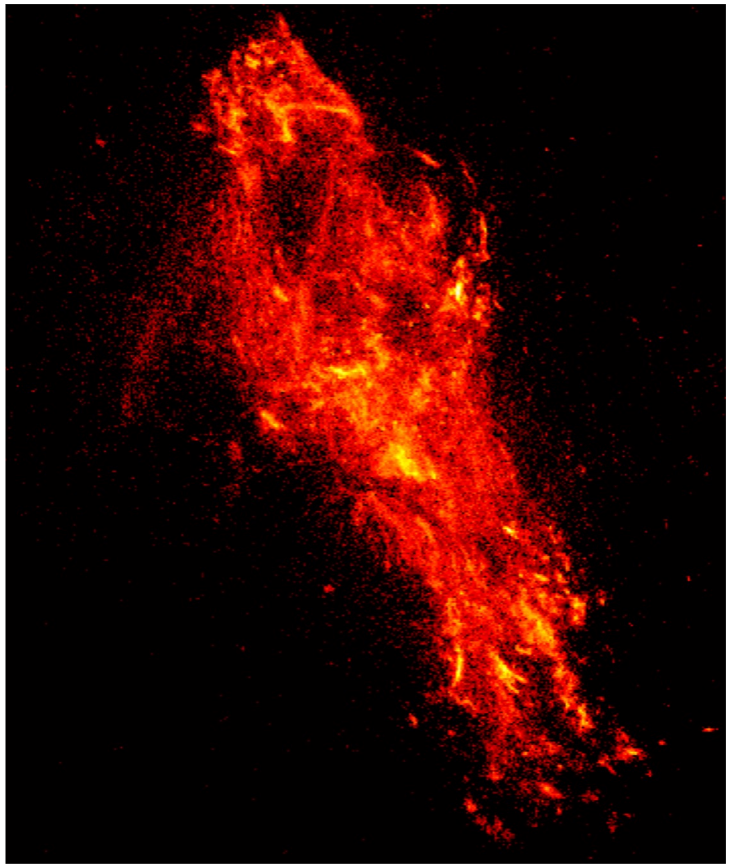
A New Generation of Earthquake Catalogs

Felix Waldhauser

Kaiwen Wang, Eric Beauce, Theresa Sawi, David Schaff, Ben Holtzman, Nate Groebner, Weiqiang Zhu (UCB), and the SCOPED gang

Lamont-Doherty Earth Observatory
Columbia University

CRESCENT ML WS, Seattle May12-14 2025



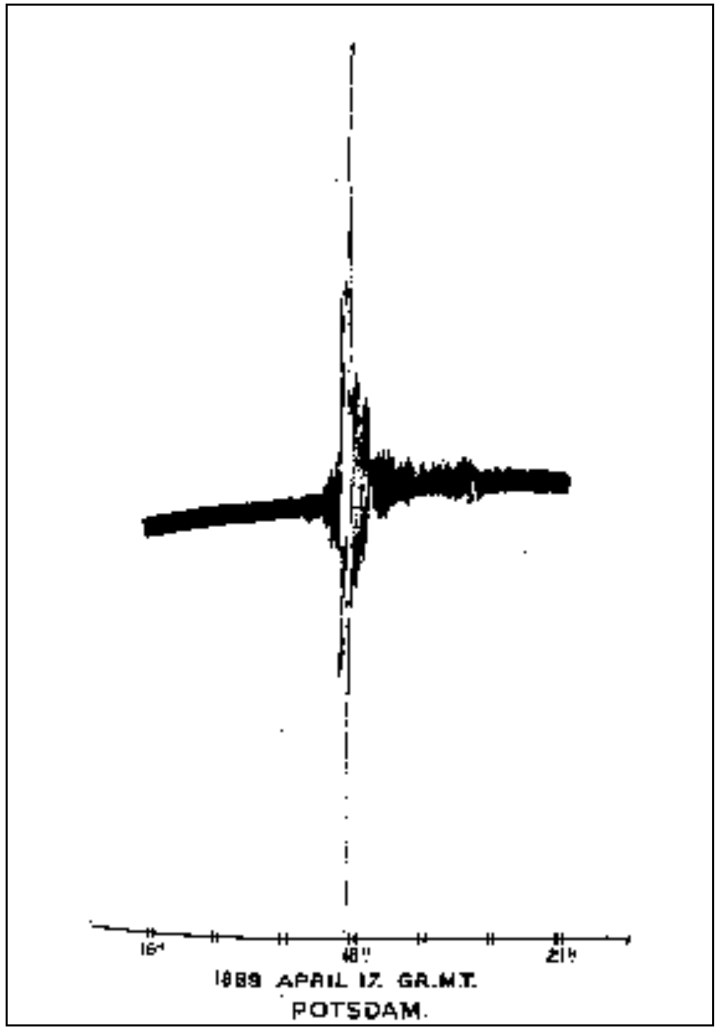
Overview

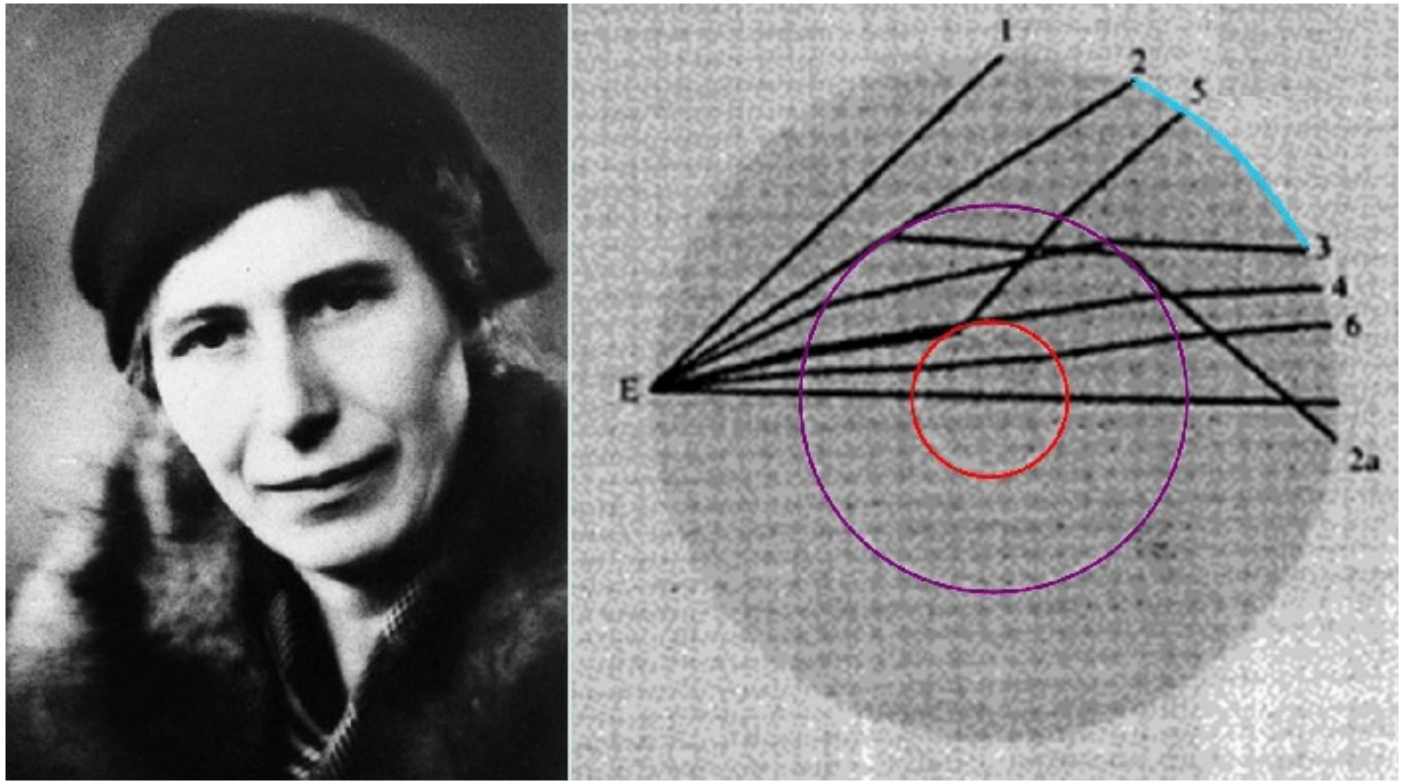
1. Advancements in large-scale, high-precision, deep-magnitude earthquake catalog production and seismic monitoring

2. Experiences from California, Italy, and Axial Seamount

3. Discoveries, opportunities, challenges

First Teleseismic Seismogram in 1889

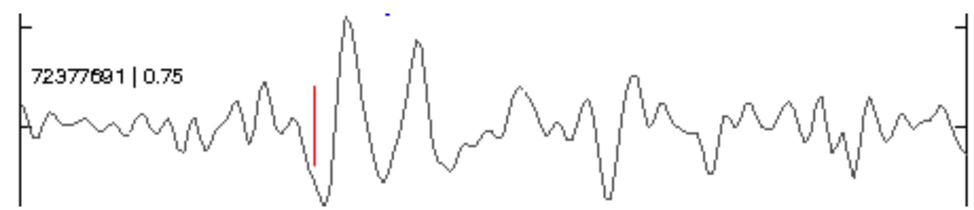




Discovery, by Inge Lehmann in 1936, that inner core is solid

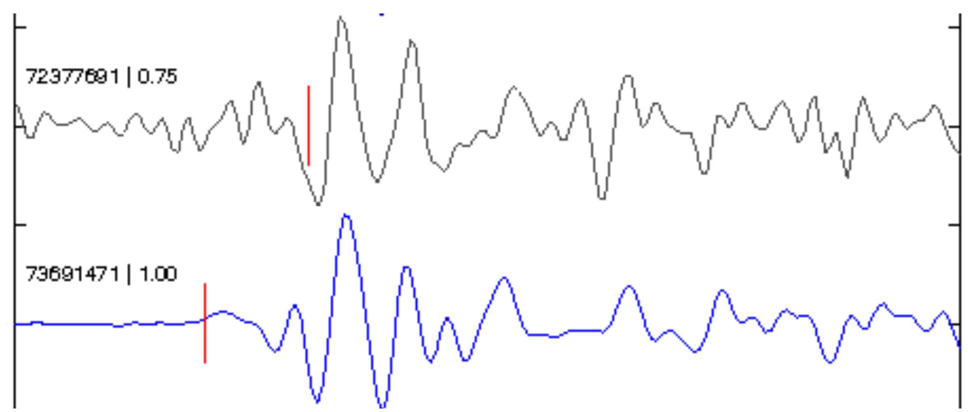
Earthquake monitoring

1. Detection
2. Arrival time picking
3. Association
4. Discrimination
5. Location, magnitude



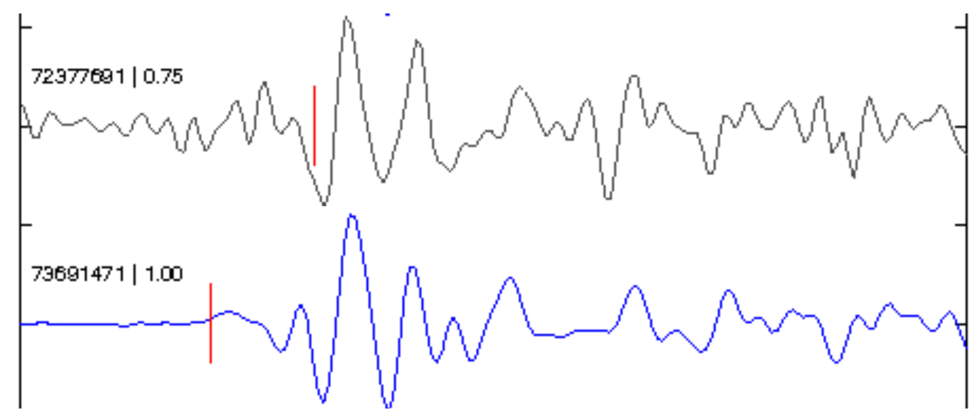
Earthquake relocation

1. Detection
2. Arrival time picking
3. Association
4. Discrimination
5. Location, magnitude
6. Relative time measurement
7. Relative location, magnitude
8. Template matching



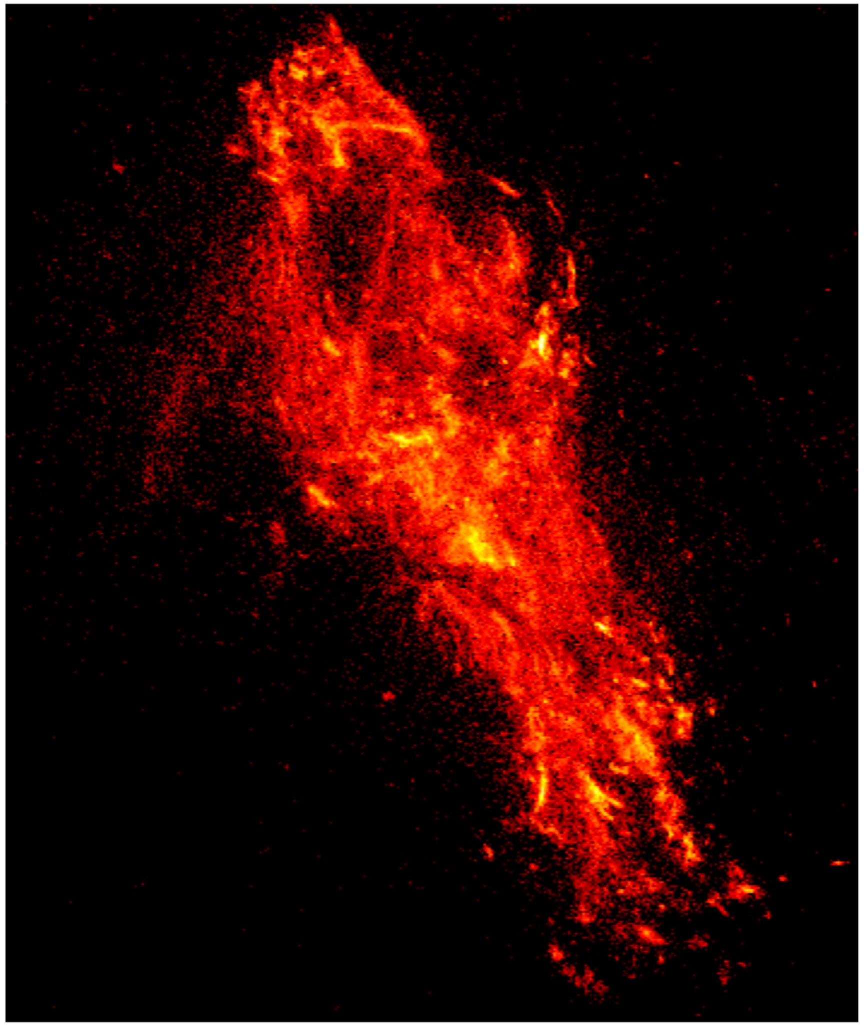
Earthquake monitoring with machine-learning

1. Detection
2. Arrival time picking
3. Association
4. Discrimination
5. Location, magnitude
6. Relative time measurement
7. Relative location, magnitude
8. Template matching



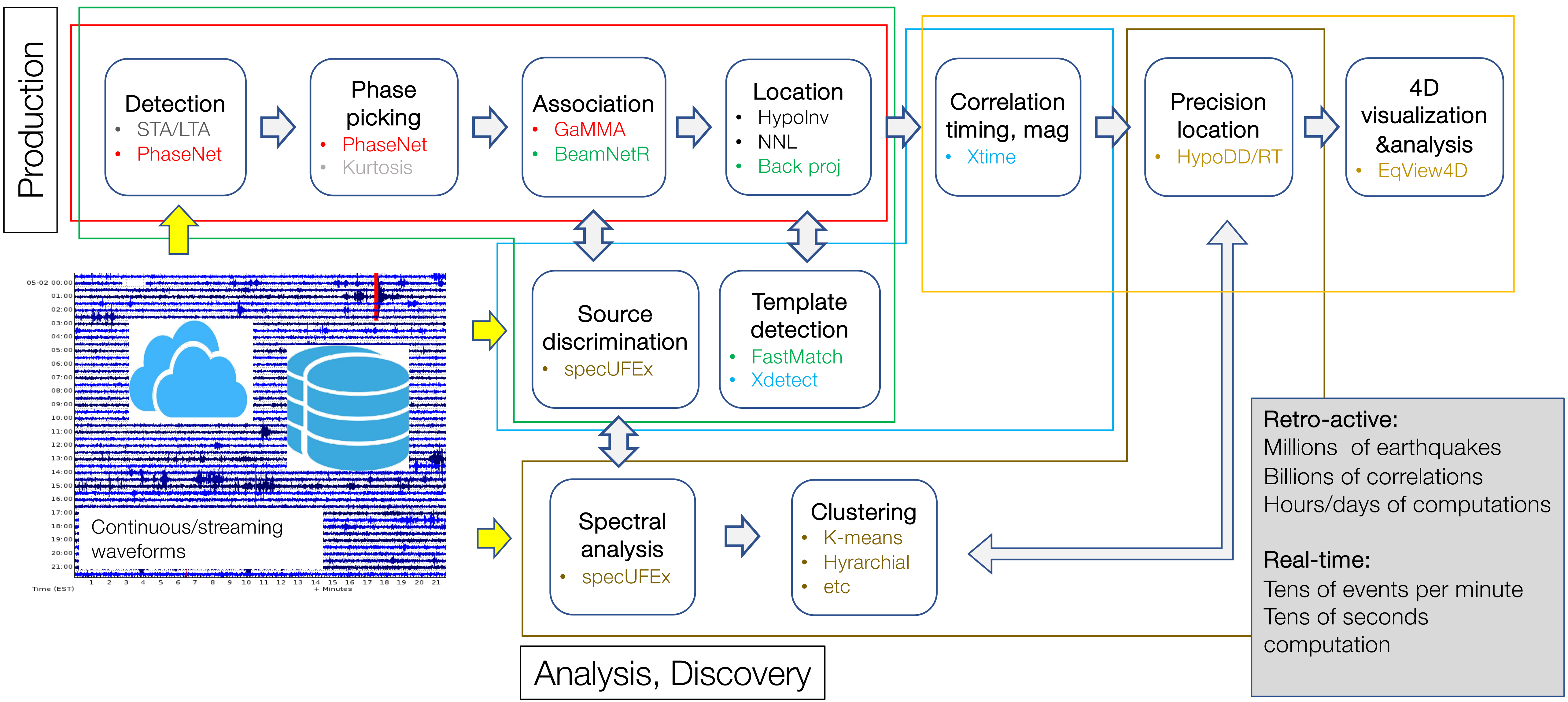
Earthquake monitoring with machine-learning

1. Detection
2. Arrival time picking
3. Association
4. Discrimination
5. Location, magnitude
6. Relative time measurement
7. Relative location, magnitude
8. Template matching



Seismic Computational Platform for Empowering Discovery.

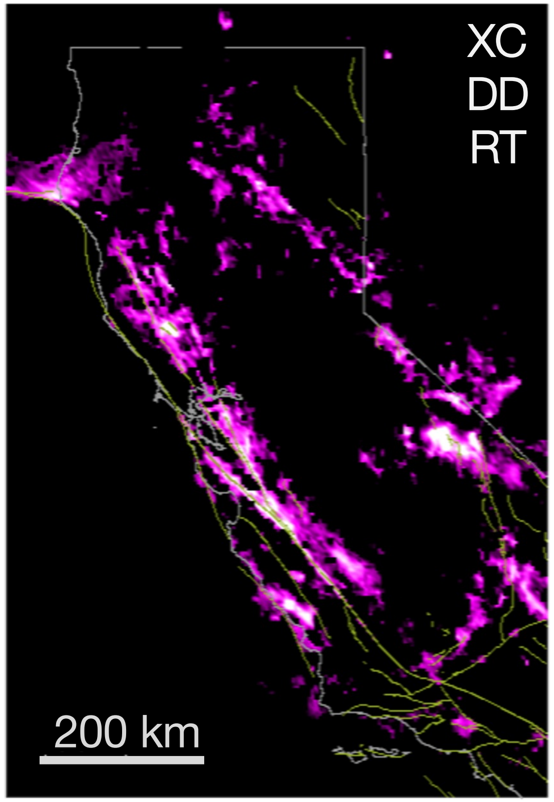
Modules and workflows for retro-active and real-time earthquake catalog production



High-resolution (10s of m) catalog production at scale

California

Waldhauser et al 2008+



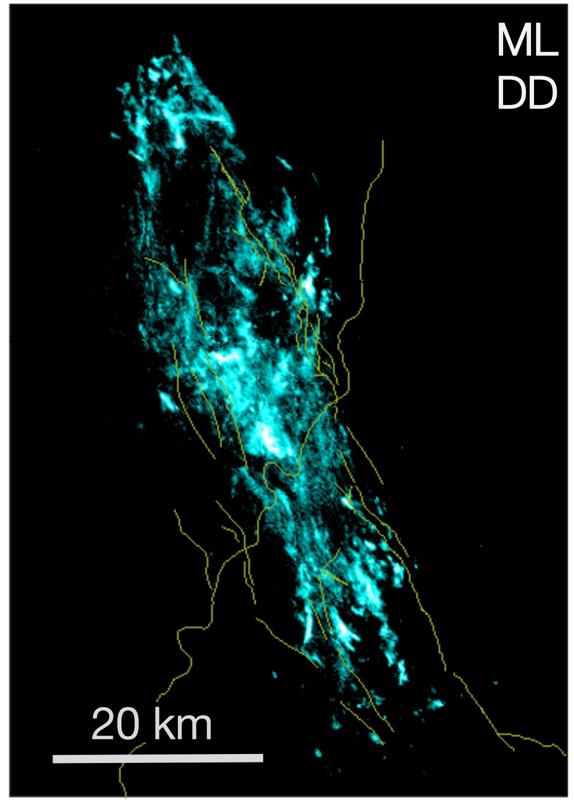
https://nocalDD.ldeo.columbia.edu

100s of kms
1,000 stations
1 million eqs in 40 years

→

Central Italy

Tan et al., 2021

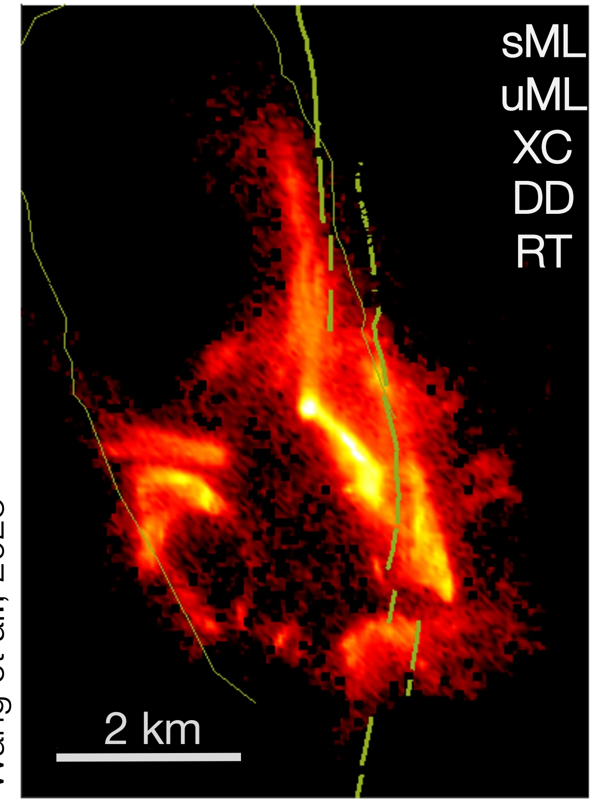


10s of kms
130 stations
1 million eqs in 1 year

→

Axial Seamount

Wang et al., 2023

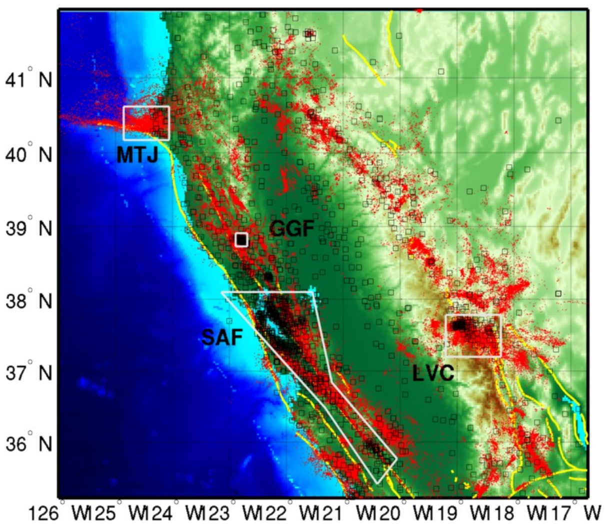


https://axialDD.ldeo.columbia.edu

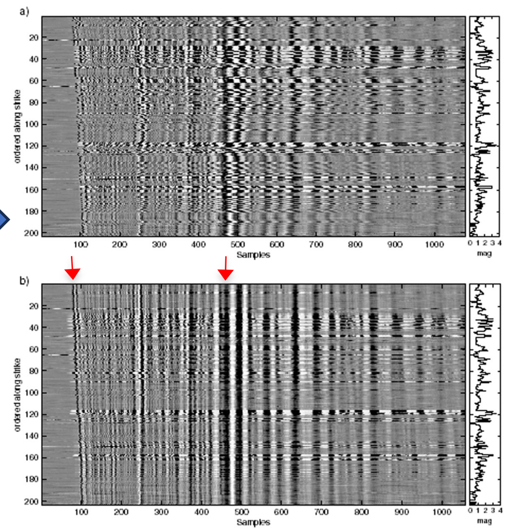
kilometers
7 stations
150k+ eqs in days/weeks

Massive-Scale X-correlation and Relocation in California

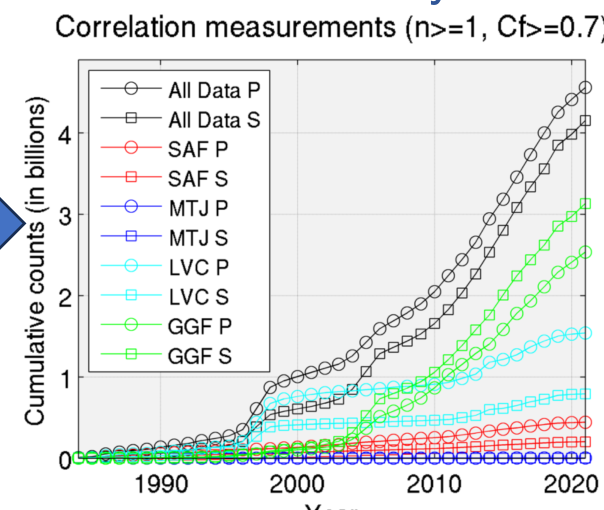
NCSN network



Cross-correlation



Correlation delay times

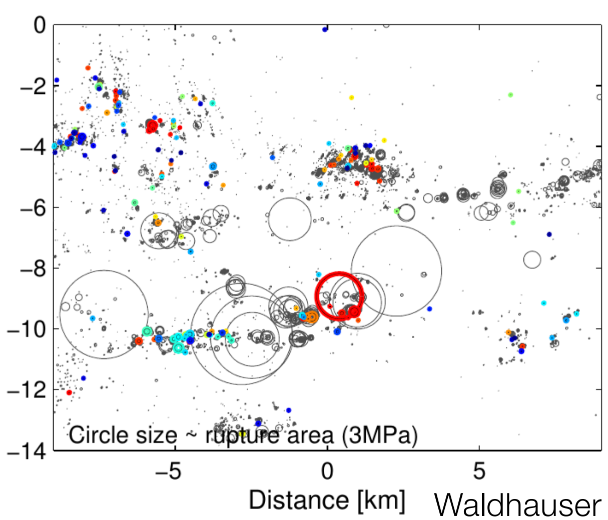


Schaff & Waldhauser (2005)

<u>NCSN archive 1984-2021:</u>

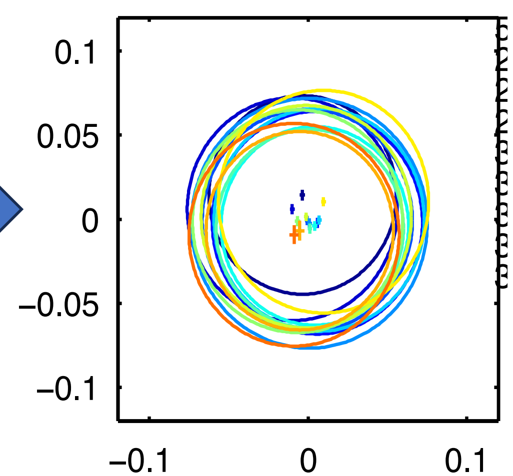
- 1,000 stations
- 1,000,000 earthquakes
- 50 million seismograms
- 20 million phase picks (mostly P)

High-resolution eq locations



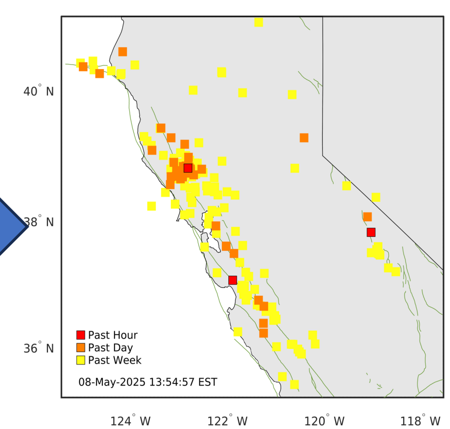
Waldhauser & Schaff (2008)

Repeating earthquakes



Waldhauser & Schaff (2021)

Real-time monitoring



Waldhauser (2009)

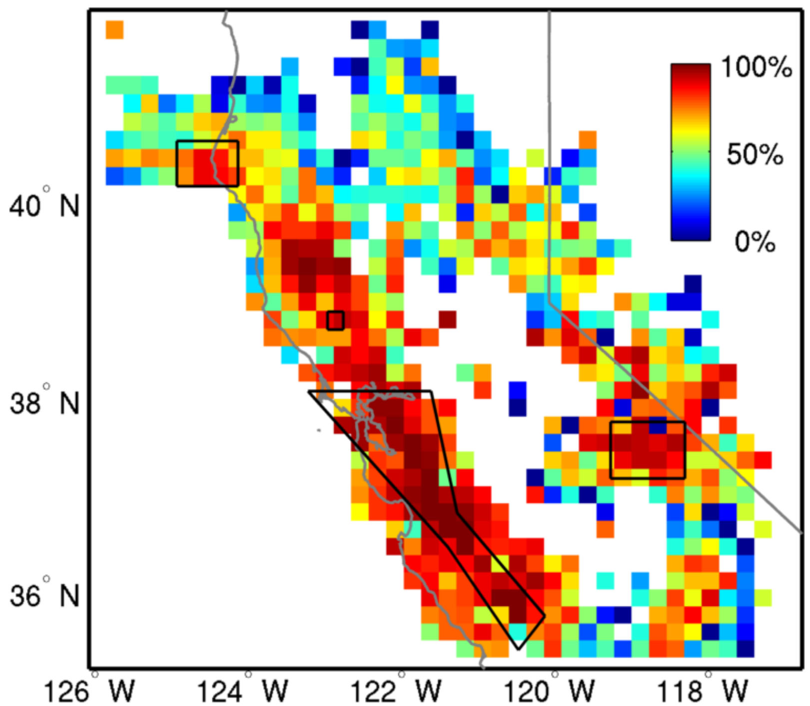
<u>Hypocenter relocation:</u>

- 100+ billion correlations
- 7.5 billion correlation delay times (Cf >0.7) (P&S!)
- 50 billion DD equations
- Resolution: 10s – 100s meters (10-100 X improvement)

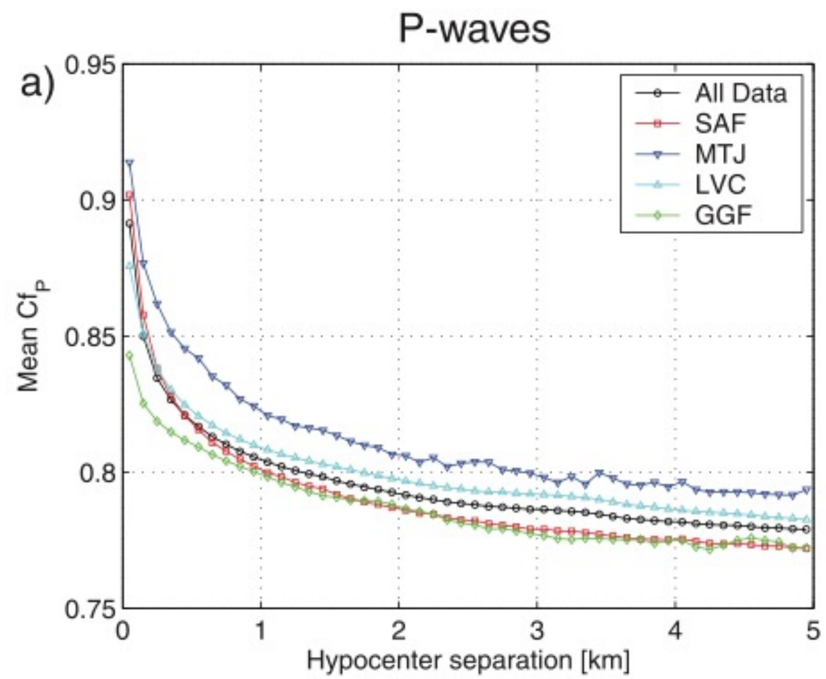
https://nocalDD.ldeo.columbia.edu

Waveform correlation measurements

Percentage of events with correlated seismograms

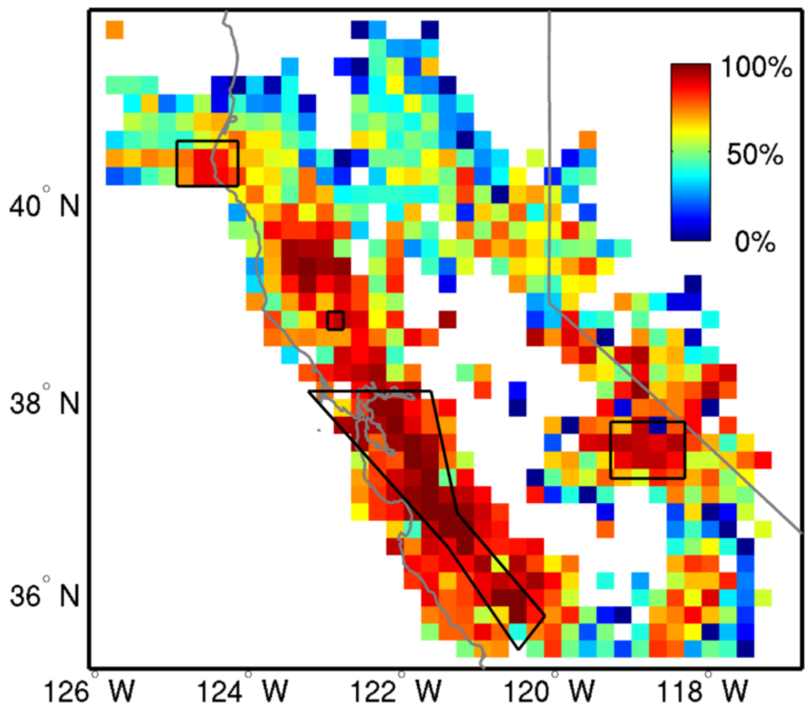


Decay of correlation coefficient with increasing hypocenter separation

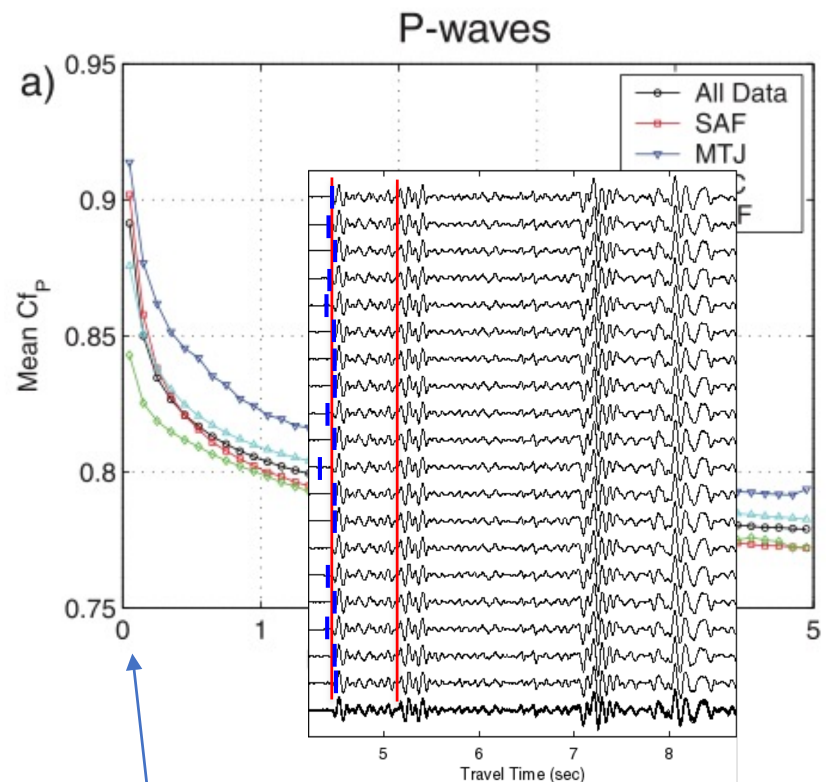


Waveform correlation measurements

Percentage of events with correlated seismograms



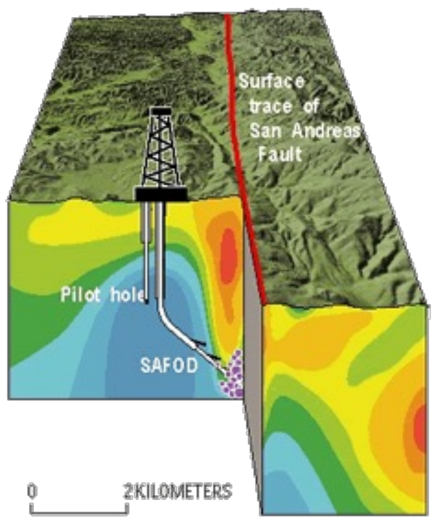
Decay of correlation coefficient with increasing hypocenter separation



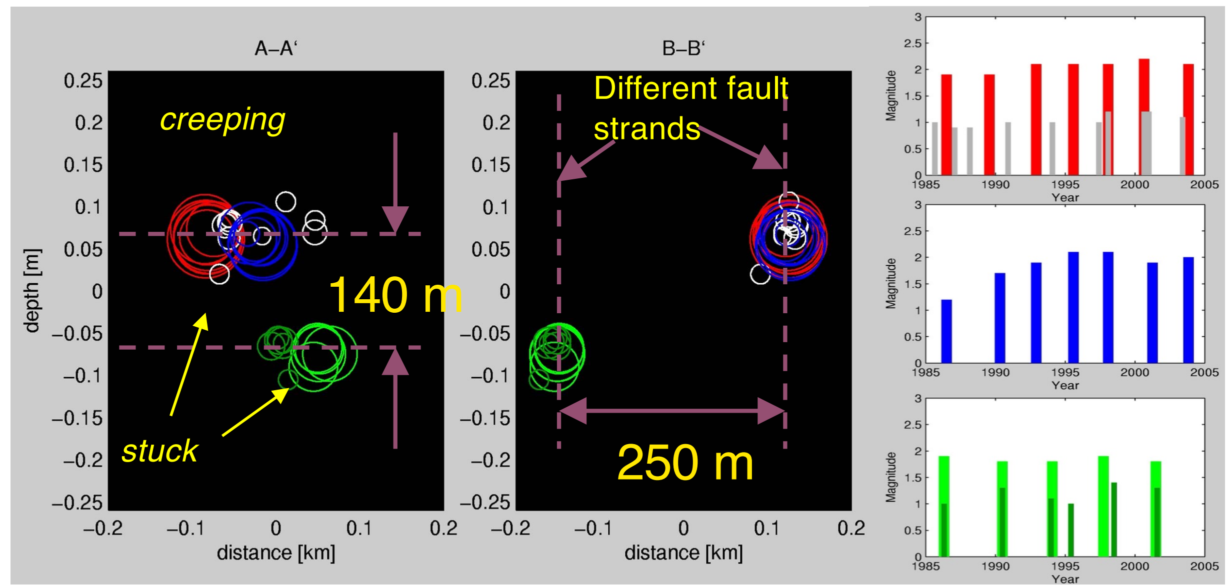
Repeating earthquakes

Repeating earthquakes in California

Repeating earthquakes rupture the same fault surface with similar magnitudes and focal mechanisms, thus generating close to identical seismograms.



In plane view | Across strike | Time vs magnitude



- Thought to represent stuck asperity in an otherwise creeping fault.
- Potential to improve hazard assessment (Field et al., 2014), earthquake forecast (Zechar et al., 2012), and seismic monitoring capabilities.
- Increasingly important role in the study of fault processes and behavior (recent review by Uchida and Bürgmann (2019).

Search for repeating earthquakes

Comprehensive, iterative, semi-automated search process:

- Highly correlated seismograms (over long windows)
- Similar magnitudes
- Co-located hypocenters (within source areas)

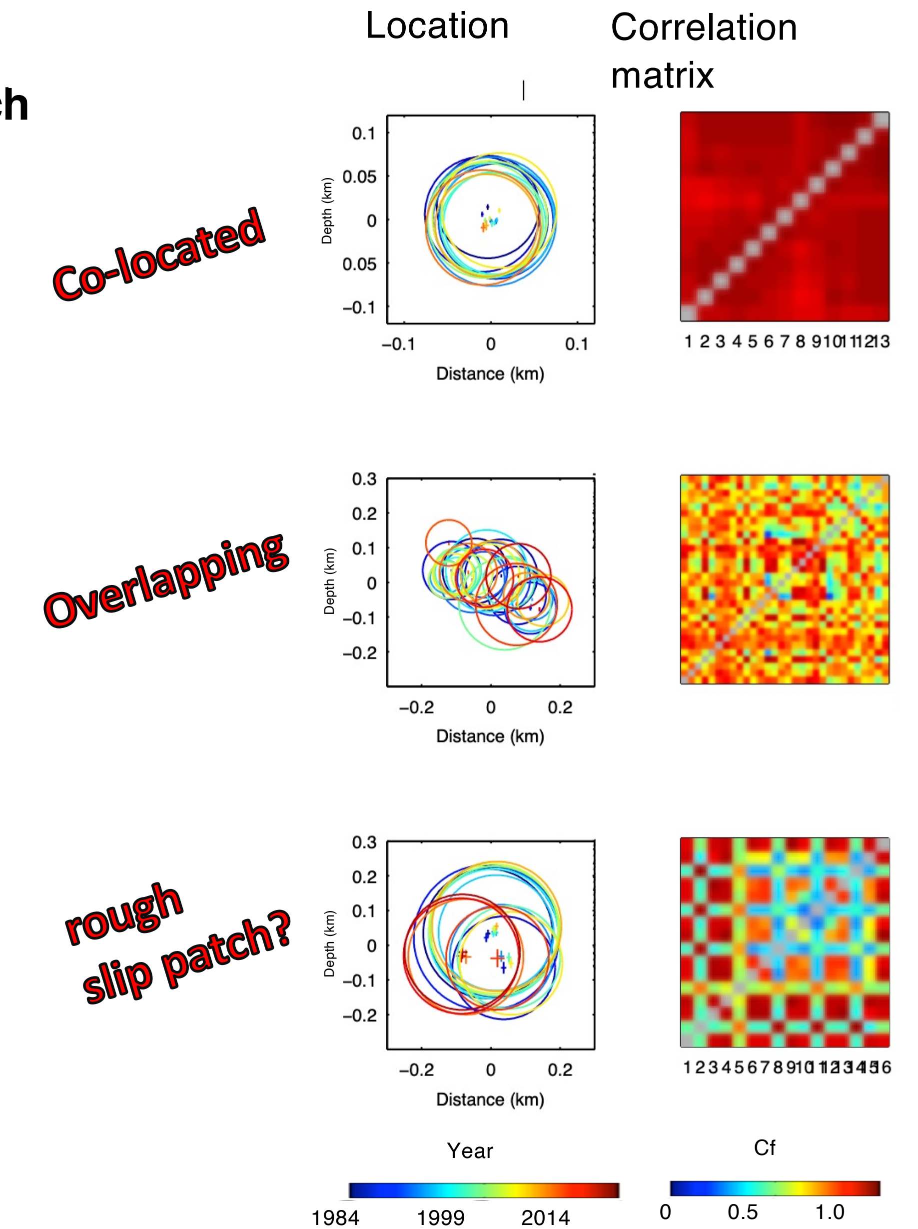
-> Isolated sequences only

Resulting catalog of repeating earthquakes:

7,713 sequences of a total of 27,675 events (1984-2014)

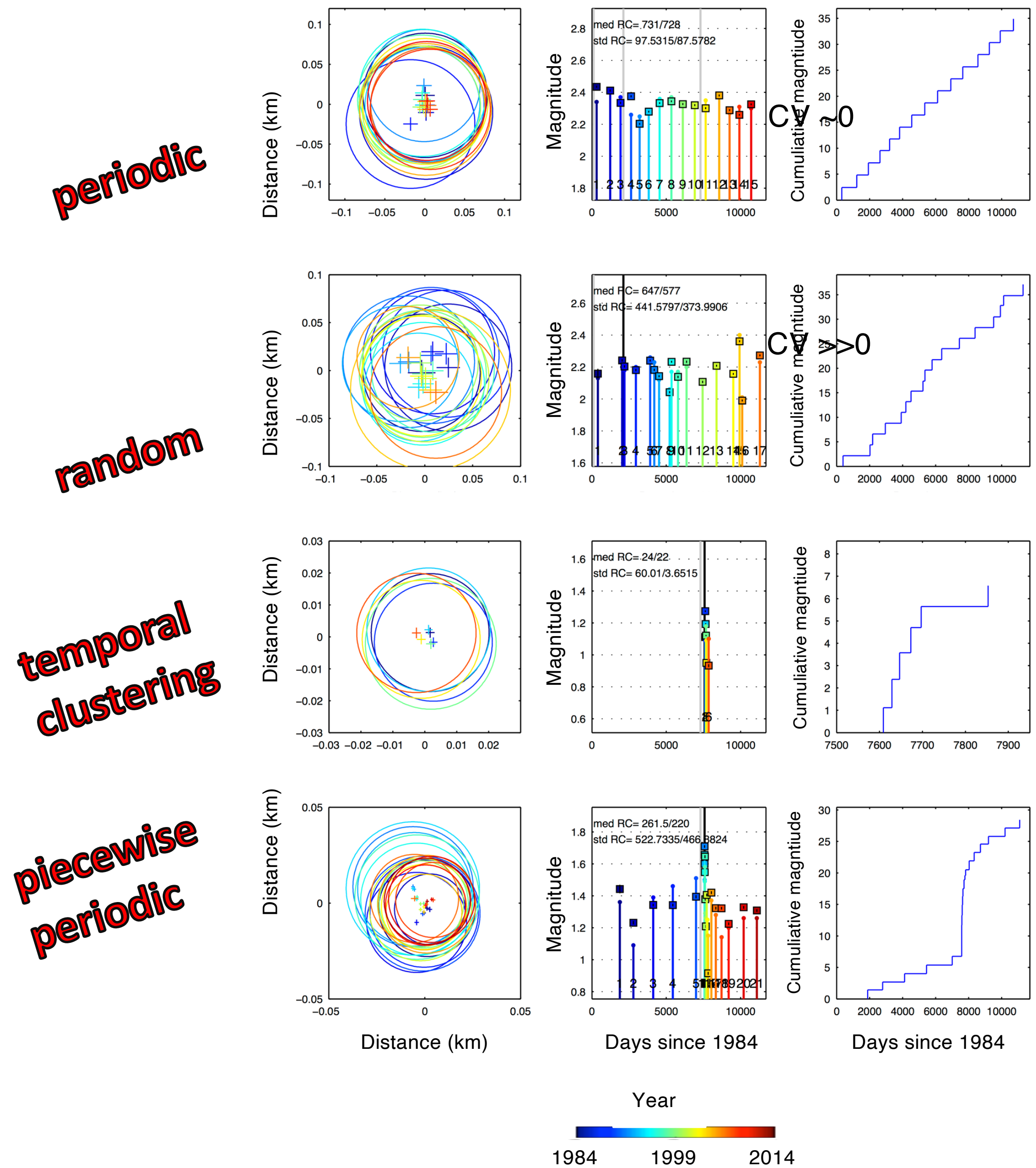
Additional measurements for each sequence:

- Differential magnitudes
- CV of recurrence interval
- Slip rates (following Nadeau & Johnson, 1998)



Waldhauser & Schaff (2021)

Temporal characteristics

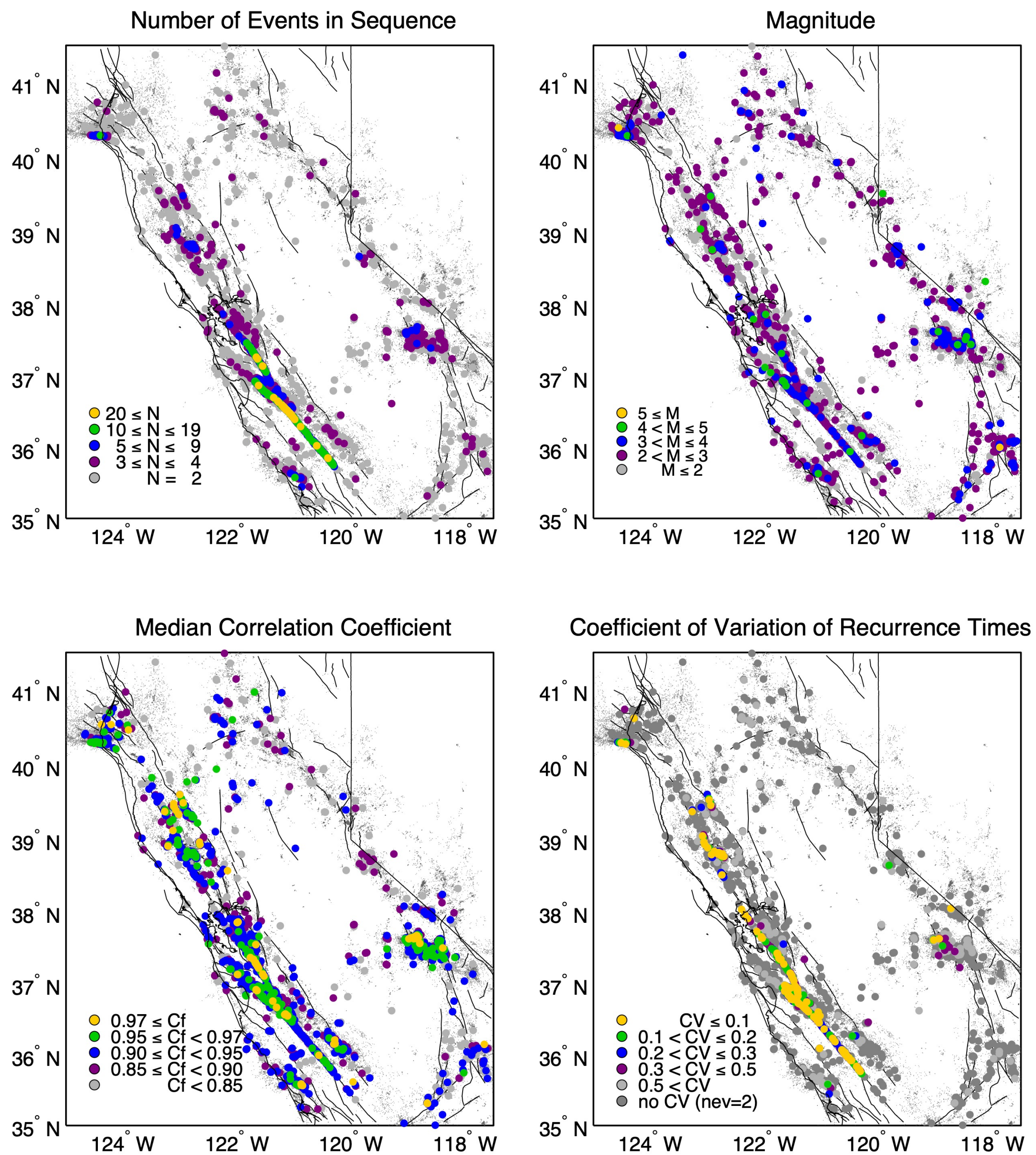


- Recurrence time coefficient of variation (CV):

$$CV = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(Tr_i - \overline{Tr}\right)^2}}{\overline{Tr}}$$

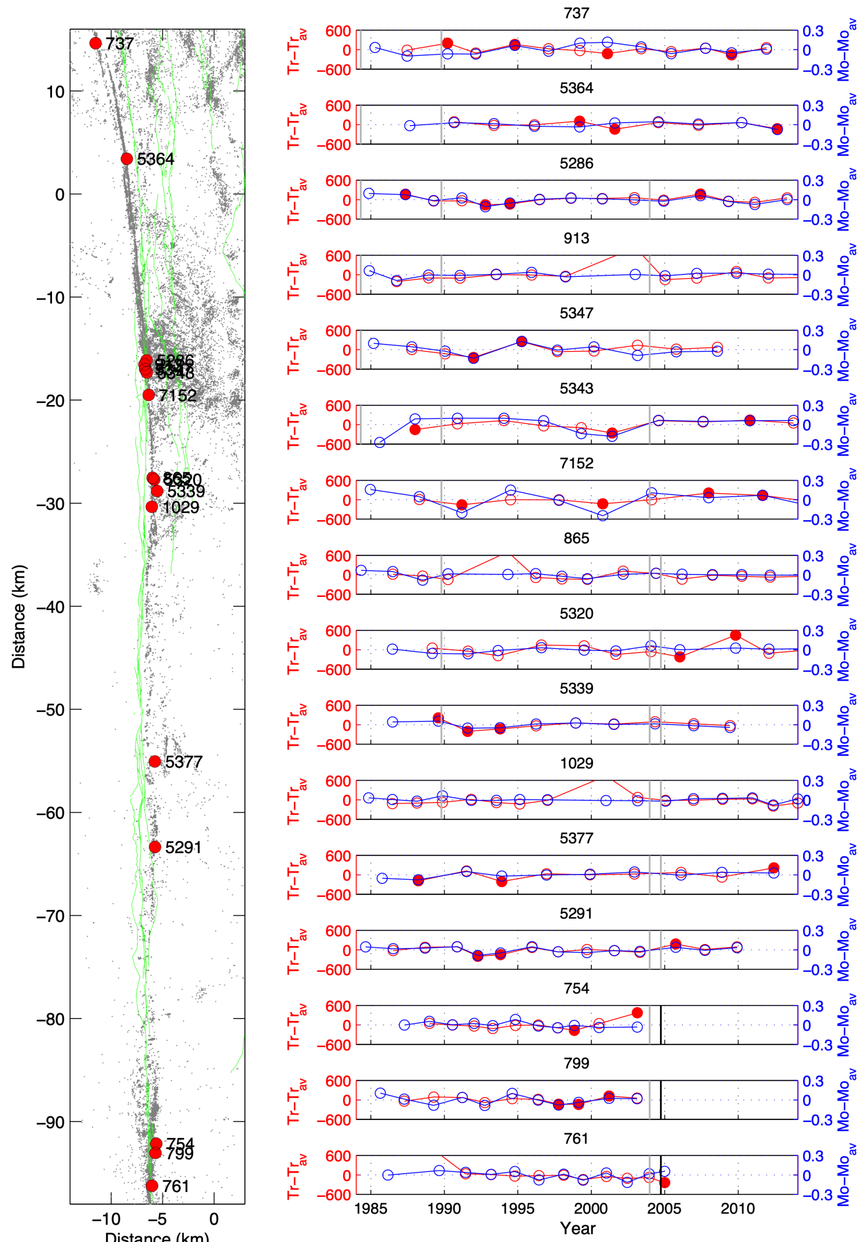
- periodic (CV~0)
- random (CV>>0)
- temporal clustering
- piecewise periodic

Geographical distribution



Waldhauser & Schaff (2021)

Moment vs. recurrence interval



Waldhauser & Schaff (2021)

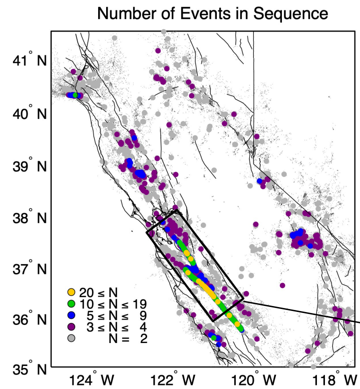
Event early 0>Tr/Tr$_{av}$ >0 Event late

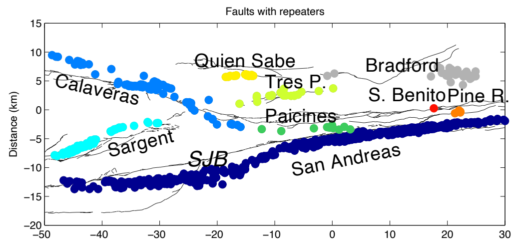
Event smaller 0>M/M$_{av}$>0 Event bigger

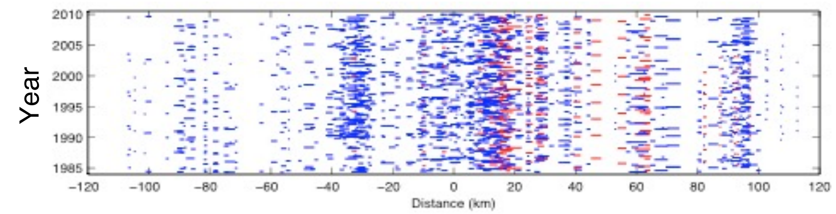
- When solid dots overlap, then early repeats have smaller magnitudes, late repeats have larger magnitudes

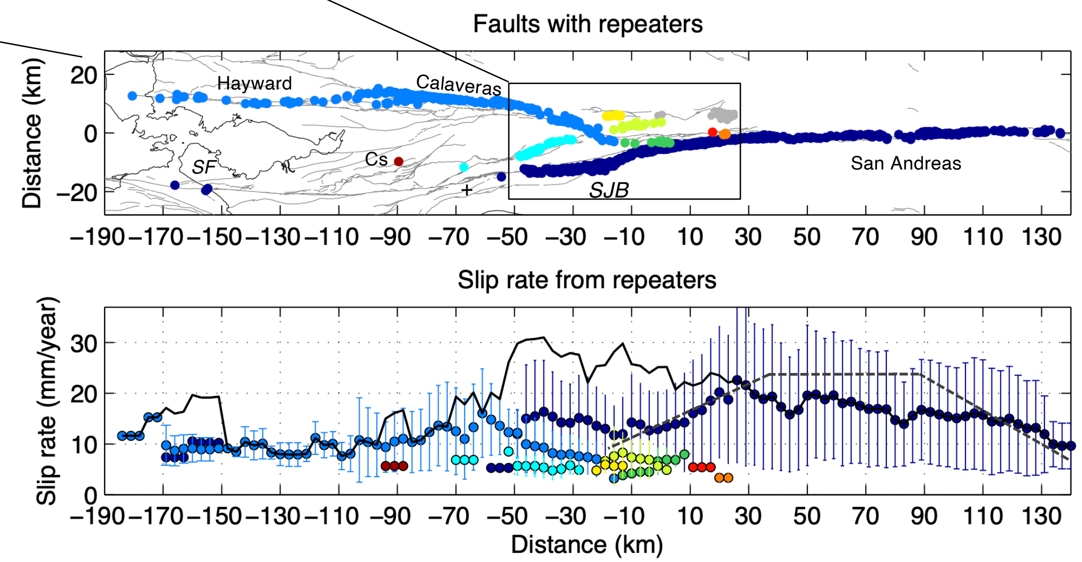
- Consistent with Rubinstein et al. (2012) we find no support of time-predictable model, in which the recurrence time scales with the size of the previous event.
- Evidence in support of the slip-predictable model, where slip in an earthquake scales with time since the last event, suggesting that knowing the recurrence time of one event lets you predict its size.

Fault slip rates and slip partitioning





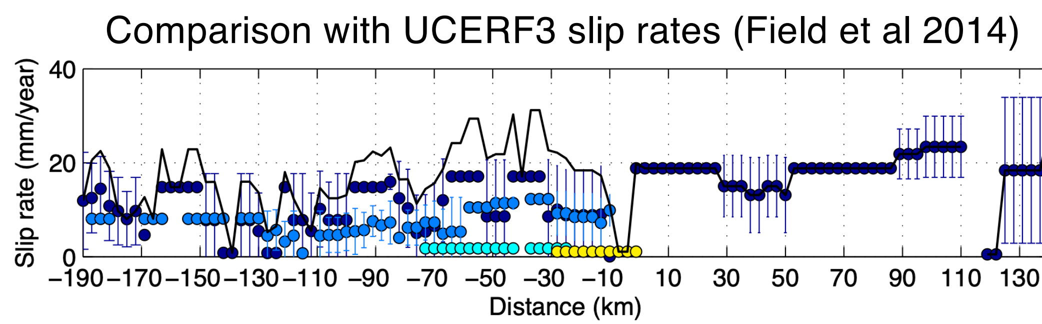




- Empirical slip rates following Nadeau & Johnson (1998):

$$S = \frac{\sum_{i=1}^{N} 10^{-2.46} {M_0^i}^{0.17}}{N}$$

$$SR = S / \left(\frac{\sum_{i=2}^{N} T_R}{N-1} \right)$$



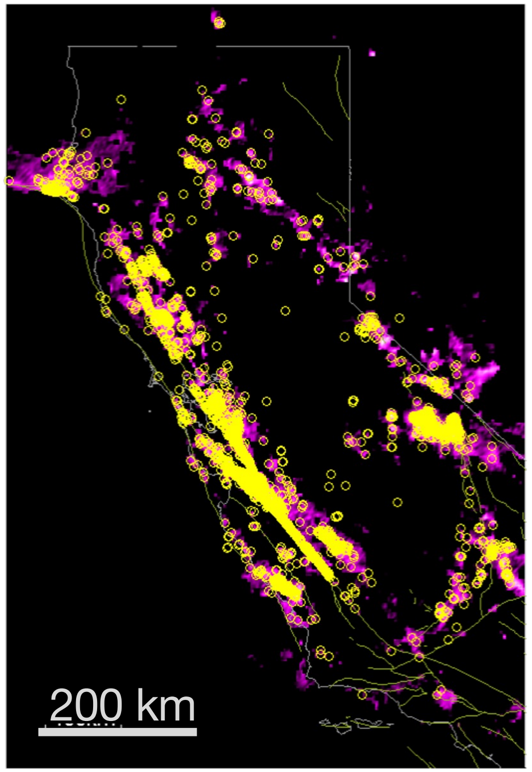
Waldhauser & Schaff (2021)

Precision monitoring in Northern California (NCSN)

NT XC DD DD_{RT} uML

Repeater catalog

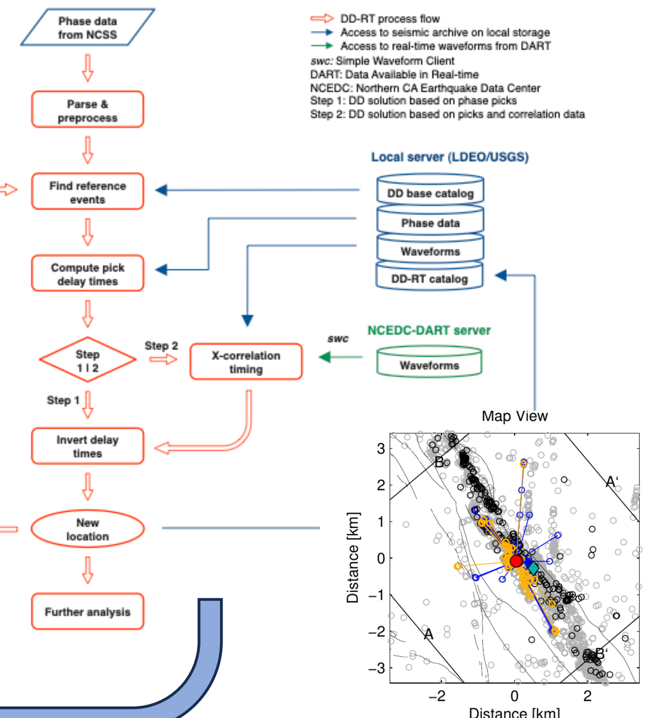
Waldhauser and Schaff (2021)



30…………years (1984-2014)
7,713……..sequences
27,675……repeaters

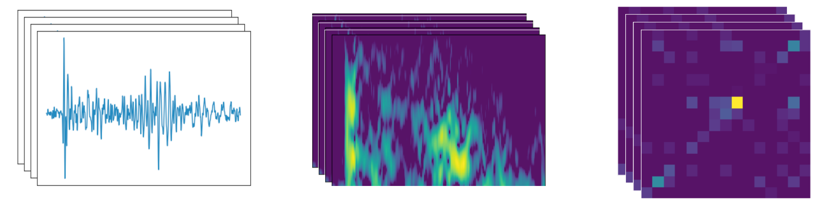
Real-time relocation

https://nocalDD.ldeo.columbia.edu



12……..years of operation
15 s….. per event
10 m…. location error
100 m…relocation shift

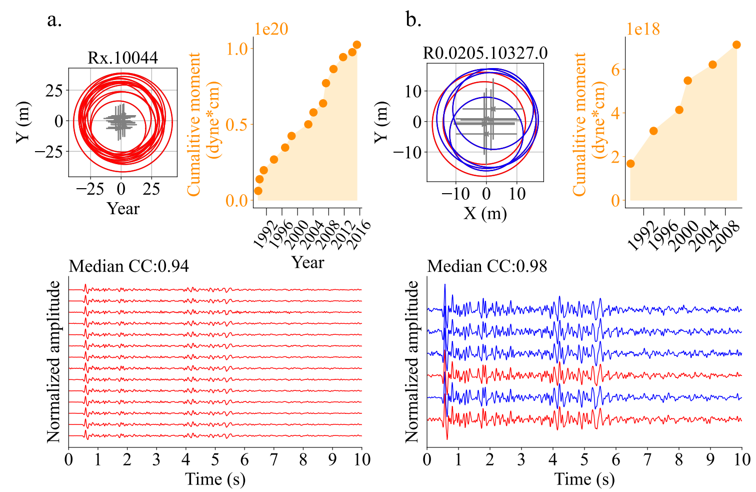
Unsupervised machine learning



waveforms → spectrograms → fingerprints

specUFEx Holtzman et al (2018)

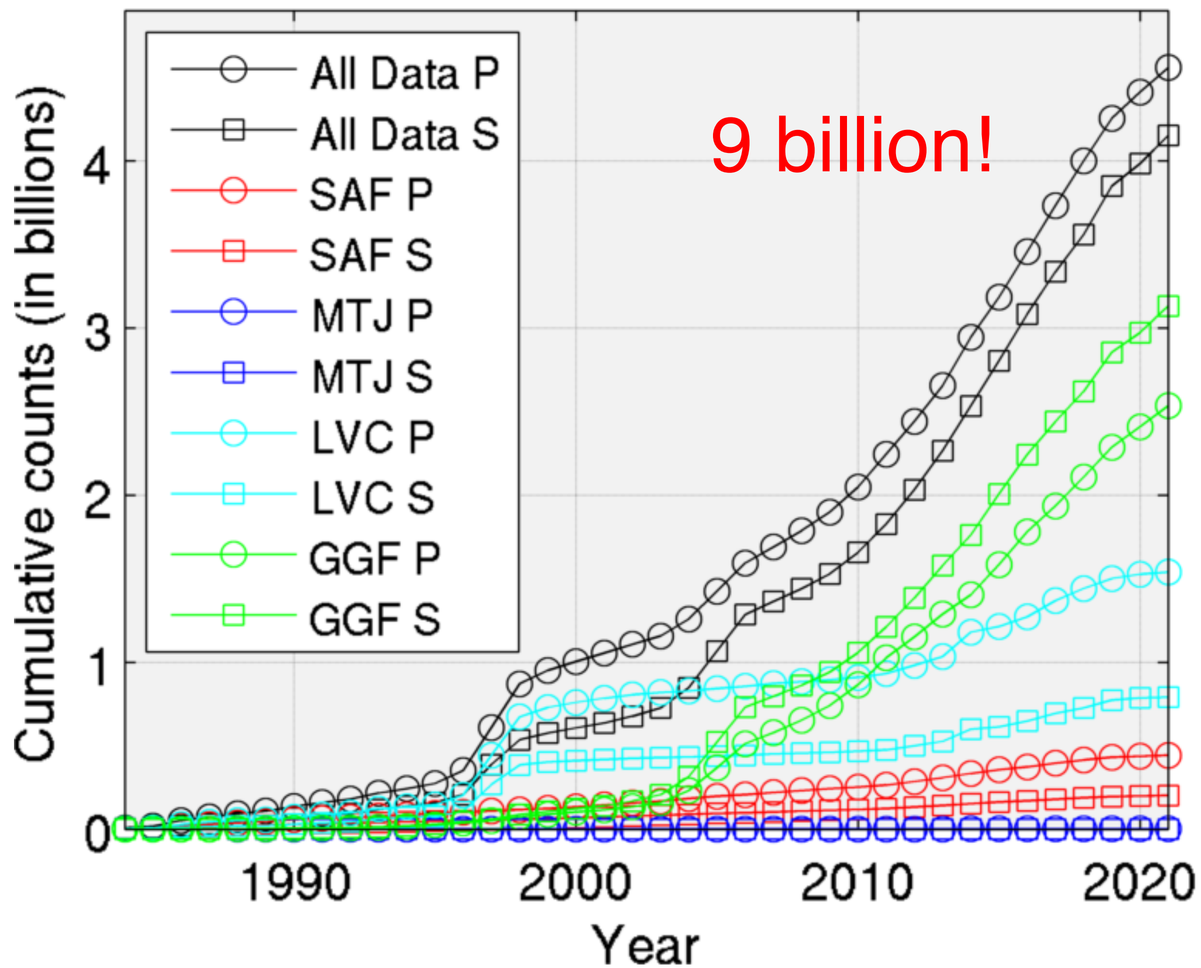
Repeating earthquakes



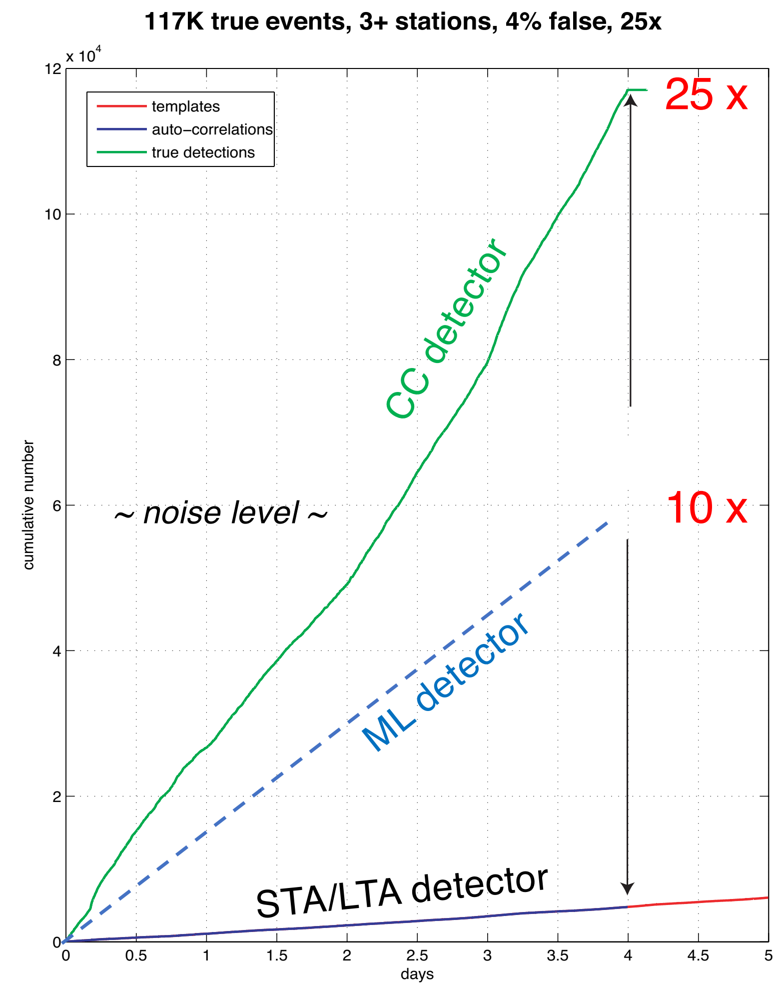
Sawi et al. (2023)

Big Data Problem

Growth in correlation measurements
(e.g., N California)

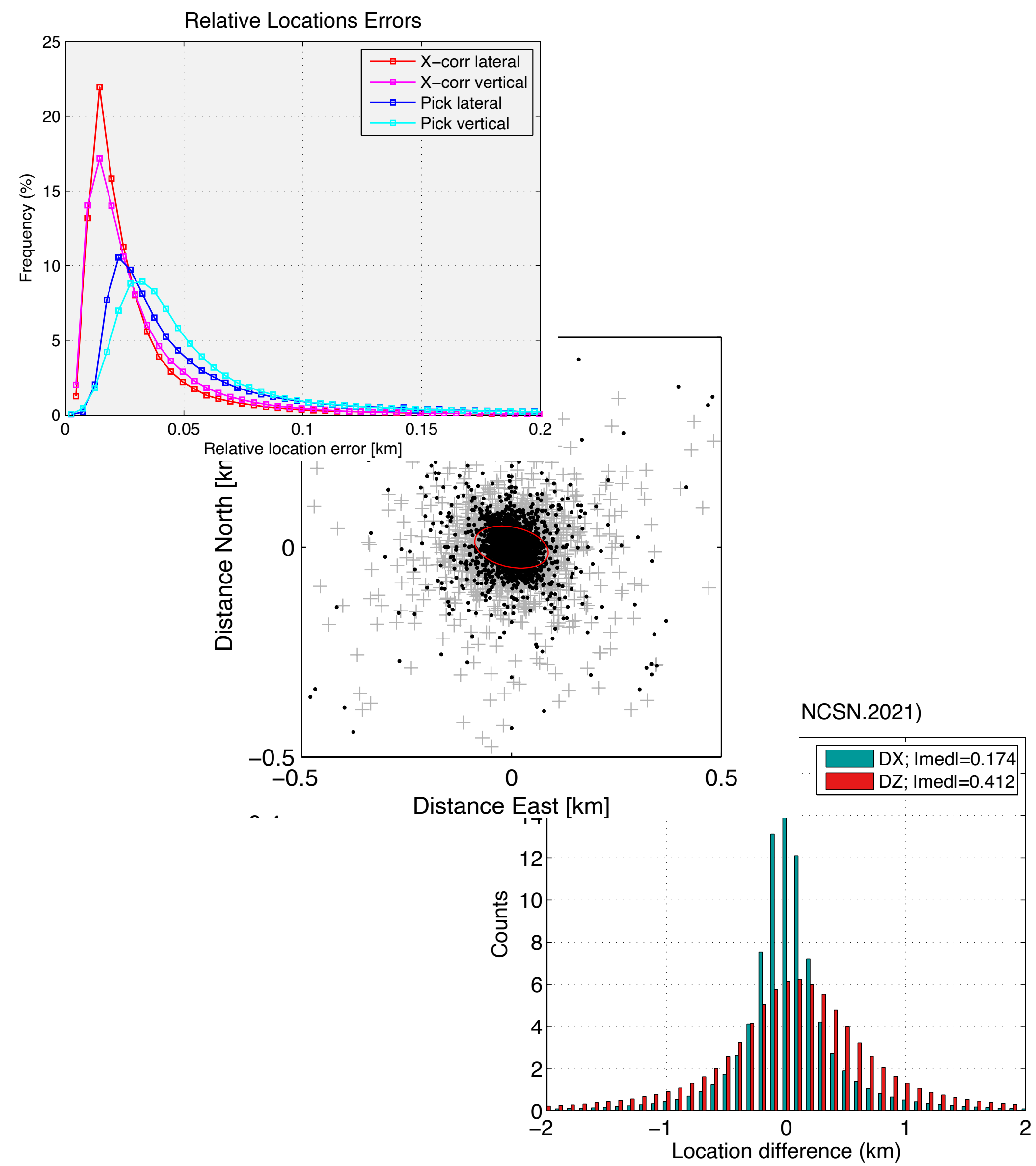


Growth in newly detected events
(e.g., Central Italy)



Evaluation of location robustness and uncertainties

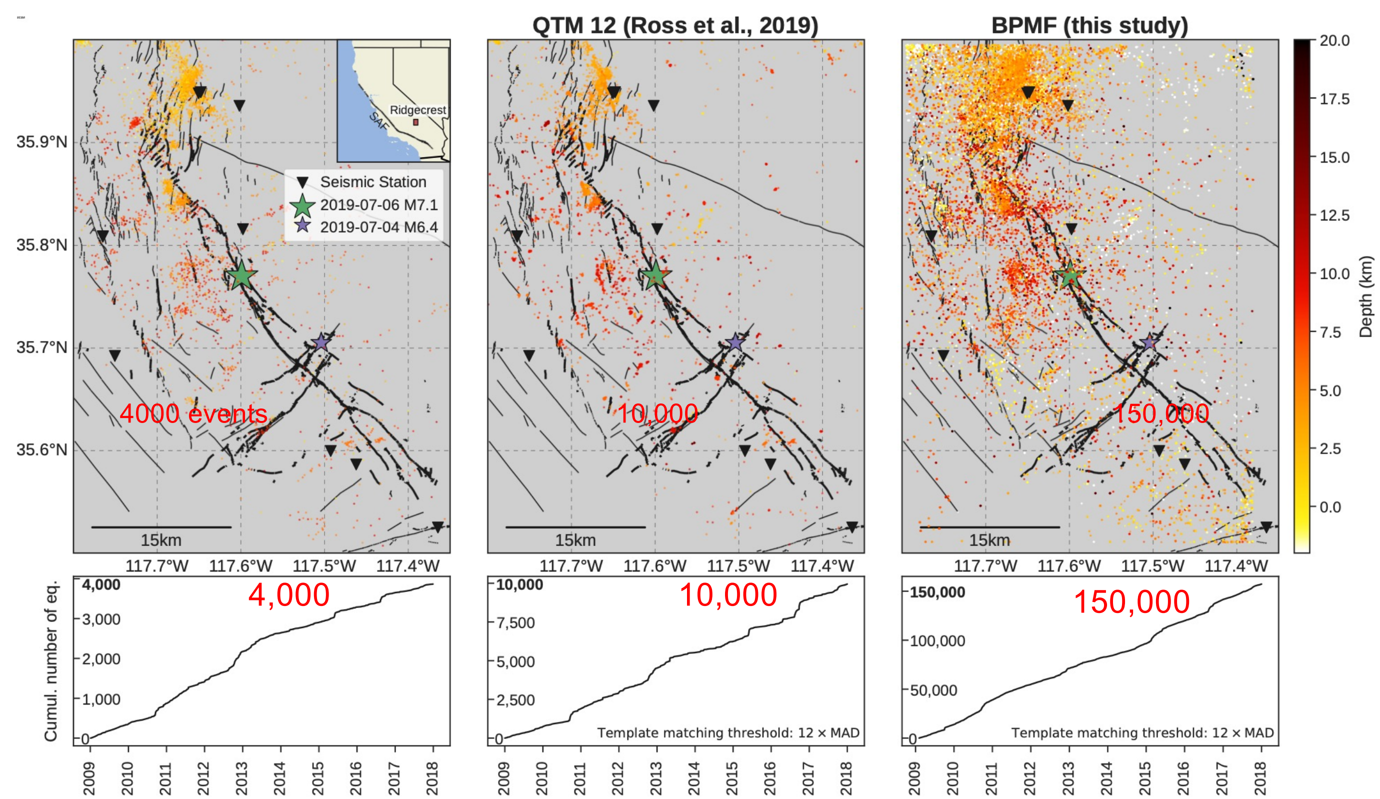
1. Bootstrap relative location errors and other statistical analysis.
2. Use known repeating events as ground truth: 95% of DD catalog events within 10 m of repeaters. 95% of NCSN catalog events within 500 m.
3. Shift in new locations within uncertainties of original locations: Std: dX = 0.7 km; dz = 1.4 km
4. Compare pick and x-corr delay times.



Deep magnitude catalog for 10 years before the 2019 M7.1 Ridgecrest event

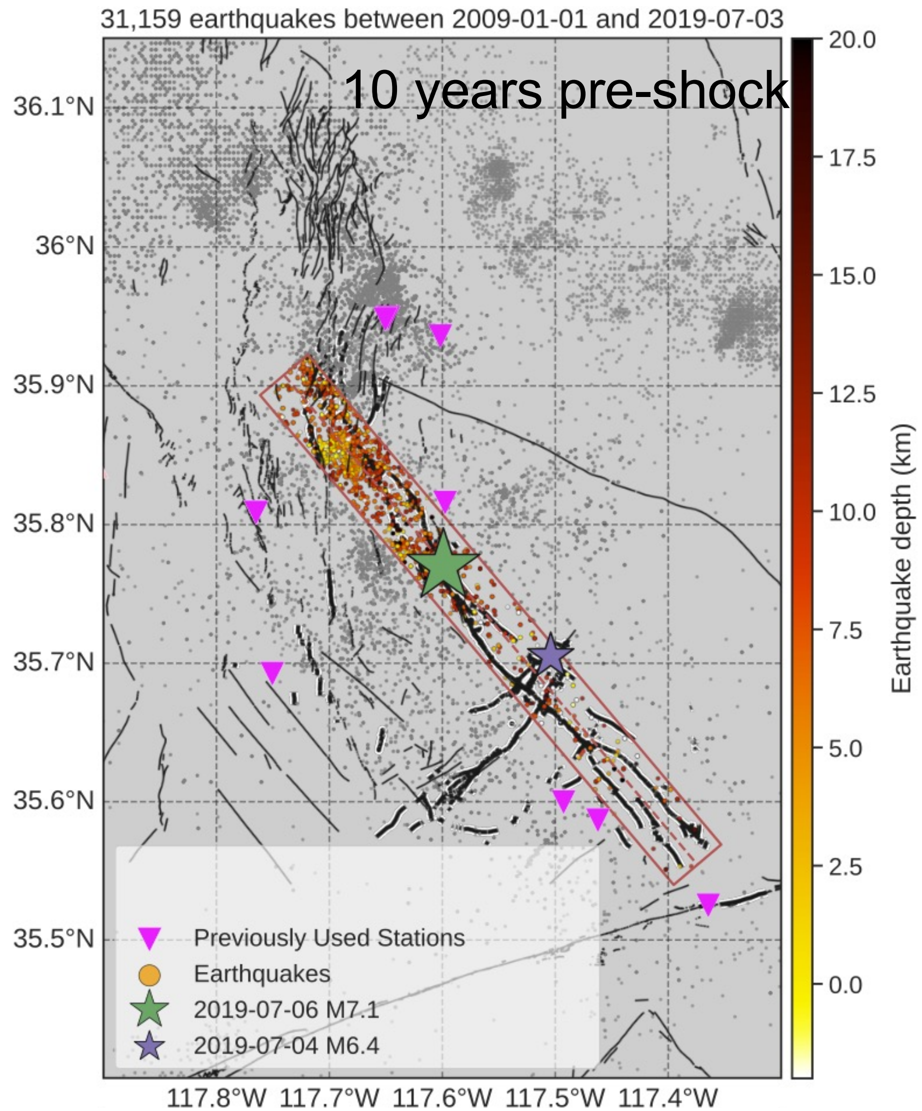
- Machine learning (PhaseNet) and template matching (FastMatch) increases the number of earthquakes in the SCSN catalog by a factor of 40!

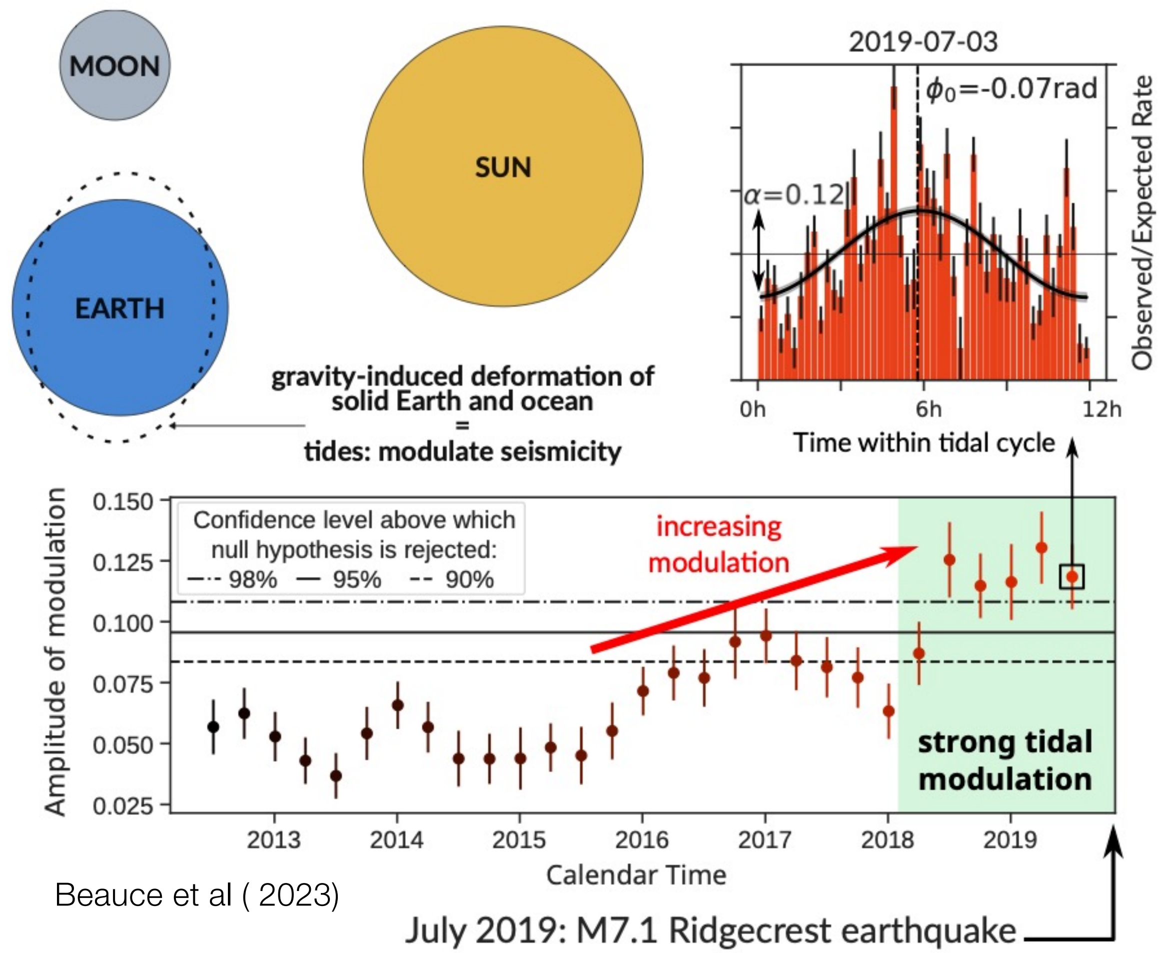
- New catalog shows that strength of tidal modulation of seismicity along the fault is continuously increasing starting about 1.5 years before the mainshock.



Beauce et al. (2023)

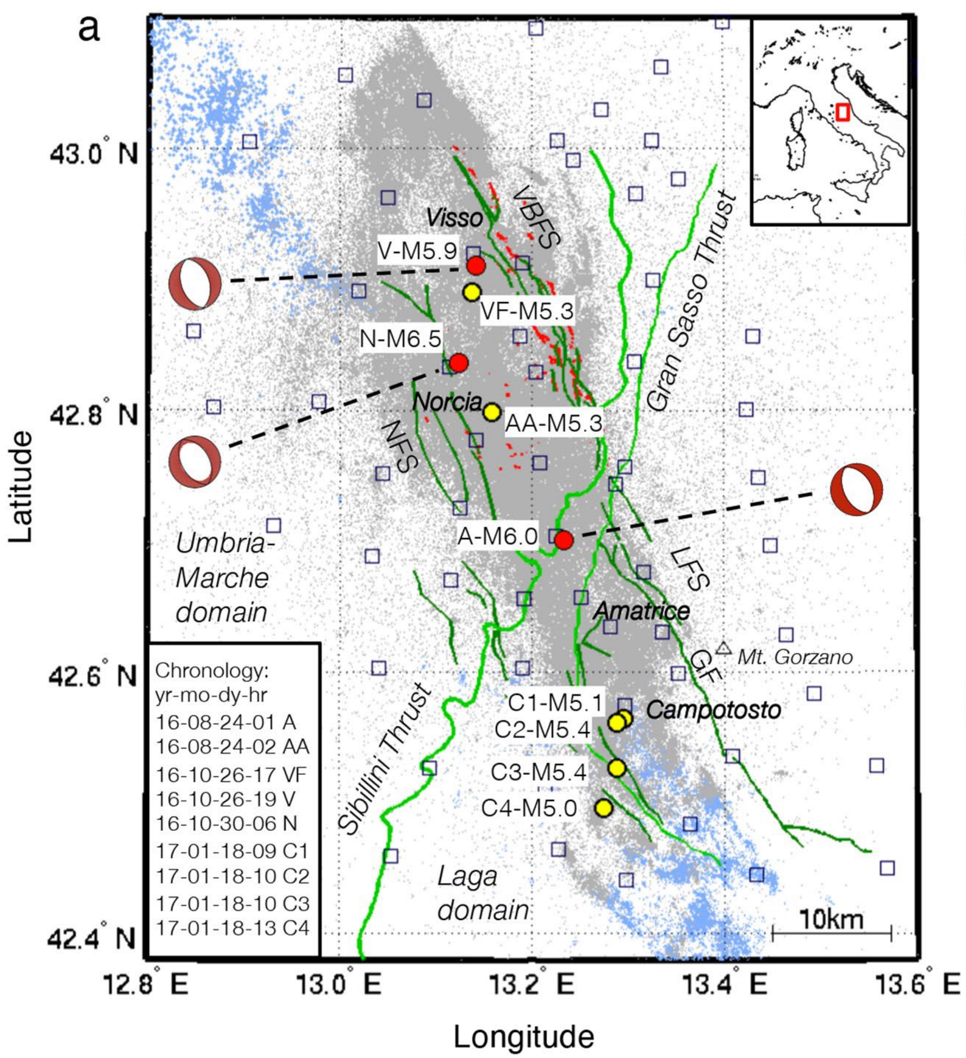
Increase in tidal modulation starting about 1.5 years before the mainshock

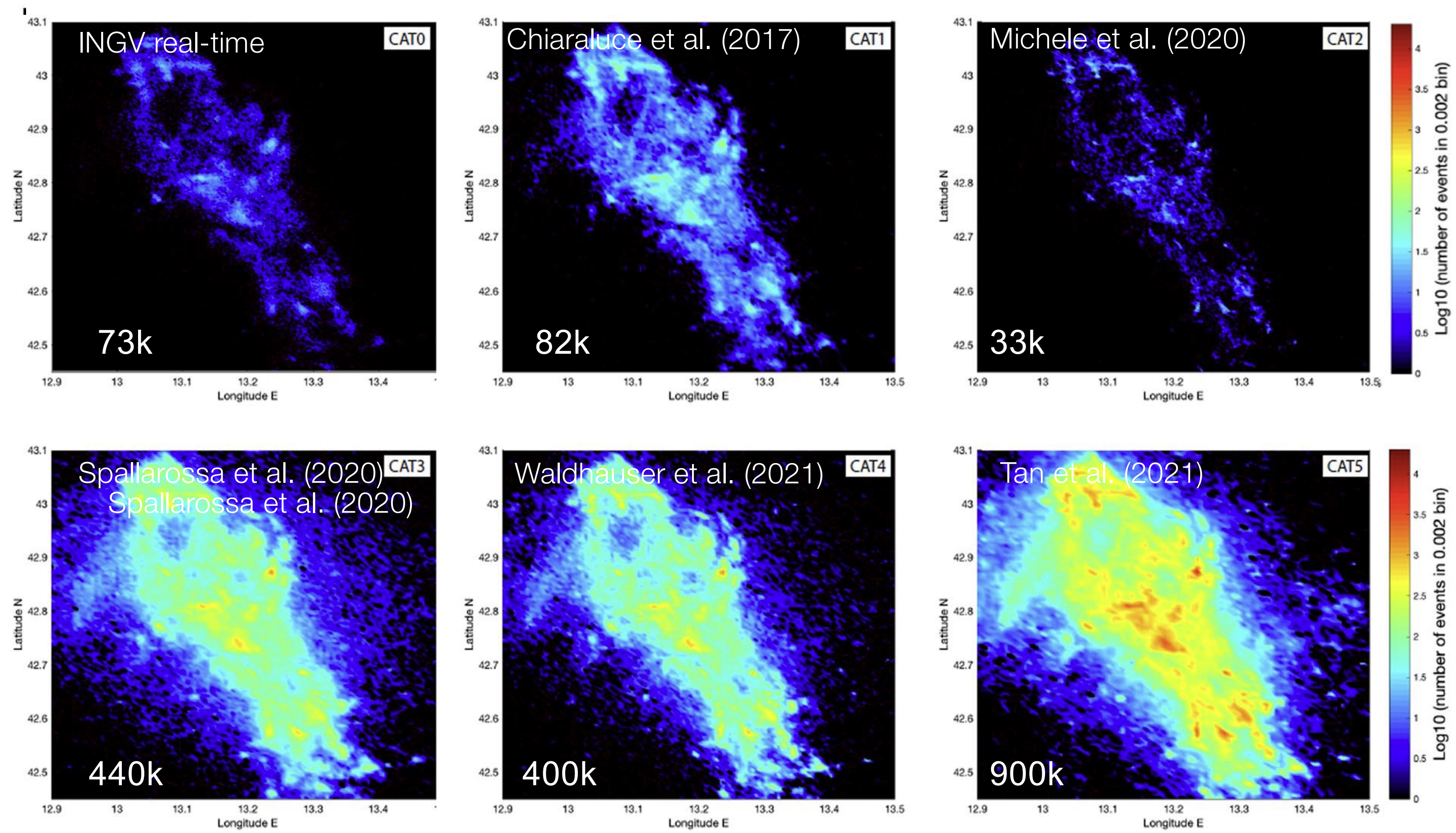




Beauce et al. (2023)

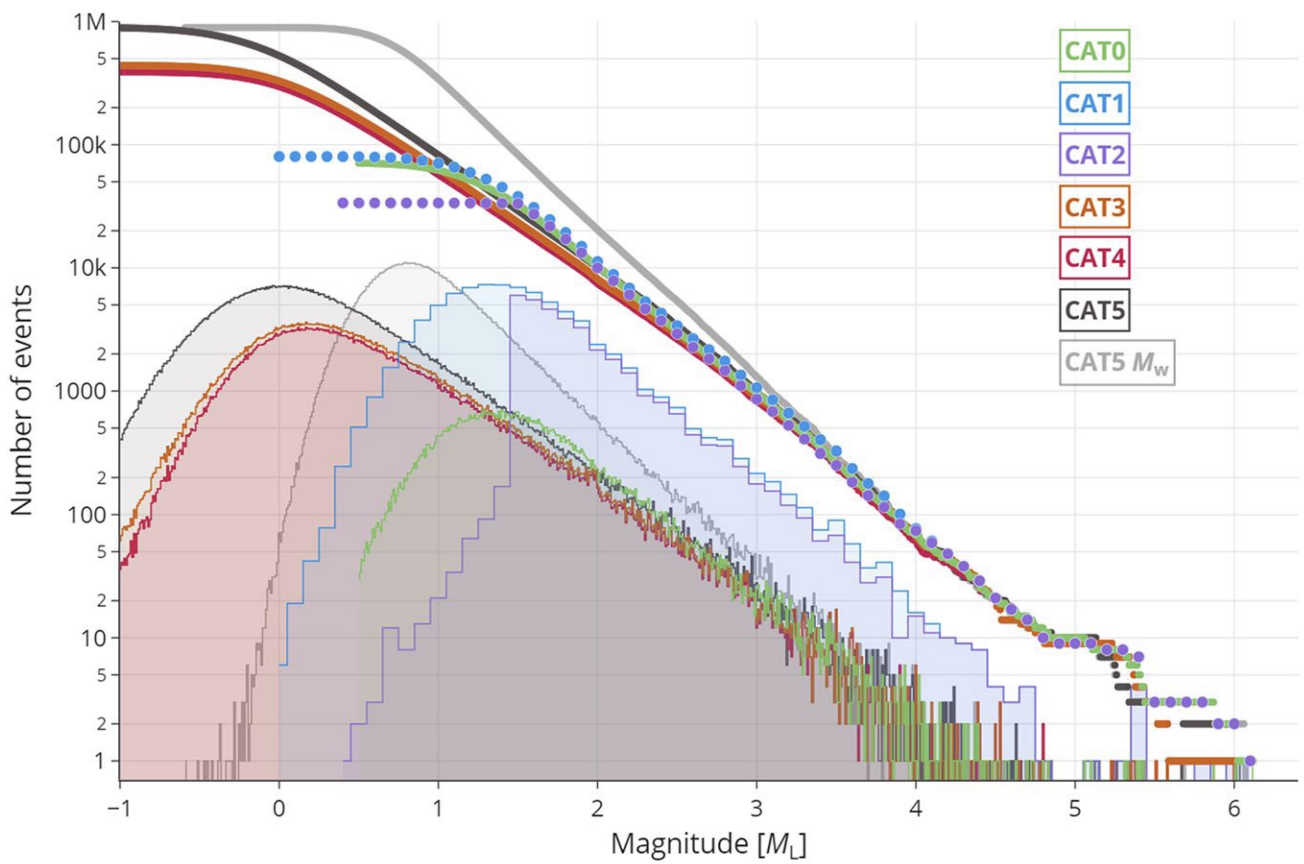
Amatrice sequence, Central Italy, 2016-2017
Evolution of catalogs





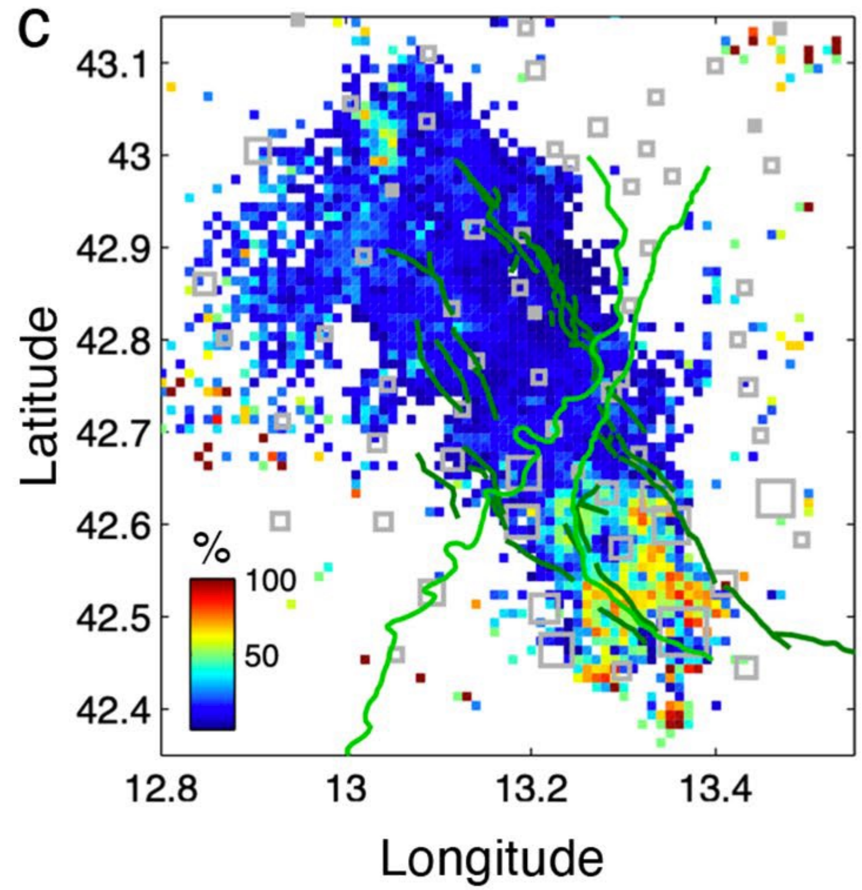
Chiaraluce et al (2022)

Evolution of catalogs: 2016-2017 Central Italy sequence



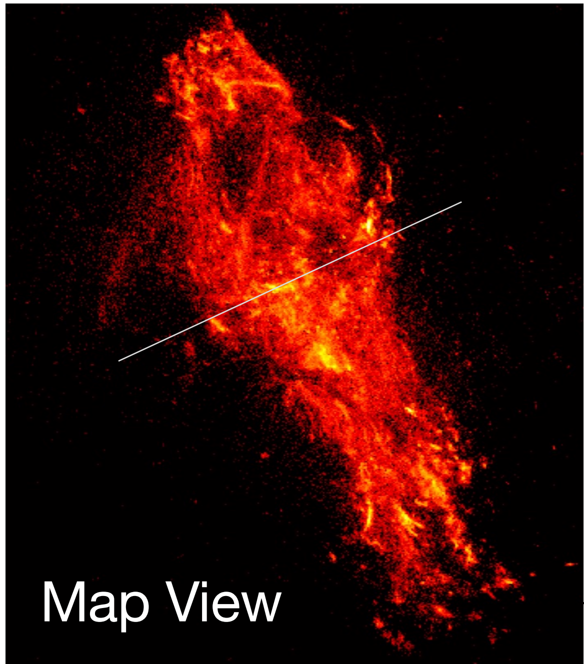
Chiaraluce et al. (2022)

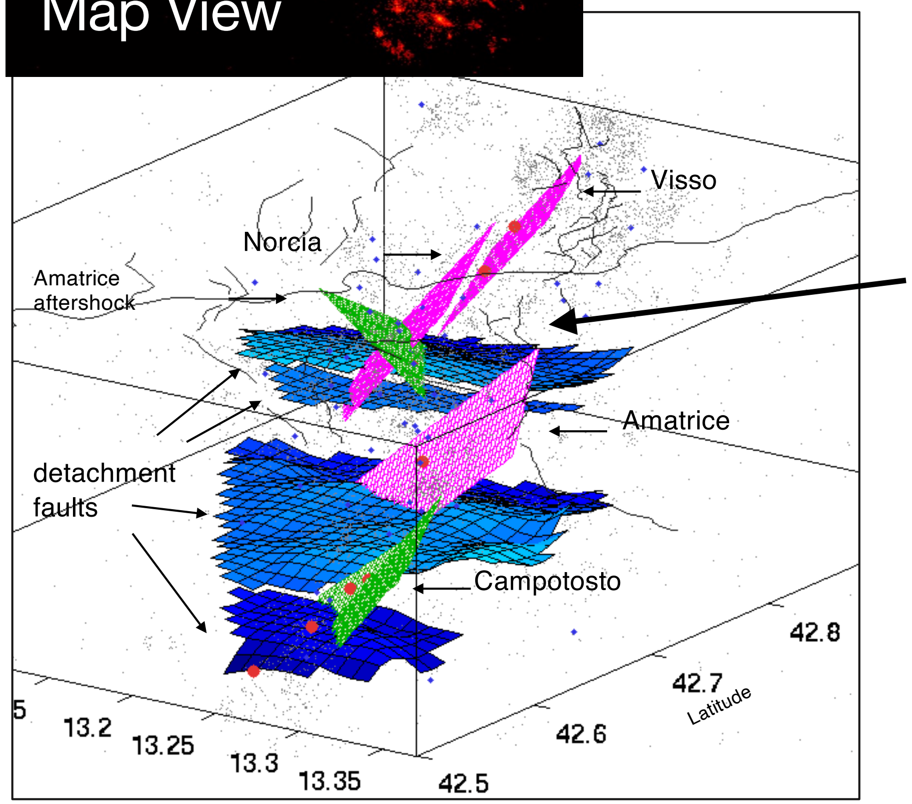
% of events with correlated waveforms



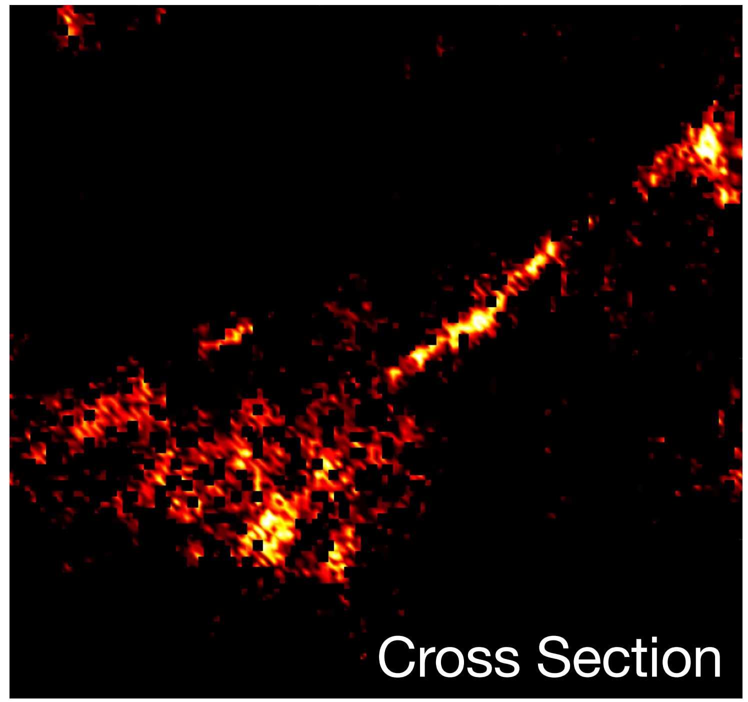
Waldhauser et al. (2021)

Fault geometry, structure, mechanics

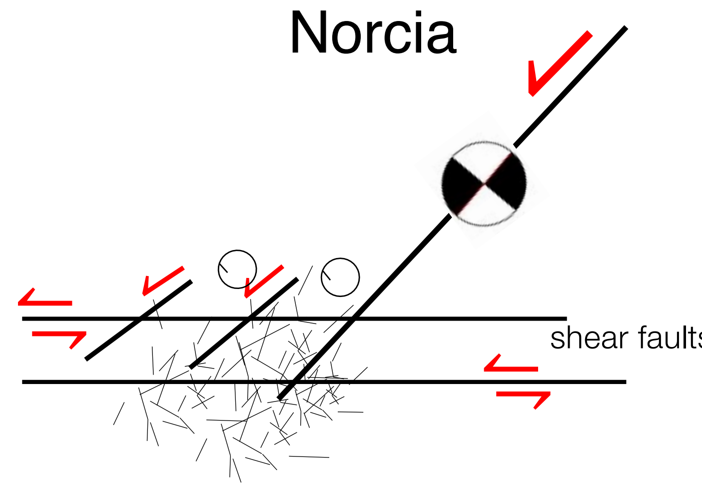




Mt. Vettore normal fault



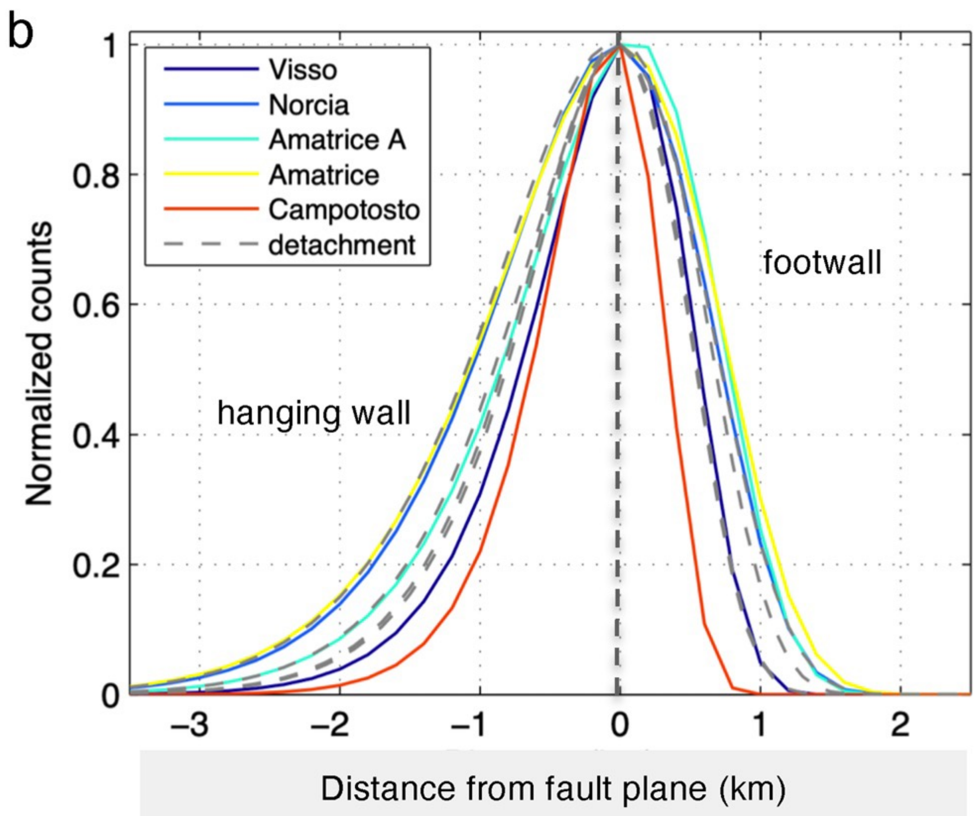
Bookshelf structure



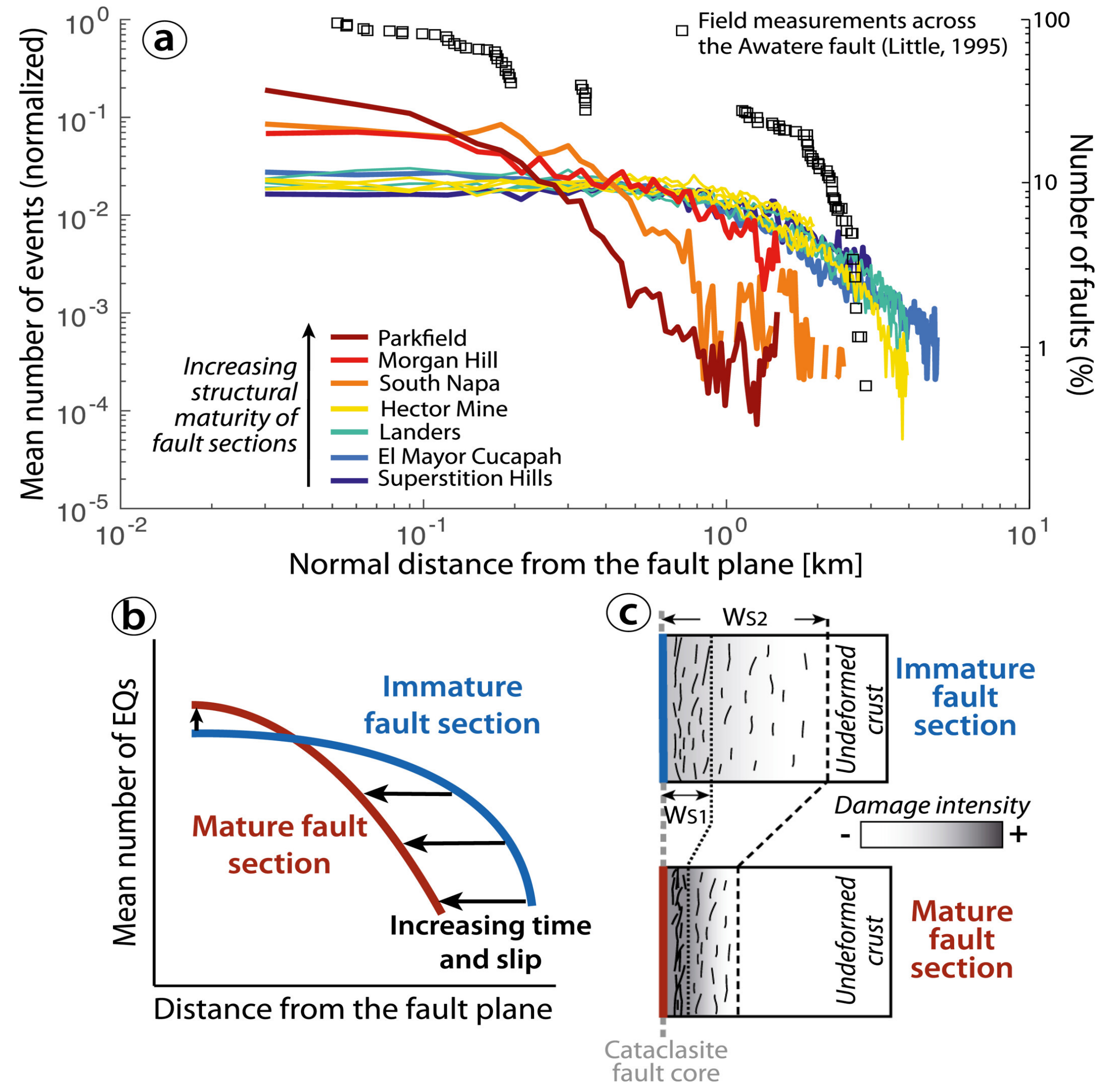
Earthquake density plot showing narrow Mt. Vettore normal fault (red) and bookshelf faults truncated by the detachment horizon (blue).

Waldhauser etal 2021

Fault zone width and inferred complexity

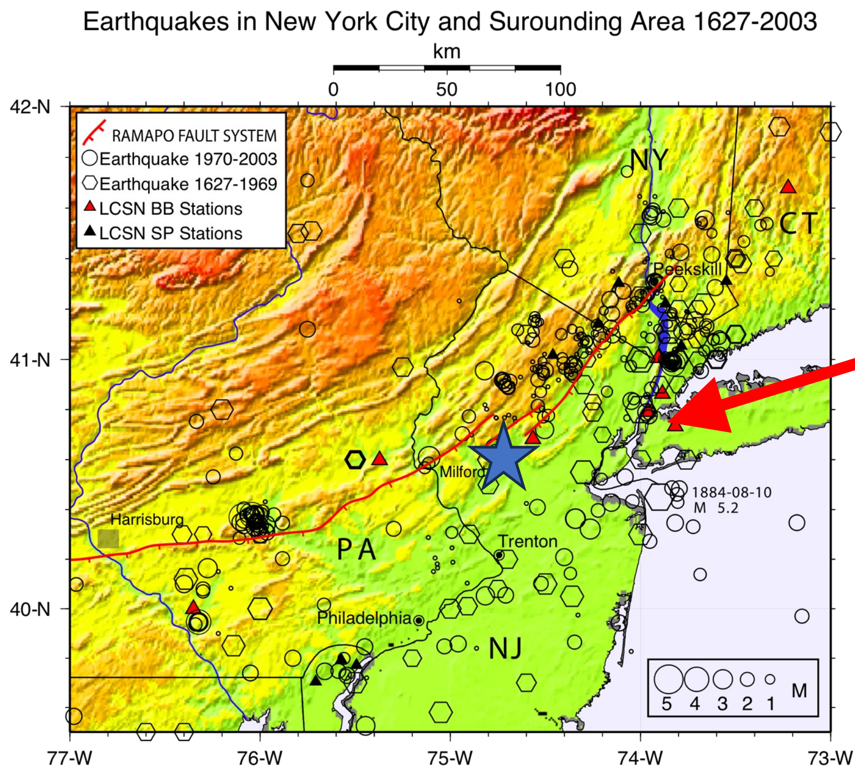


Waldhauser et al. (2021)



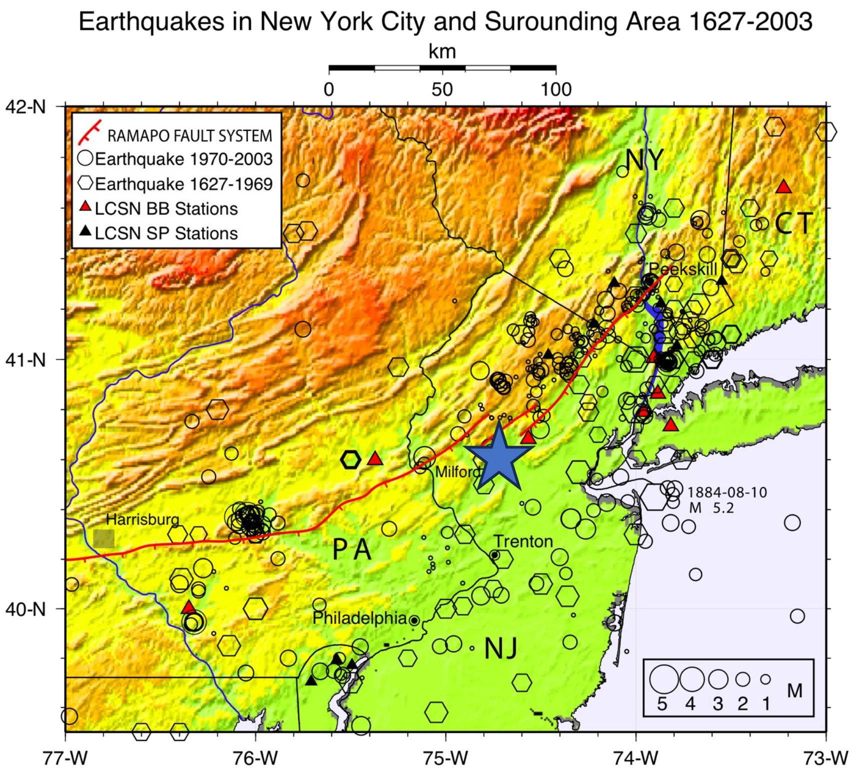
Perrin et al. (2021)

The 5 April 2024 Mw 4.8 New Jersey earthquake



- 45 miles from Lamont
- 80+ media responses
- Record DYFI reports
- 150+ aftershocks
- M3.7 aftershock on April 5, 6 PM

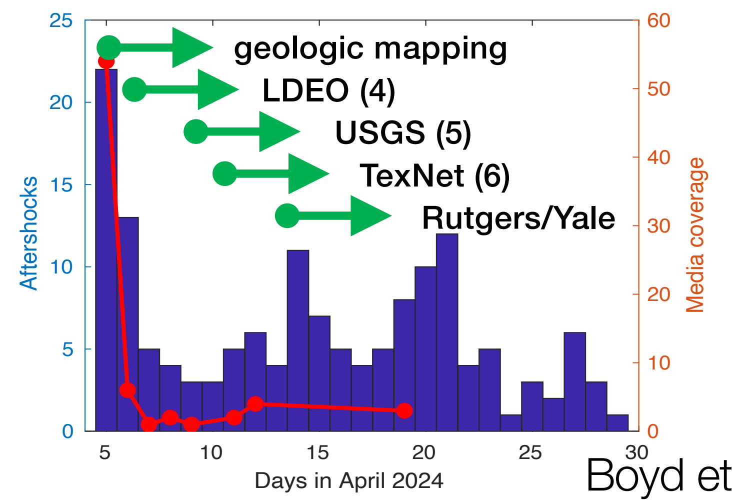
The 5 April 2024 Mw 4.8 New Jersey earthquake



(a)

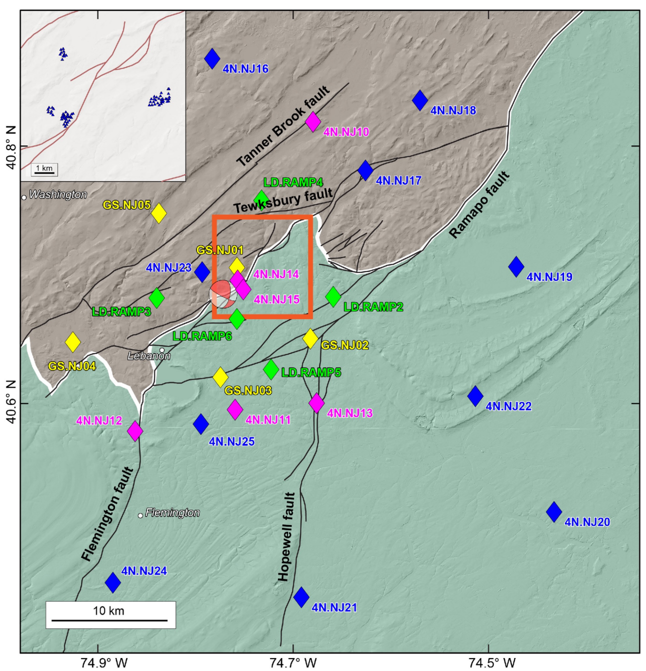
(b)

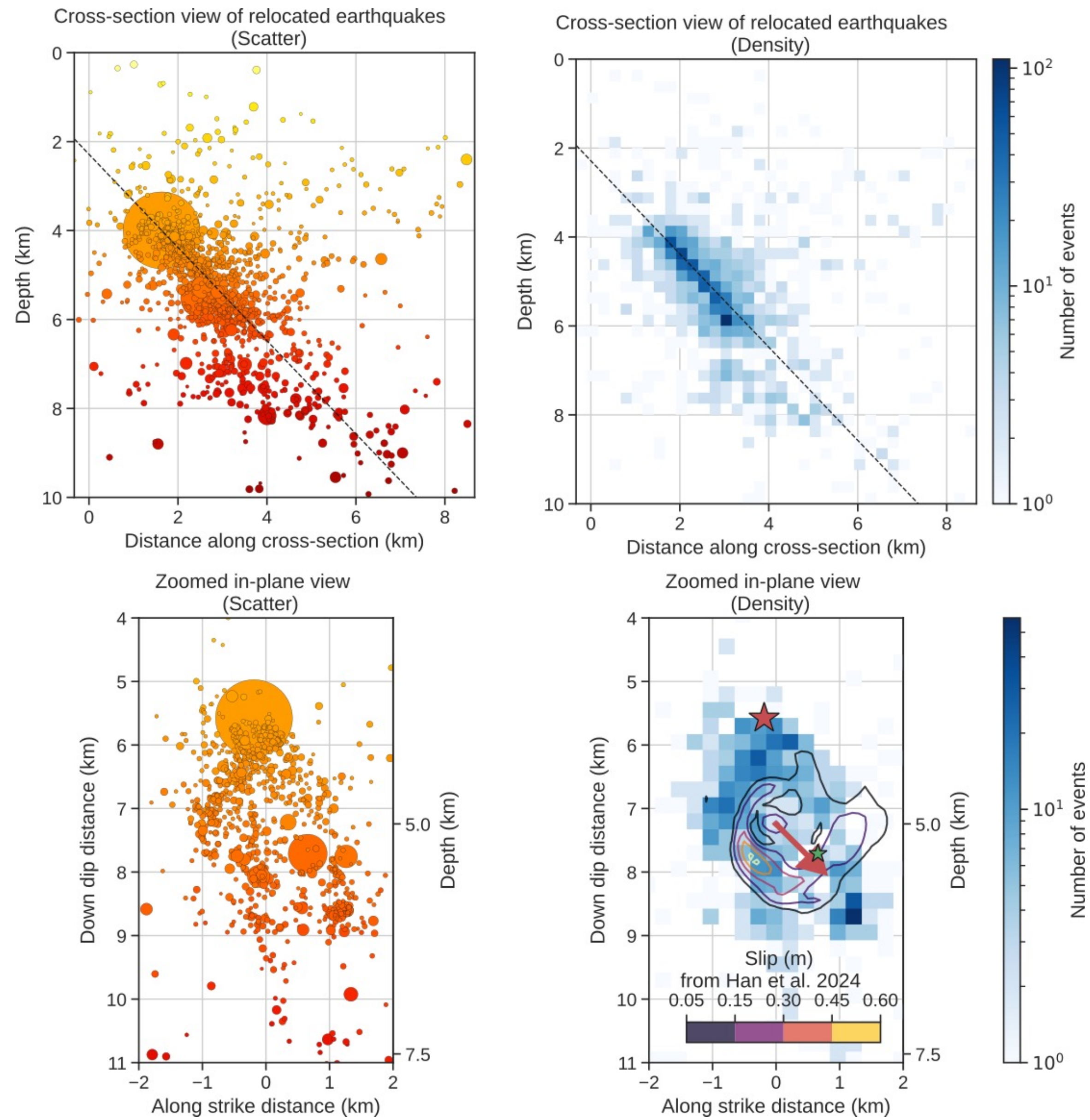
Boyd et al (2024)



Boyd et al (2024)

The 5 April 2024 Mw 4.8 New Jersey earthquake



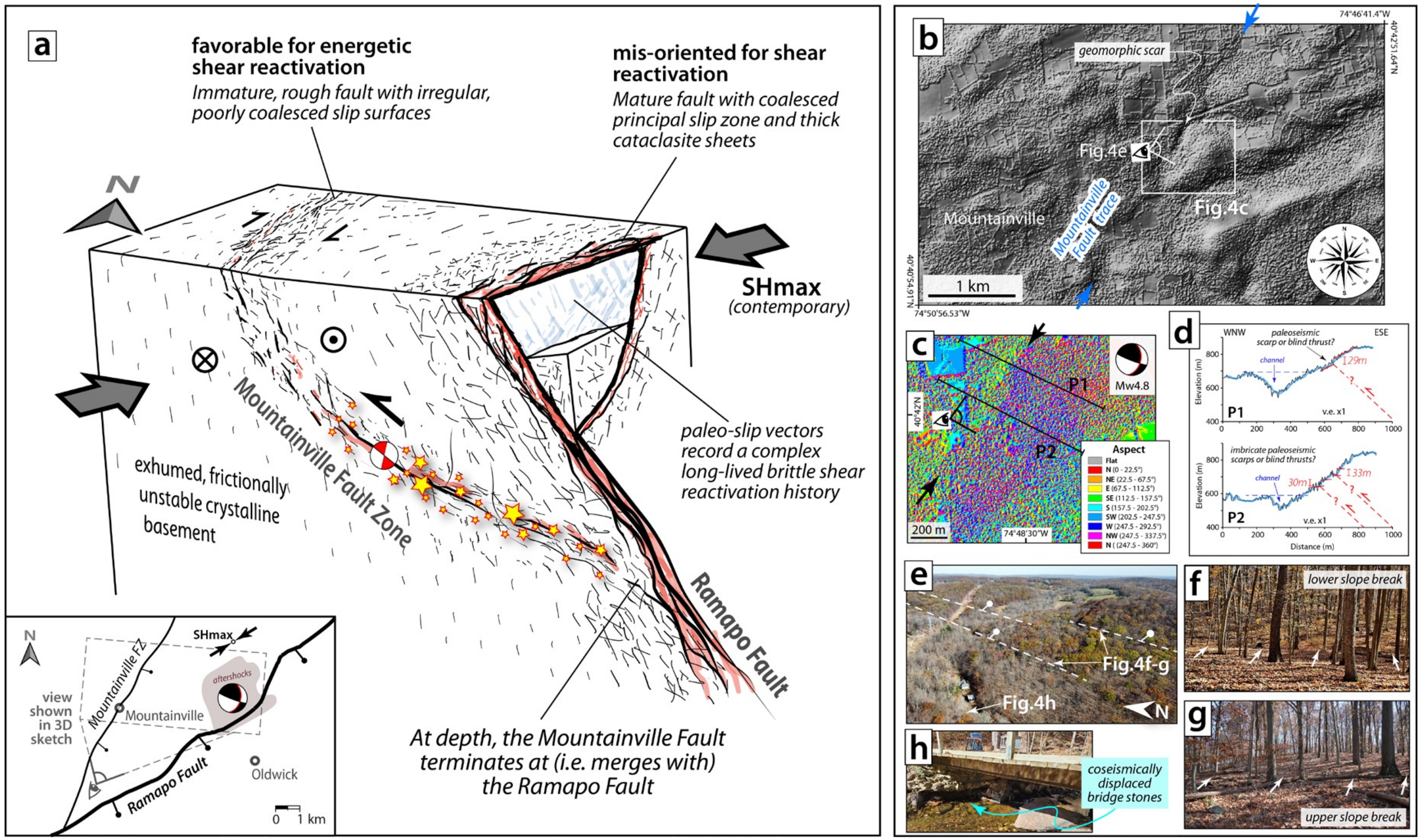


- back-projection
- machine-learning
- template-matching
- cross-correlation
- double-differencing

⇒ 2,000 aftershocks vs. ~200 by USGS!

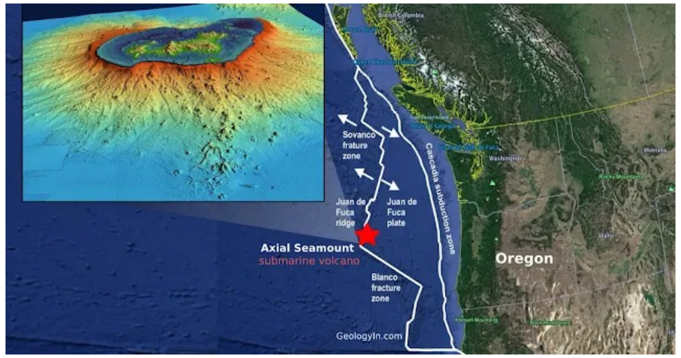
Beauce et al (2025)

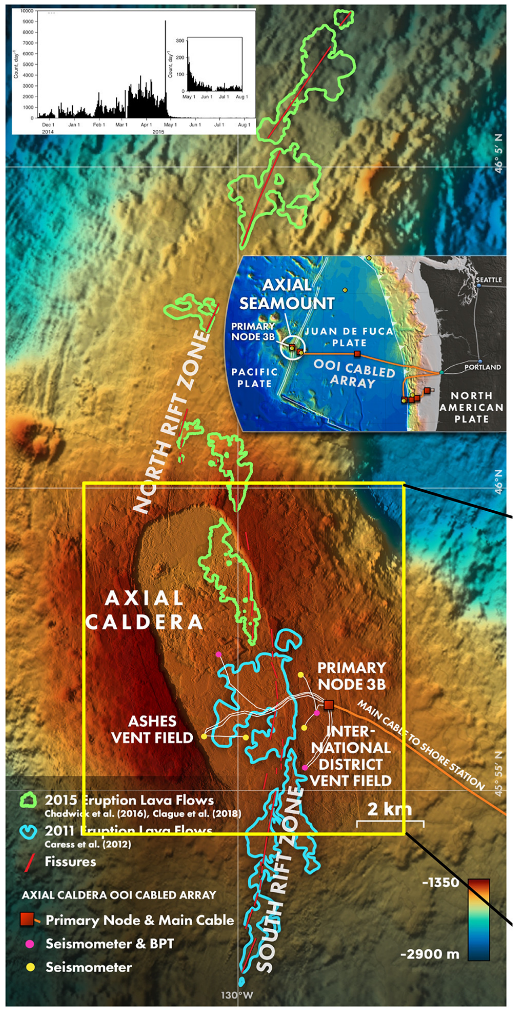
The 5 April 2024 Mw 4.8 New Jersey earthquake



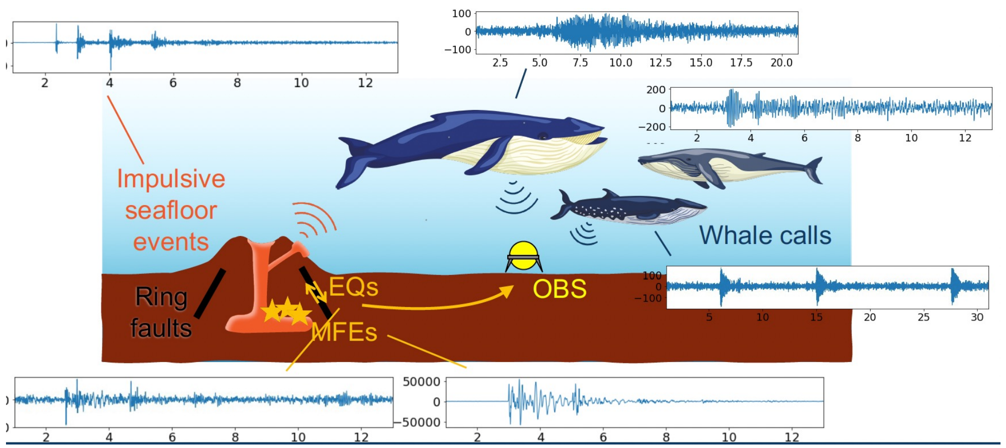
Kolawole et al (2025)

Axial Seamount (North Pacific)



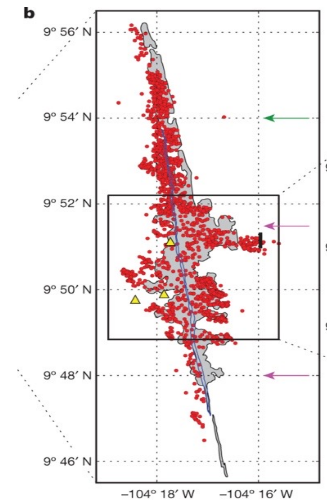
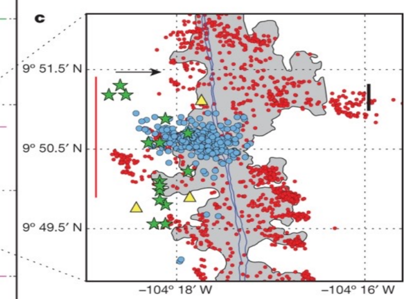


Wilcock et al (2016)

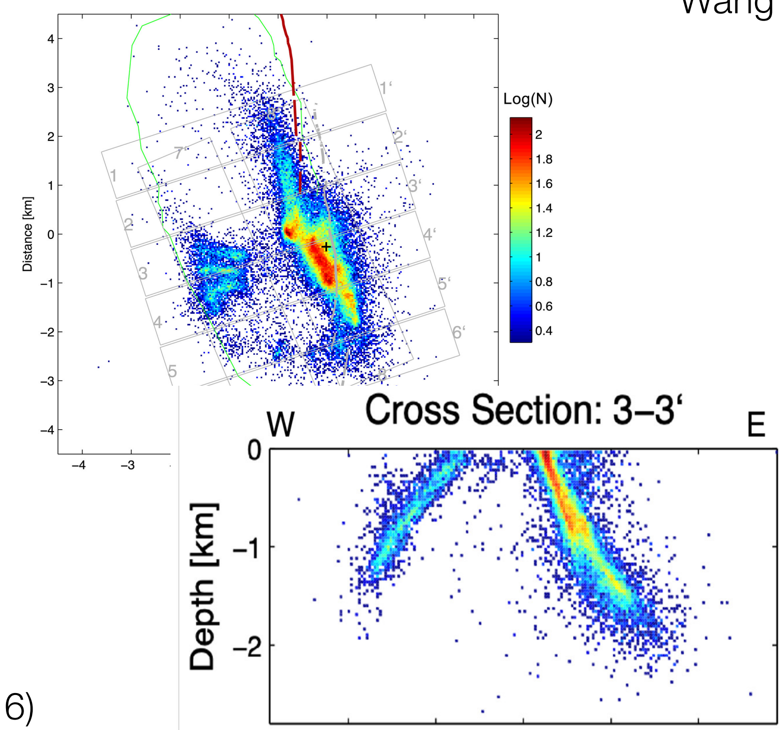


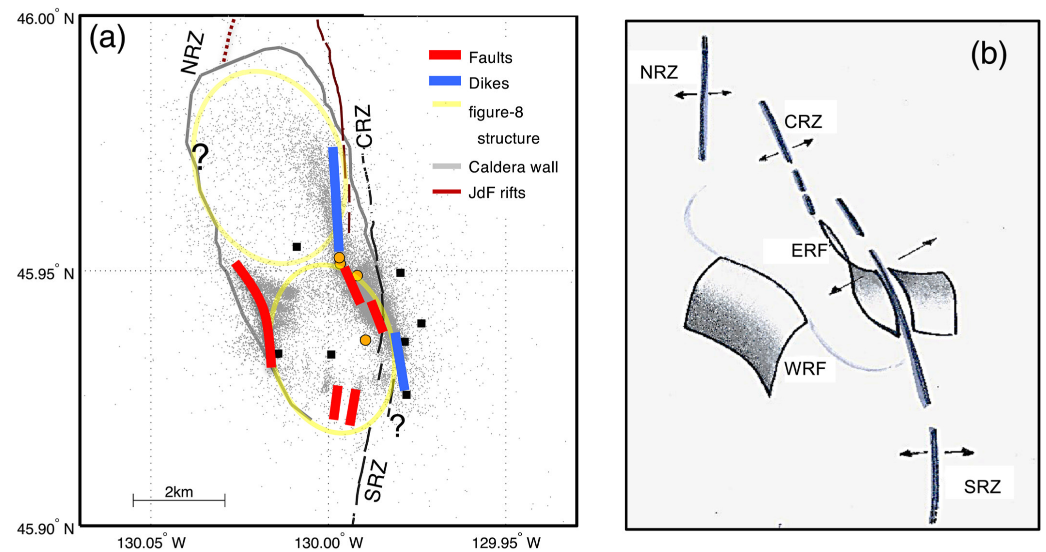
Wang et al (2024)

Impulsive seafloor Events (@EPR)

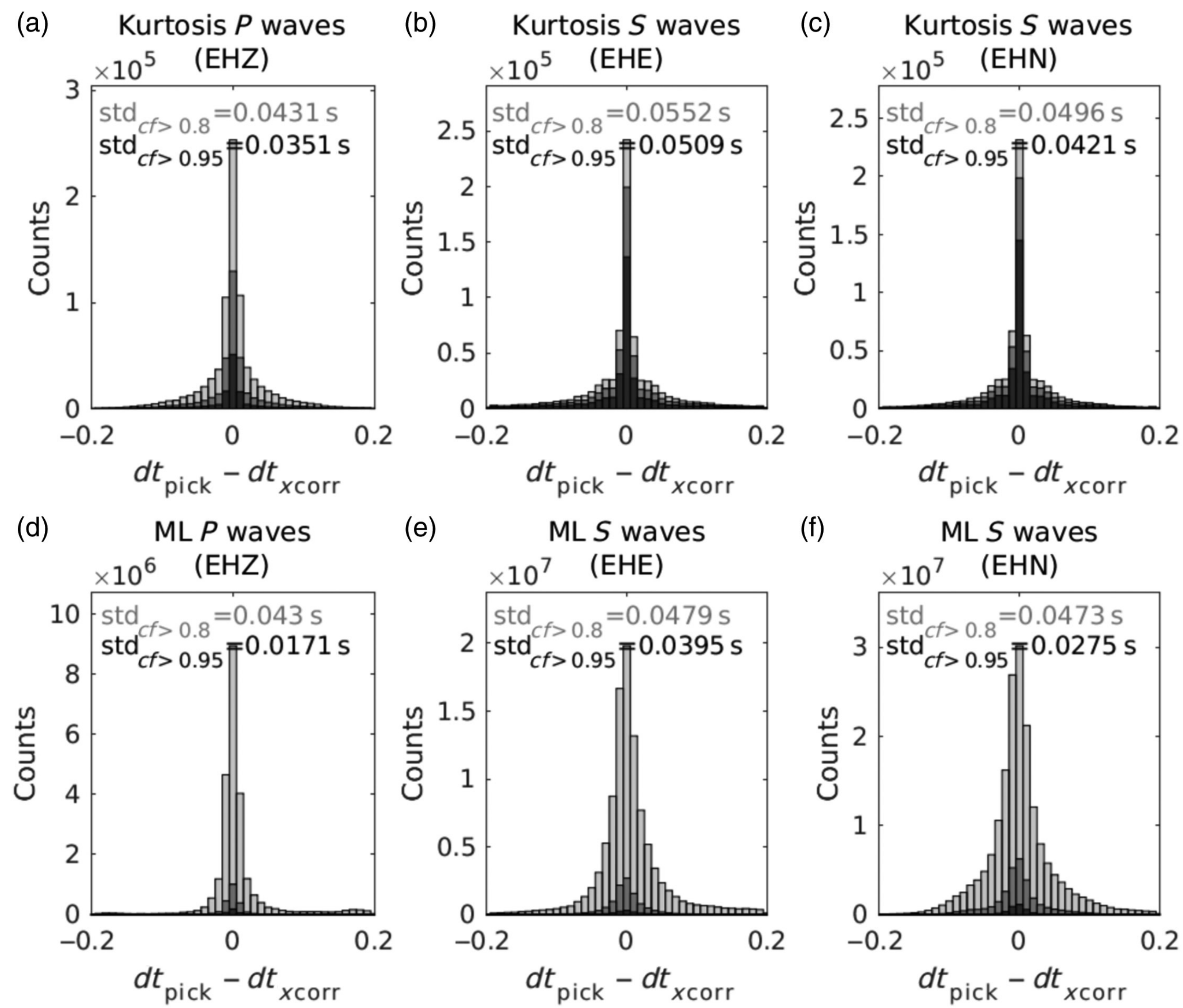
Tan et al (2016)



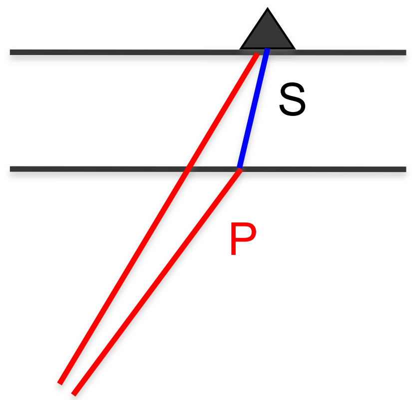


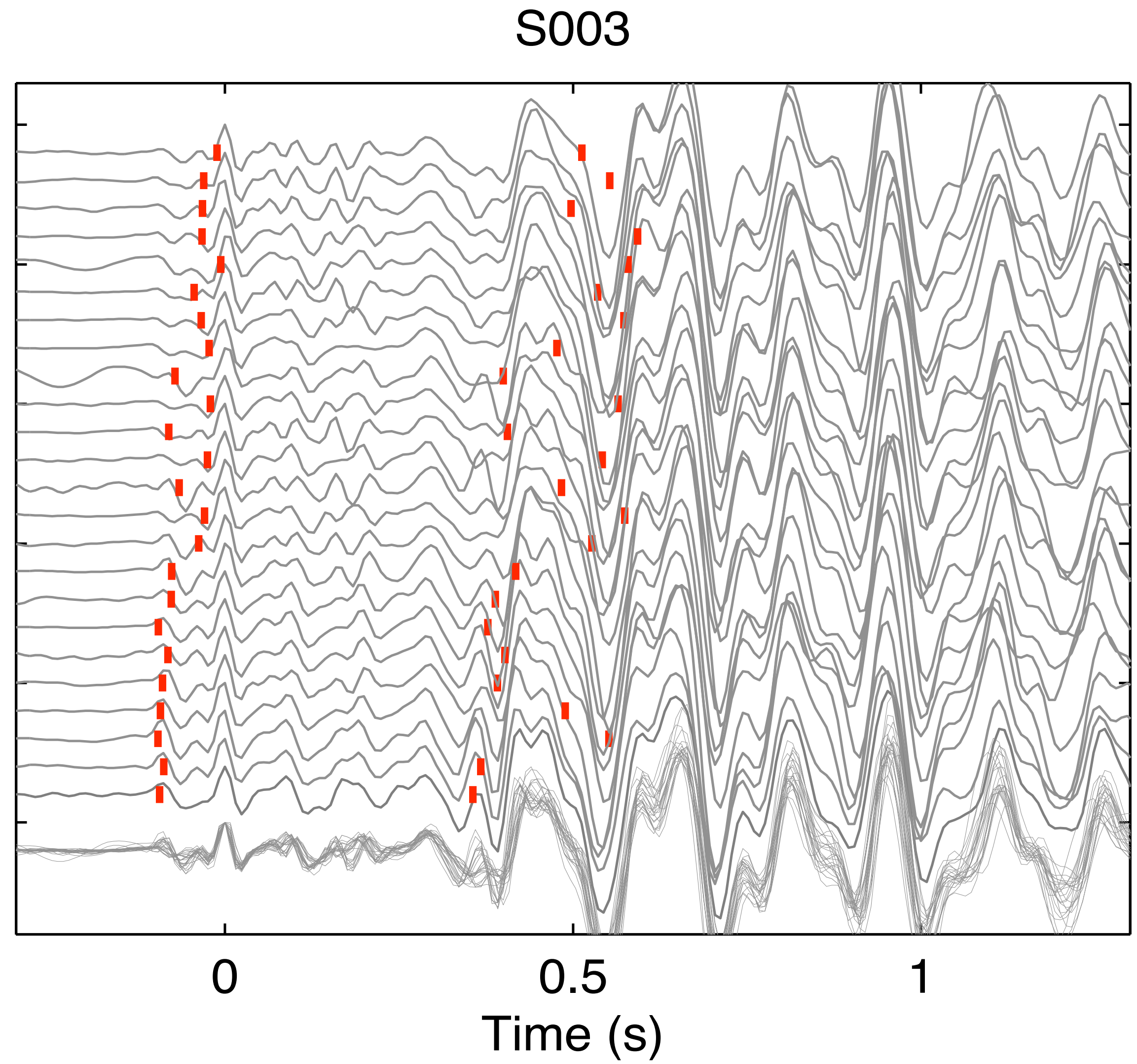
Waldhauser et al. (2020)

Picker performance: Kurtosis vs PhaseNet picks

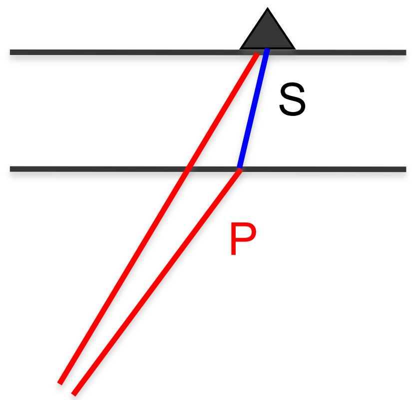


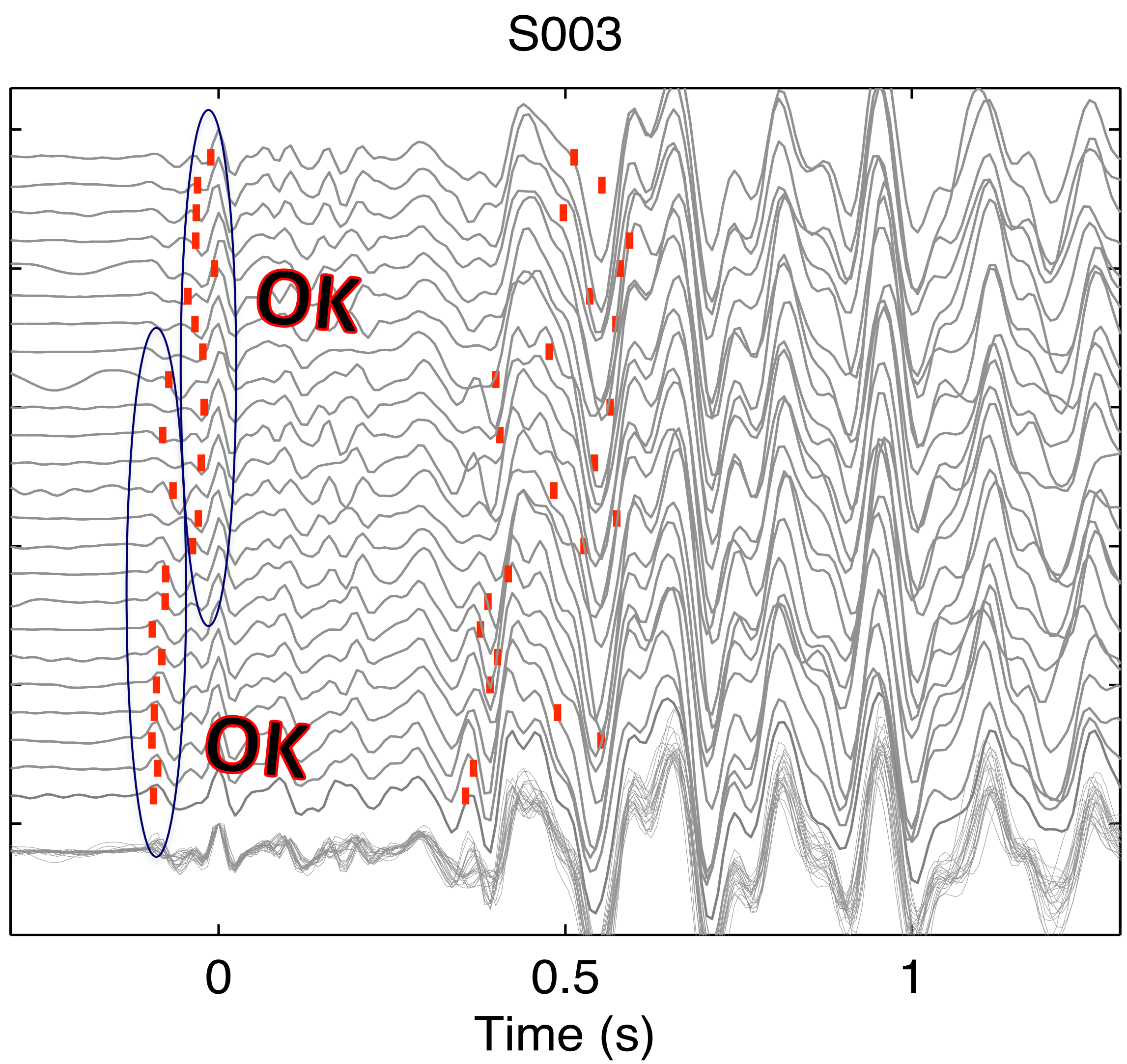
Near-surface phase conversions/reflections





Near-surface phase conversions/reflections

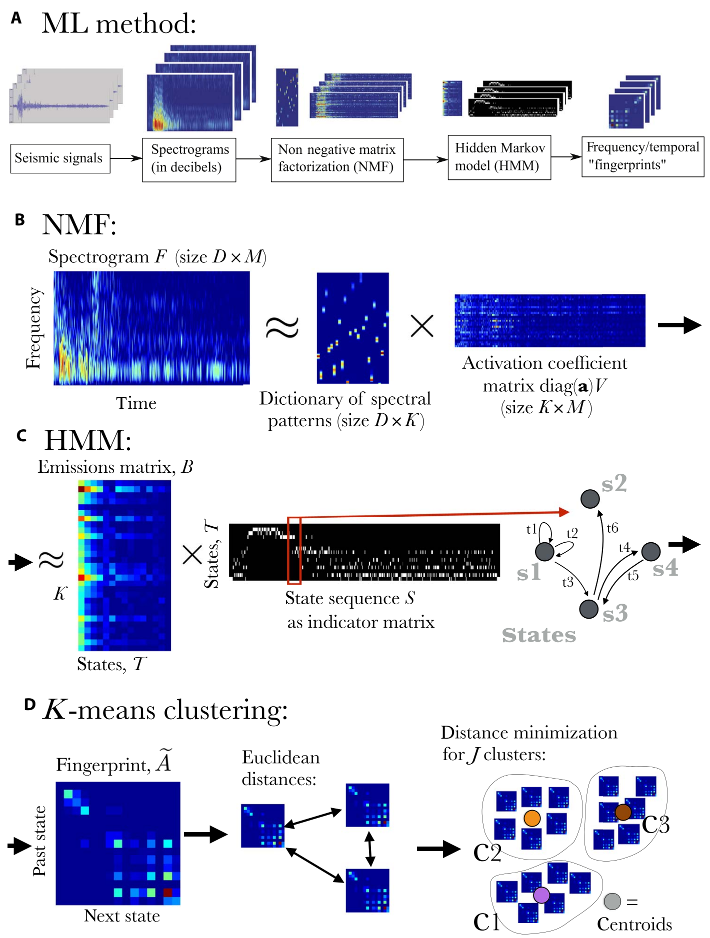




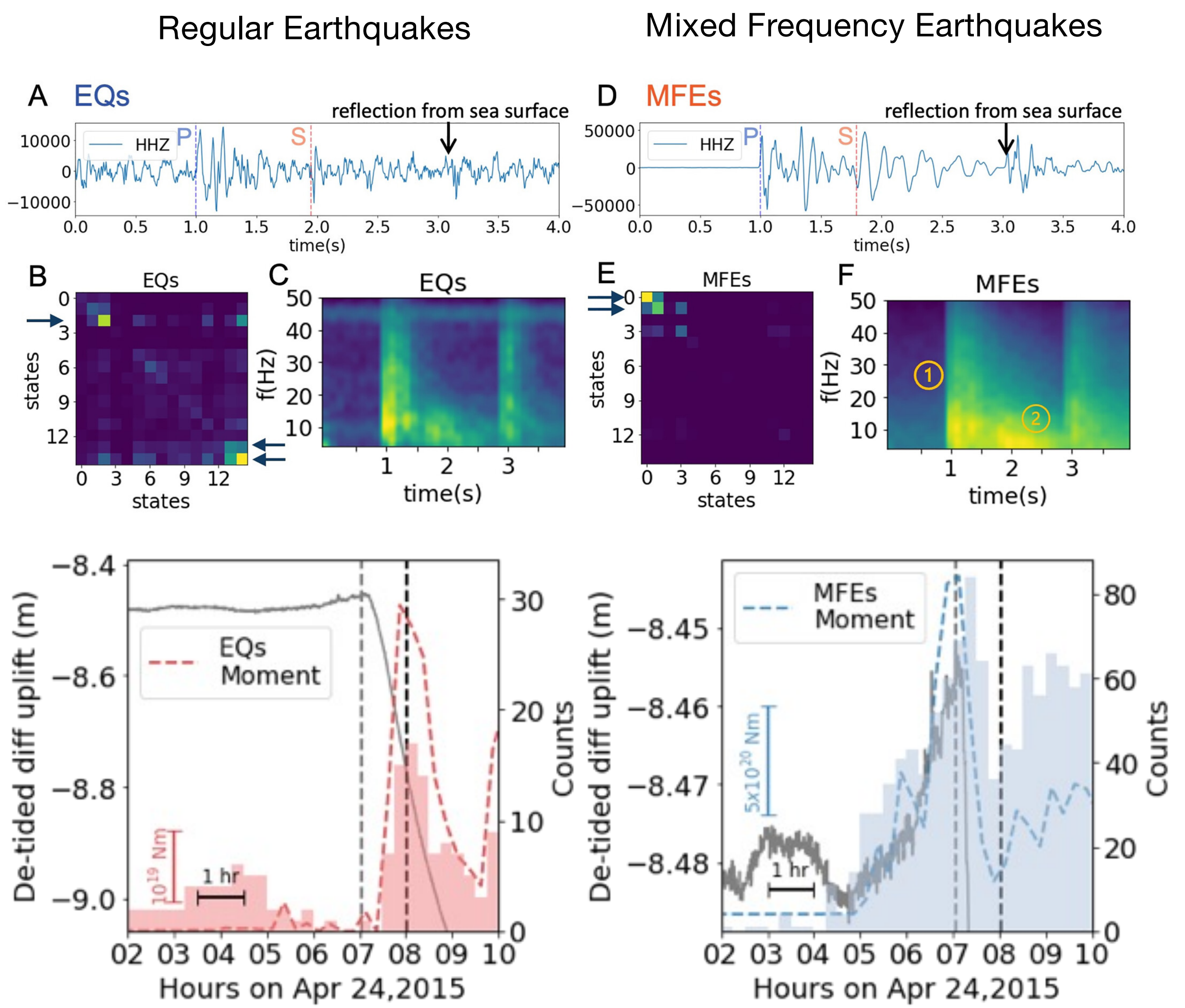
- Consistent late or early picking is OK (except maybe for P-S conversions), but not across.

Unsupervised machine learning detects volcanic precursors

Unsupervised spectral feature extraction and clustering (specUFEx)

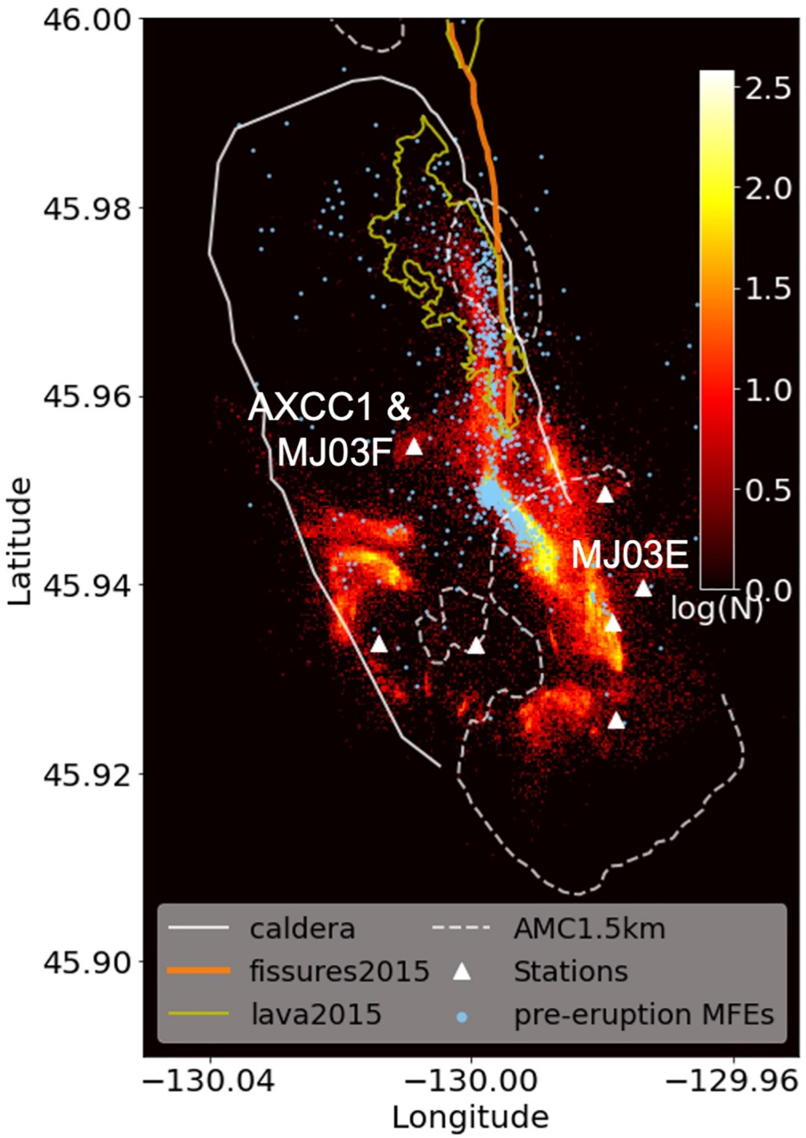


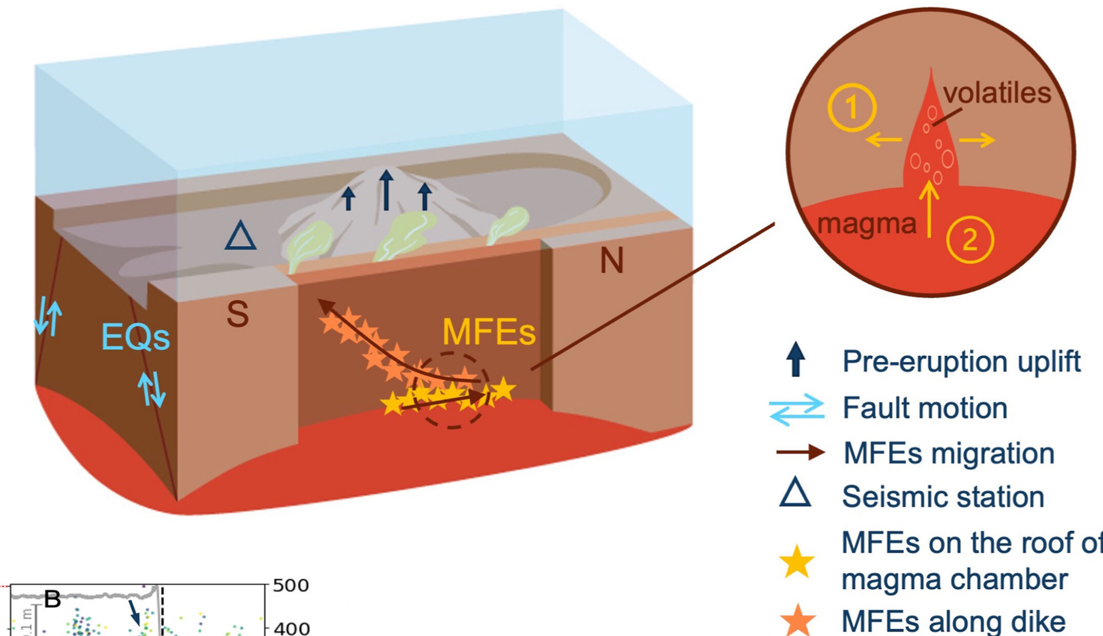
Holtzman, Paisley et al. (2018)

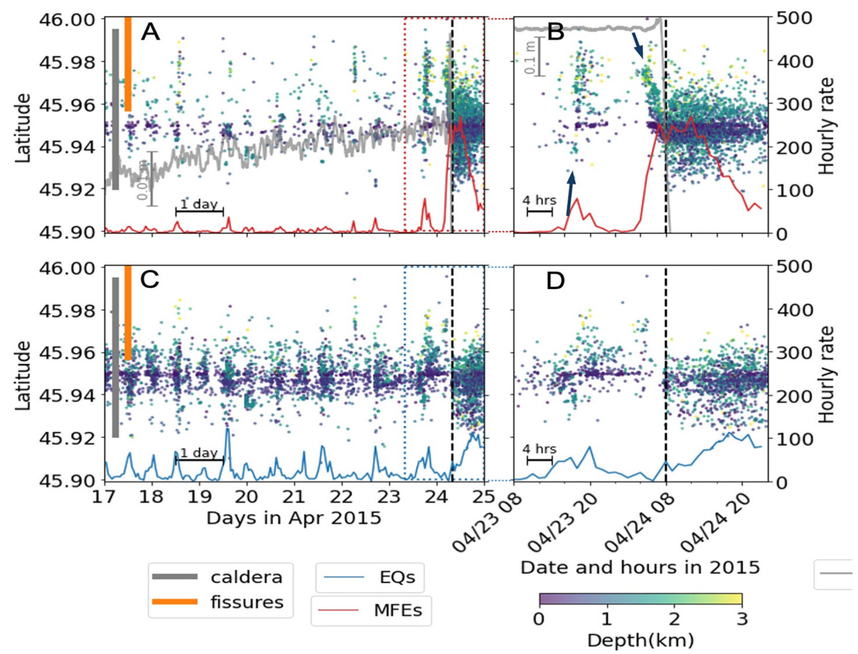


Kaiwen Wang et al (2024)

Interpretation: MFEs track movement of volatiles or magma







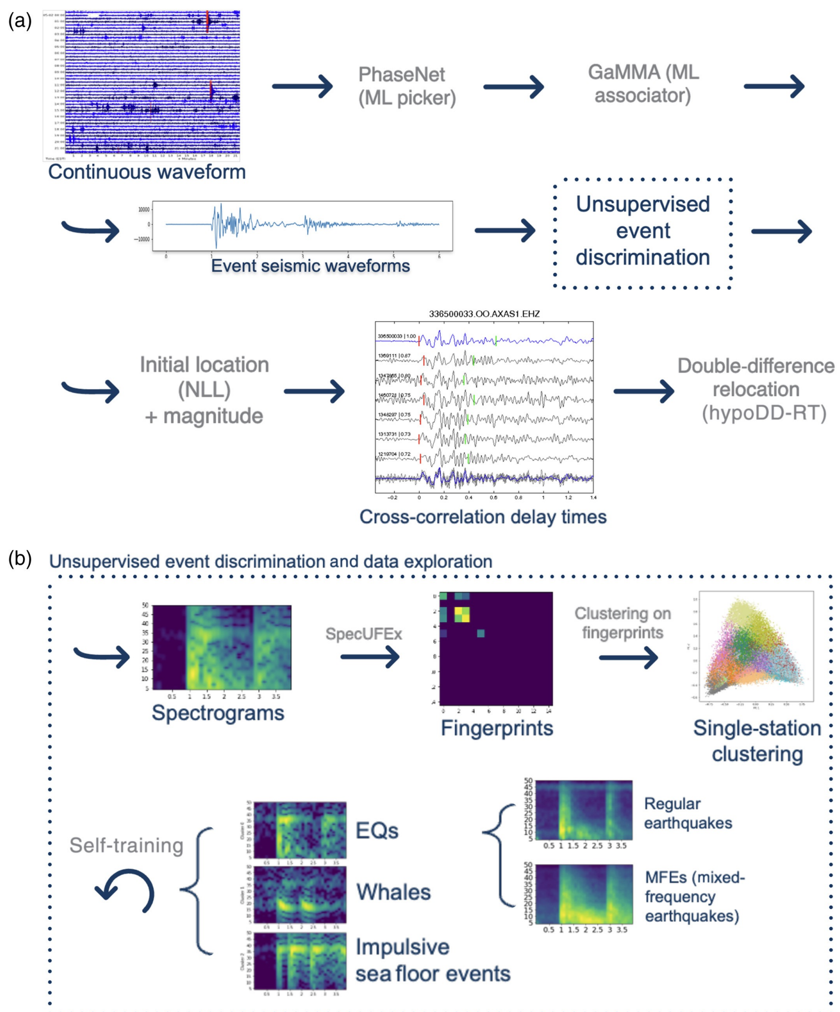
MFE

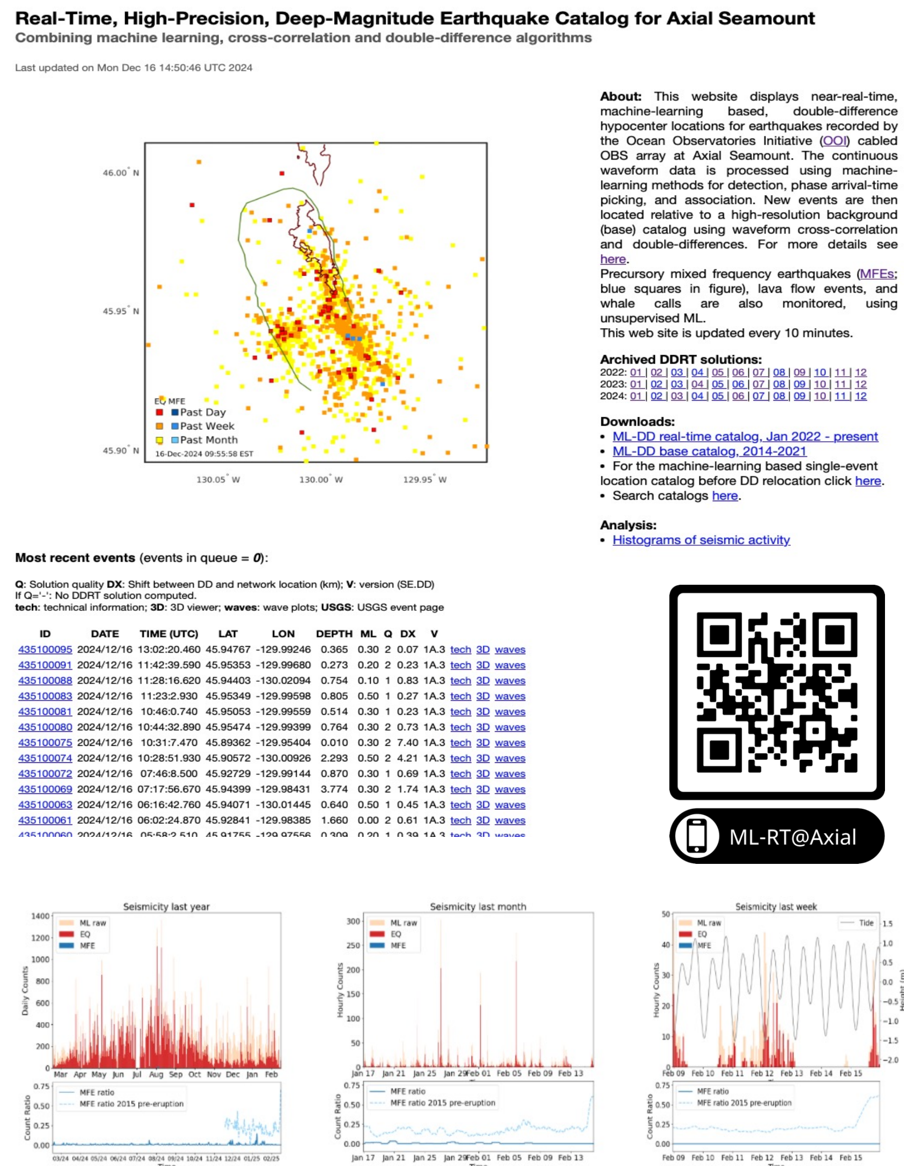
EQ

Kaiwen Wang et al (2024)

Real-time implementation: *https://axialDD.ldeo.columbia.edu*

ML-DD real-time workflow





Real-Time, High-Precision, Deep-Magnitude Earthquake Catalog for Axial Seamount

Combining machine learning, cross-correlation and double-difference algorithms

Last updated on Mon Dec 16 14:50:46 UTC 2024

About: This website displays near-real-time, machine-learning based, double-difference hypocenter locations for earthquakes recorded by the Ocean Observatories Initiative (OOI) cabled OBS array at Axial Seamount. The continuous waveform data is processed using machine-learning methods for detection, phase arrival-time picking, and association. New events are then located relative to a high-resolution background (base) catalog using waveform cross-correlation and double-differences. For more details see here.
Precursory mixed frequency earthquakes (MFEs; blue squares in figure), lava flow events, and whale calls are also monitored, using unsupervised ML.
This web site is updated every 10 minutes.

Archived DDRT solutions:
2022: 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12
2023: 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12
2024: 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12

Downloads:
- ML-DD real-time catalog, Jan 2022 - present
- ML-DD base catalog, 2014-2021
- For the machine-learning based single-event location catalog before DD relocation click here.
- Search catalogs here.

Analysis:
- Histograms of seismic activity

Most recent events (events in queue = *0*):

Q: Solution quality **DX**: Shift between DD and network location (km); **V**: version (SE.DD)
If Q='-': No DDRT solution computed.
tech: technical information; **3D**: 3D viewer; **waves**: wave plots; **USGS**: USGS event page

ID	DATE	TIME (UTC)	LAT	LON	DEPTH	ML	Q	DX	V			
435100095	2024/12/16	13:02:20.460	45.94767	-129.99246	0.365	0.30	2	0.07	1A.3	tech	3D	waves
435100091	2024/12/16	11:42:39.590	45.95353	-129.99680	0.273	0.20	2	0.23	1A.3	tech	3D	waves
435100088	2024/12/16	11:28:16.620	45.94403	-130.02094	0.754	0.10	1	0.83	1A.3	tech	3D	waves
435100083	2024/12/16	11:23:2.930	45.95349	-129.99598	0.805	0.50	1	0.27	1A.3	tech	3D	waves
435100081	2024/12/16	10:46:0.740	45.95053	-129.99559	0.514	0.30	1	0.23	1A.3	tech	3D	waves
435100080	2024/12/16	10:44:32.890	45.95474	-129.99399	0.764	0.30	2	0.73	1A.3	tech	3D	waves
435100075	2024/12/16	10:31:7.470	45.89362	-129.95404	0.010	0.30	2	7.40	1A.3	tech	3D	waves
435100074	2024/12/16	10:28:51.930	45.90572	-130.00926	2.293	0.50	2	4.21	1A.3	tech	3D	waves
435100072	2024/12/16	07:46:8.500	45.92729	-129.99144	0.870	0.30	1	0.69	1A.3	tech	3D	waves
435100069	2024/12/16	07:17:56.670	45.94399	-129.98431	3.774	0.30	2	1.74	1A.3	tech	3D	waves
435100063	2024/12/16	06:16:42.760	45.94071	-130.01445	0.640	0.50	1	0.45	1A.3	tech	3D	waves
435100061	2024/12/16	06:02:24.870	45.92841	-129.98385	1.660	0.00	2	0.61	1A.3	tech	3D	waves
435100060	2024/12/16	05:58:2.510	45.91755	-129.97556	0.309	0.20	1	0.39	1A.3	tech	3D	waves

Wang et al (SRL, 2024)
Waldhauser et al (2020)

Opportunities and Challenges

Opportunities

- Large archives of continuous waveforms
- Decades of expert labeled data
- Unlimited compute power
- Widely used algorithms that take advantage of all the above
- New generation of deep-magnitude earthquake catalogs that enable research and discovery
- New era of high-precision seismic monitoring

Challenges

- Ensure reliability of catalog and products
- Raise awareness of limitations
- New generation of machine-seismologists
- Machine learning of seismogenic processes

Opportunities and Challenges

Opportunities

- Large archives of continuous waveforms
- Decades of expert labeled data
- Unlimited compute power
- Widely used algorithms that take advantage of all the above
- New generation of deep-magnitude earthquake catalogs that enable research and discovery
- New era of high-precision seismic monitoring

Challenges

- Ensure reliability of catalog and products
- Raise awareness of limitations
- New generation of machine-seismologists
- Machine learning of seismogenic processes

Most annoying new ML problem

- Event IDs

Thank you!