTA Detection Statistic



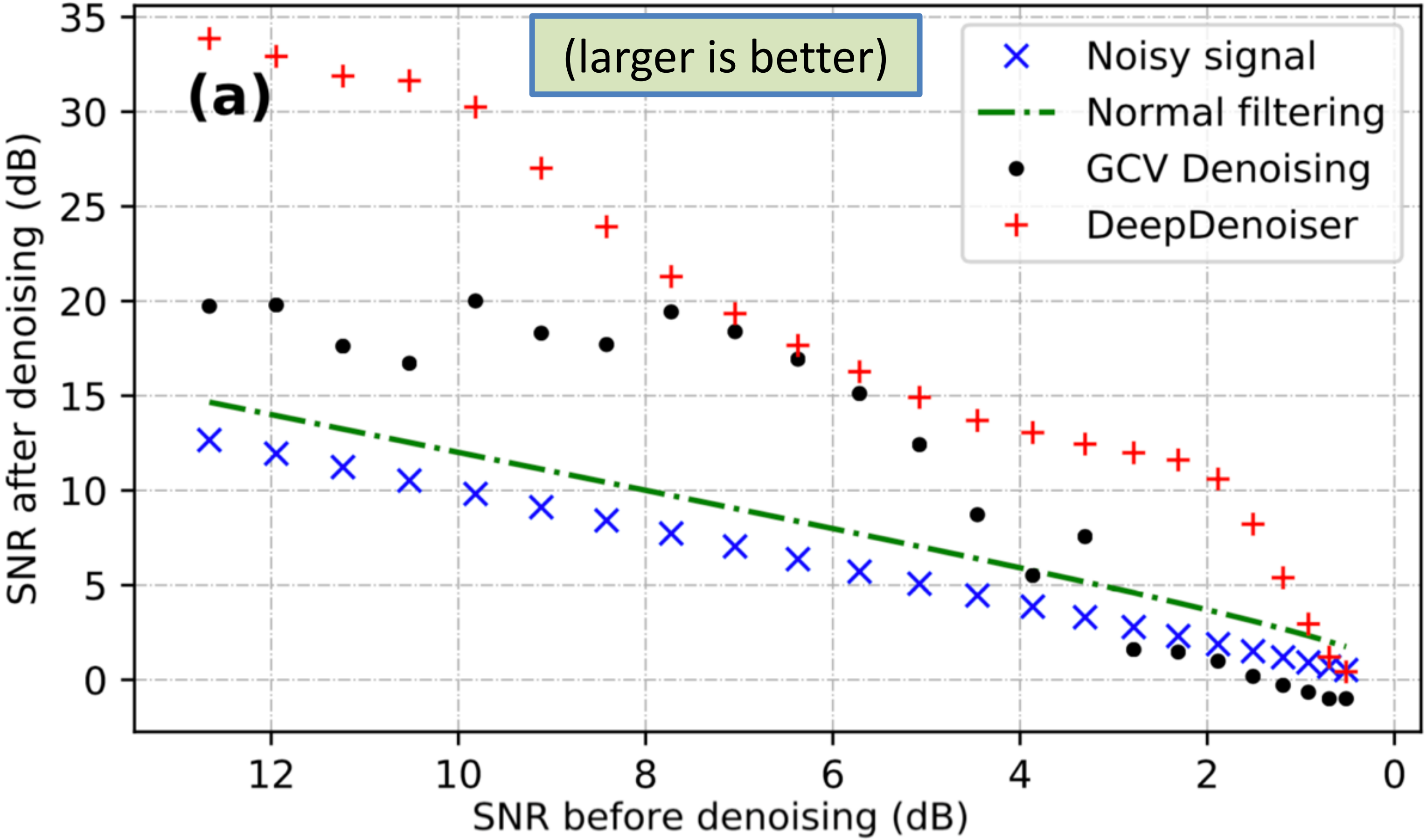
DeepDenoiser helps avoid false detections

Time Series

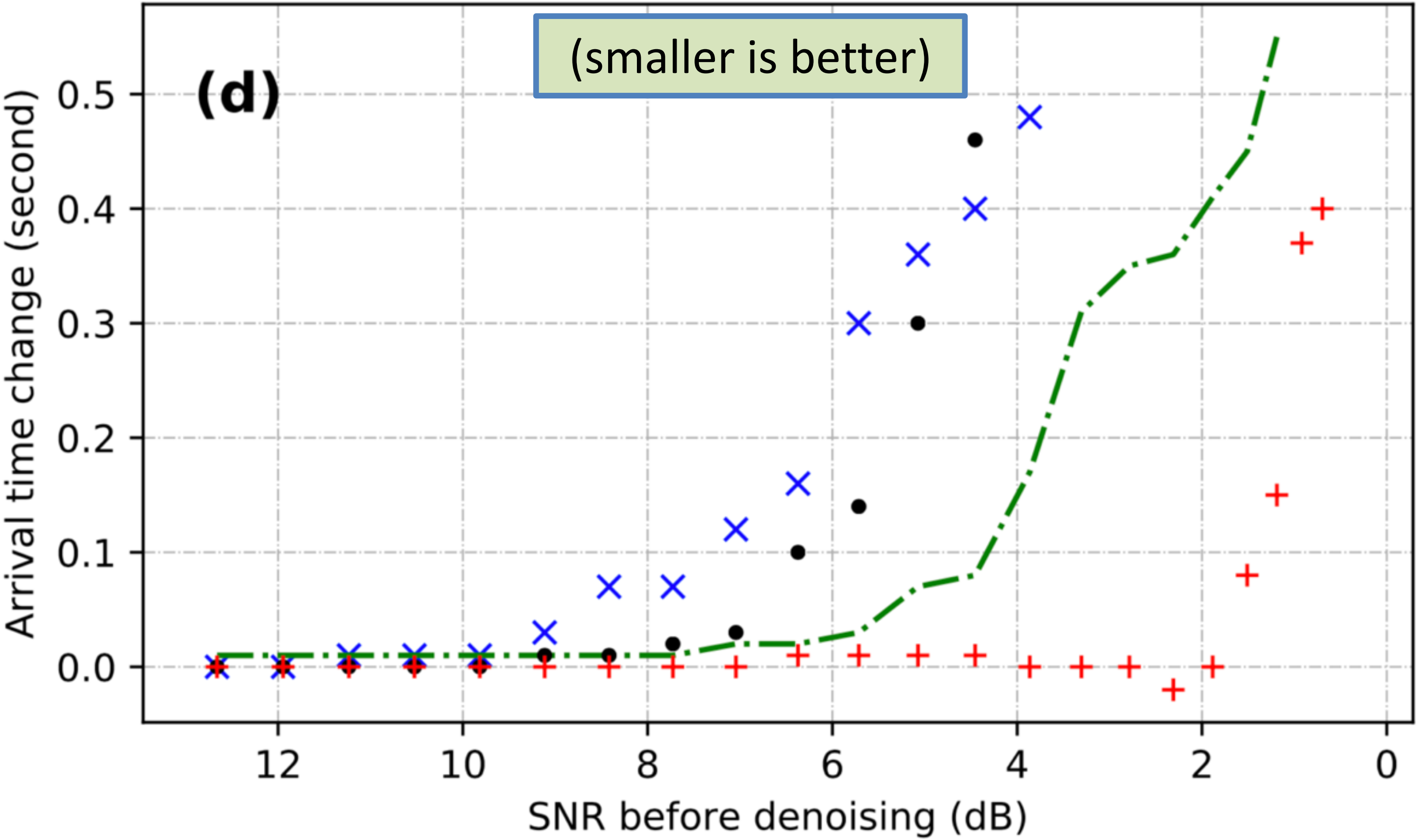
STLA/LTA Detection Statistic



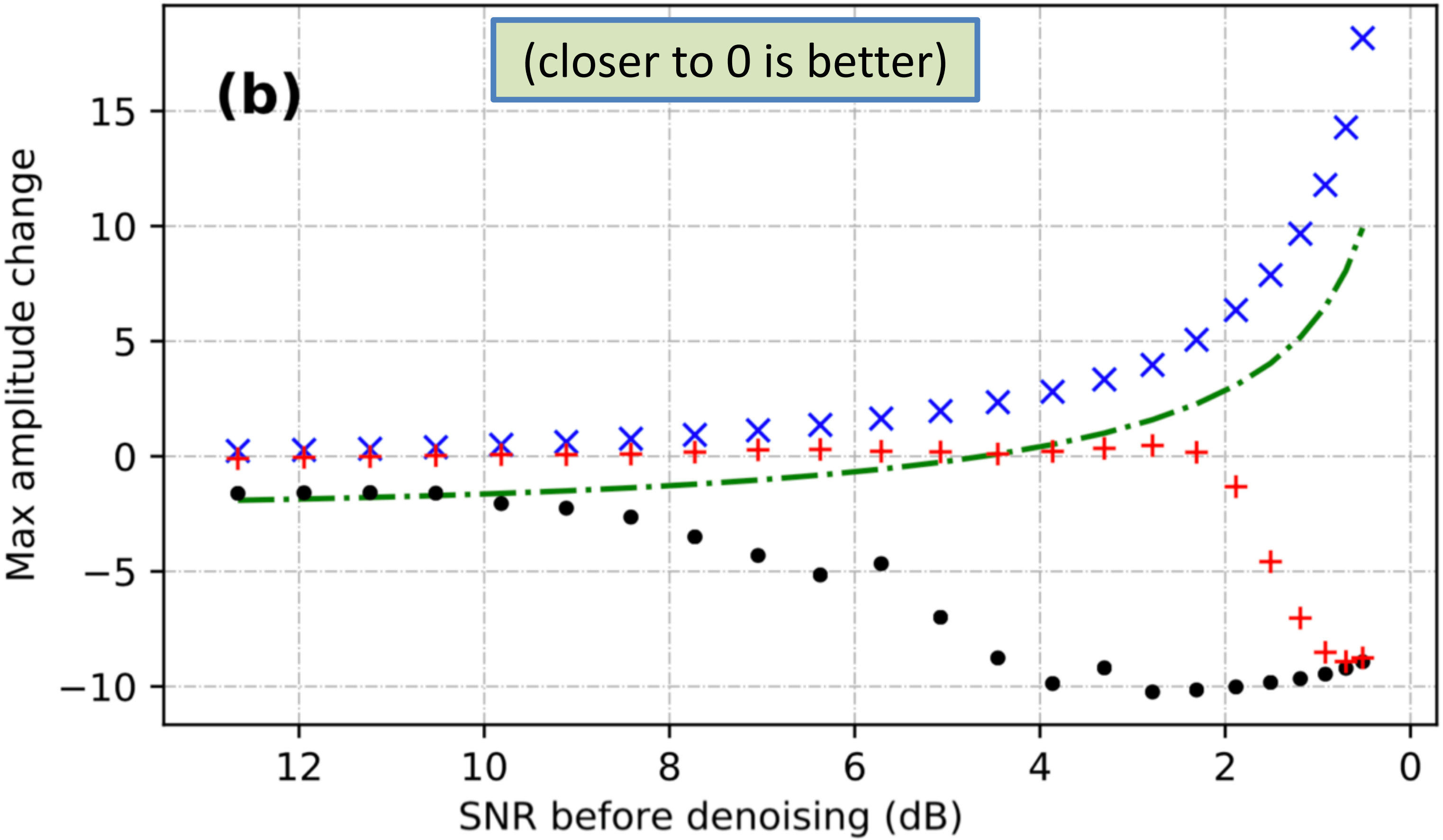
DeepDenoiser improves snr better



DeepDenoiser recovers arrival times better



DeepDenoiser recovers amplitudes better



DeepDenoiser has many potential applications

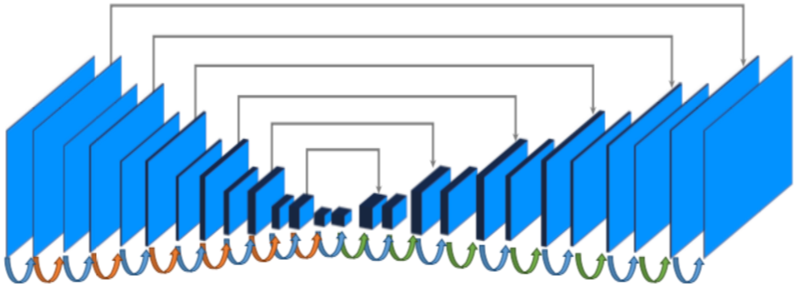
Conventional Seismic Monitoring

Urban Seismic Monitoring

- MeSONet
- Long Beach Nodal Array
- DAS

Seafloor Seismic Monitoring

- OBS
- S-Net
- DAS

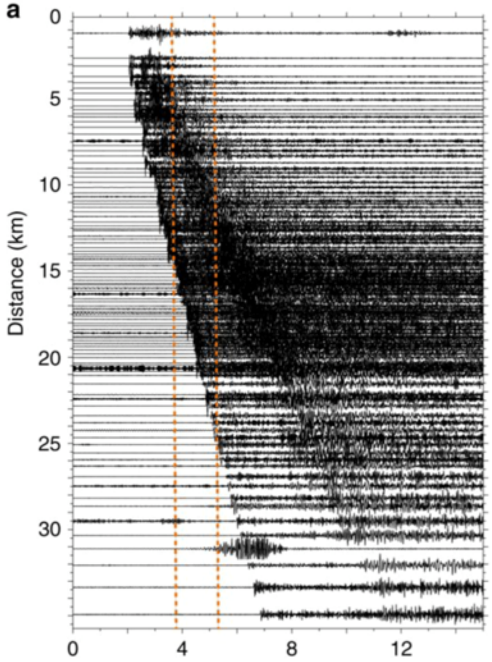


Volcano Monitoring

LIGO?

We are on the threshold of ML-Based EQ monitoring

<u>Input:</u> continuous time series of ground motion from a seismic network



Pre-processing

↓

Event Detection

↓

Phase Picking

↓

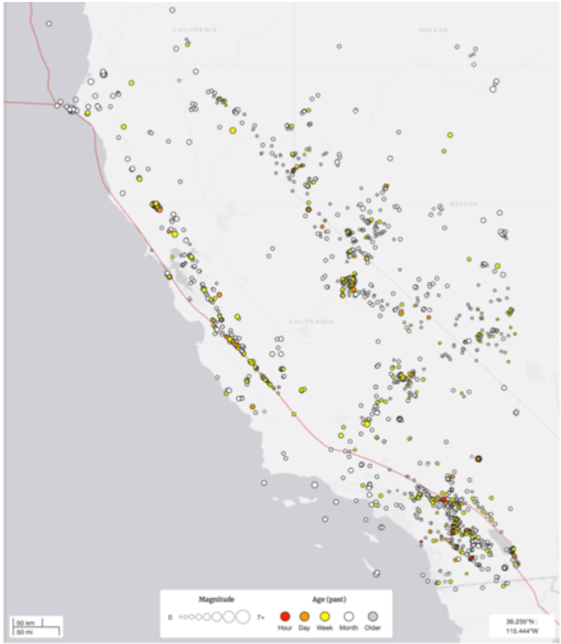
Association

↓

Location

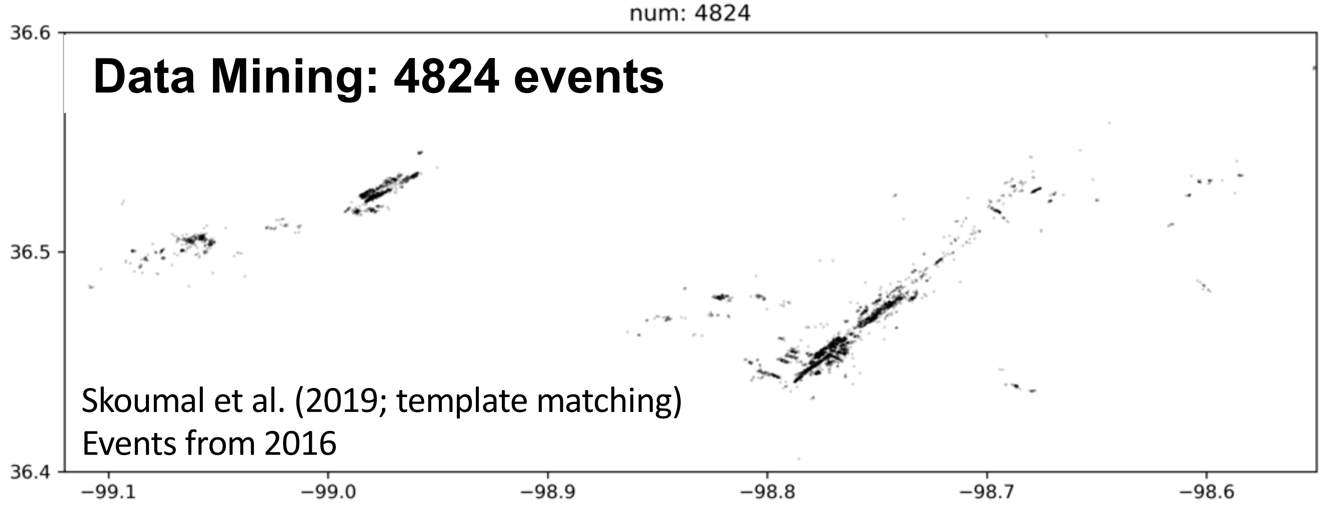
↓

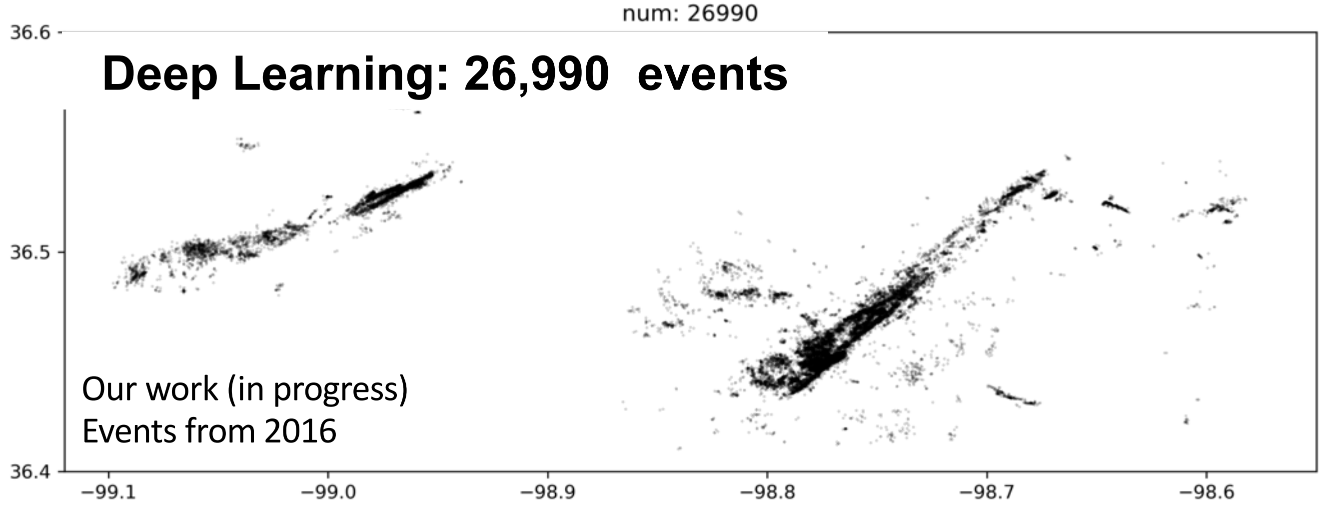
Characterization



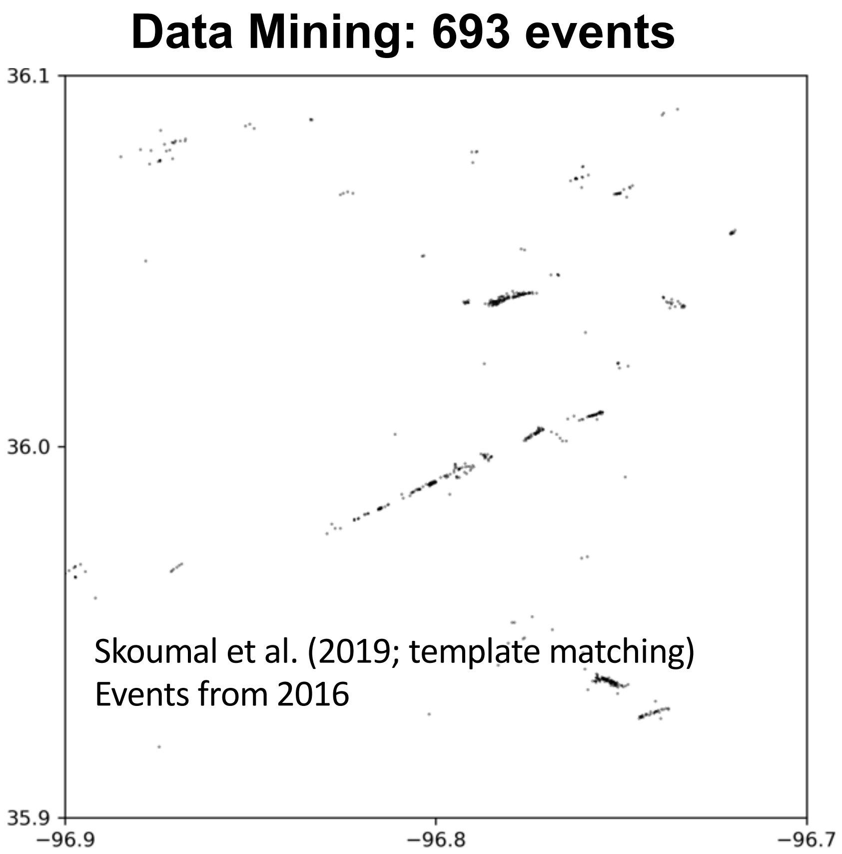
<u>Output:</u> comprehensive catalog of discrete seismic events

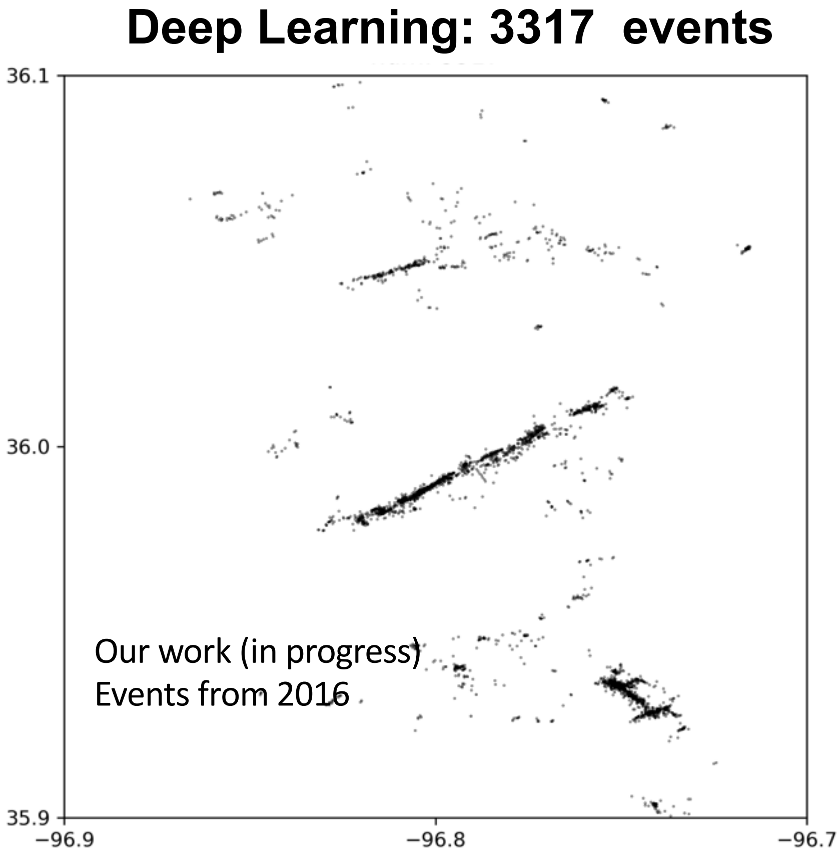
Fairview, Oklahoma





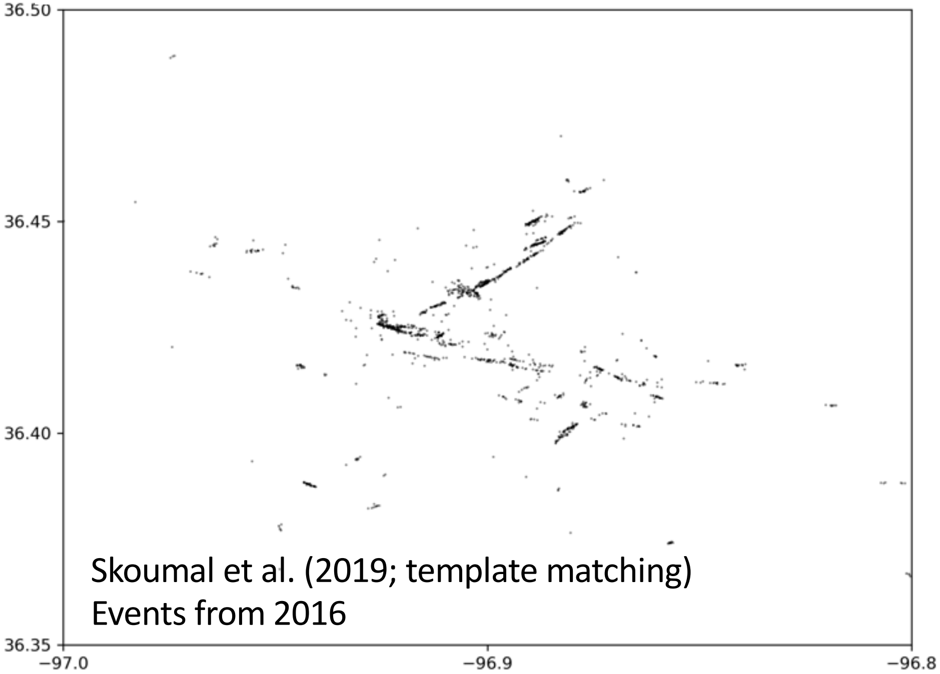
Cushing, Oklahoma



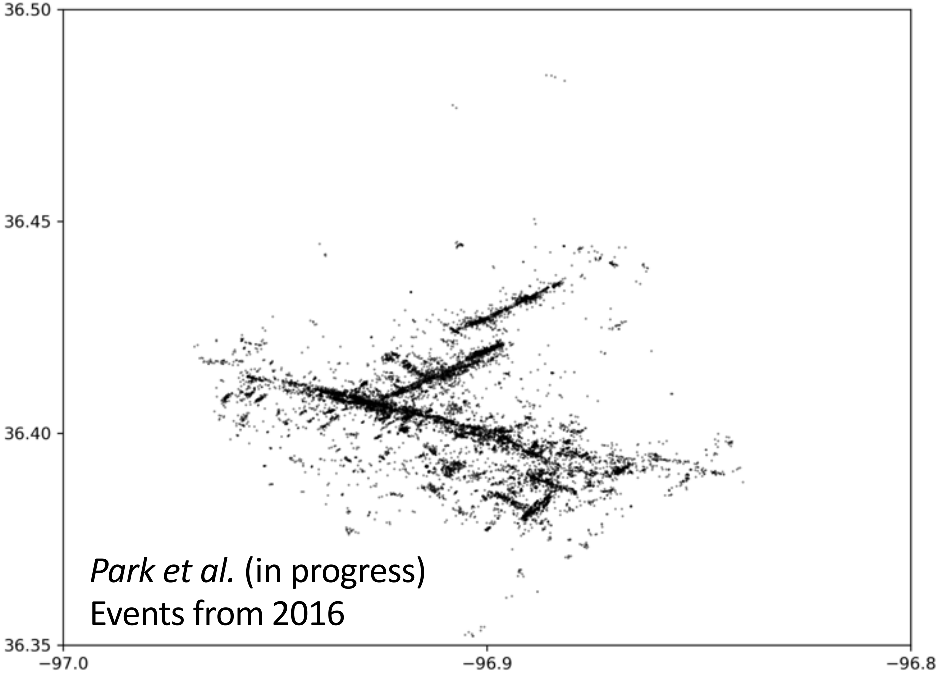


Pawnee, Oklahoma sequence

Data Mining: 1423 events

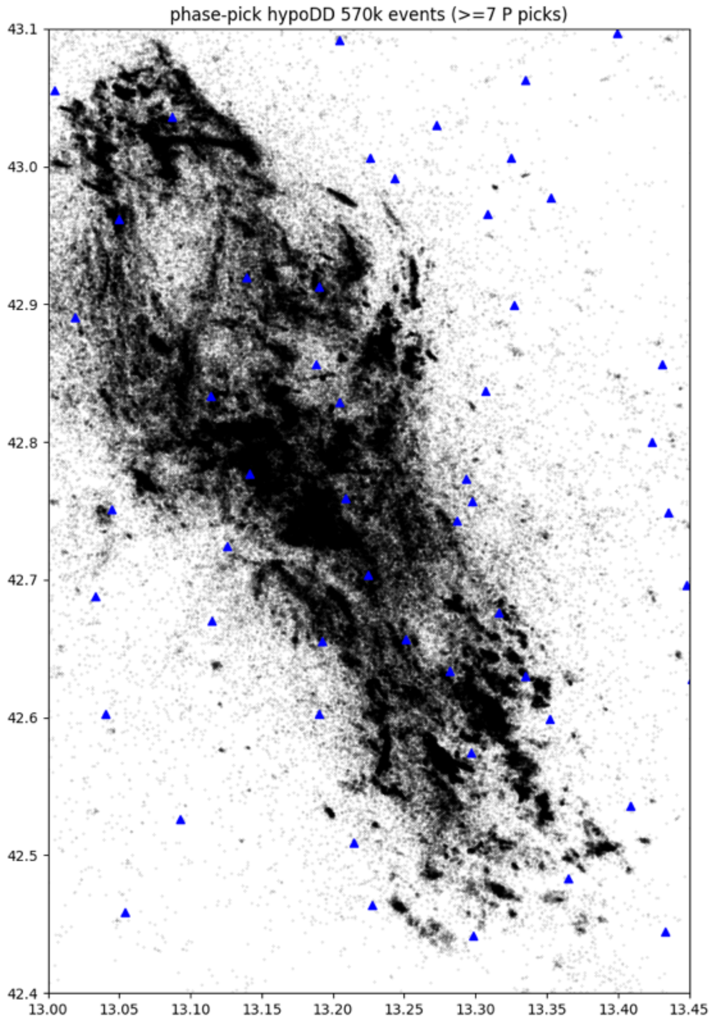


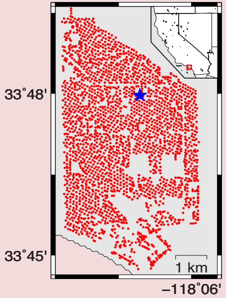
Deep Learning: 12,356 events



ML methods: faster, easier, more general, and result in a more complete picture

Work in progress for the Apennines, Italy earthquake sequence: deep-learning-based earthquake monitoring increases earthquake catalog to ~570,000 well-located events (so far) resulting in a dramatic increase in information on fault structure and earthquake processes.





Big Networks (Large-N)

Continuous Data (Large-T)

New Data Sources

Standard practice is sub-optimal

- Channels/stations treated in isolation
- Detection-association-location-characterization treated separately
- Events treated individually (e.g., STA/LTA phase picking)

Earthquake monitoring with AI (ready to go)

- Machine learning for detection and arrival-time picking
- AI to extract meaning from results
- Concatenation of detection/association/location/characterization

Earthquake monitoring with AI (coming soon)

- Recursive improvement
- The network as the sensor
- Merging of detection-association-location-characterization

I hope you find this useful

FAST available at:

https://github.com/stanford-futuredata/FAST

DeepDenoiser and PhaseNet available at:

https://github.com/wayneweiqiang?tab=repositories