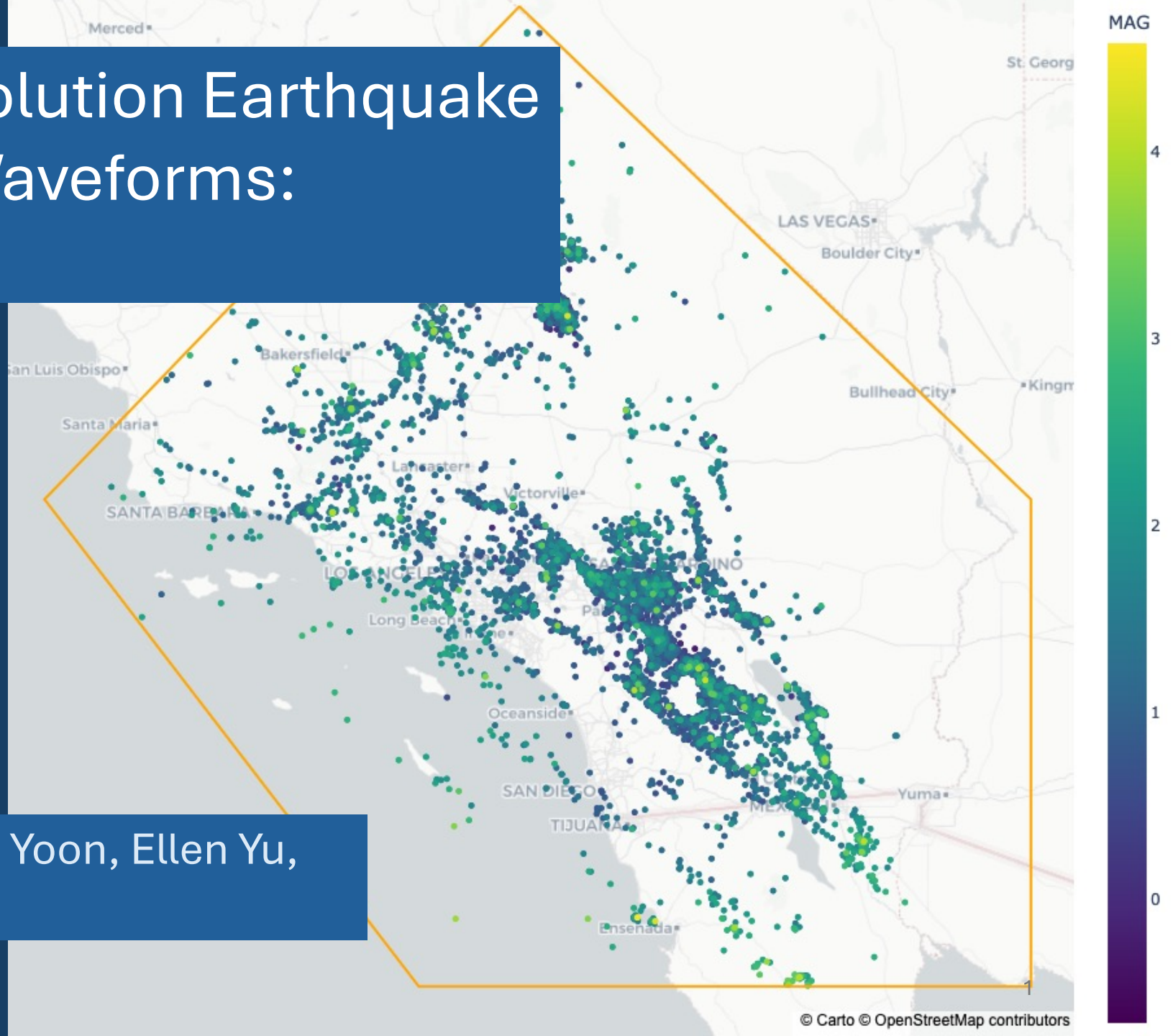
Building a High-Resolution Earthquake Catalog from Raw Waveforms:
A Step-by Step Guide

SSA Workshop
Baltimore, MD

April 14, 2025 (1 – 5 pm)

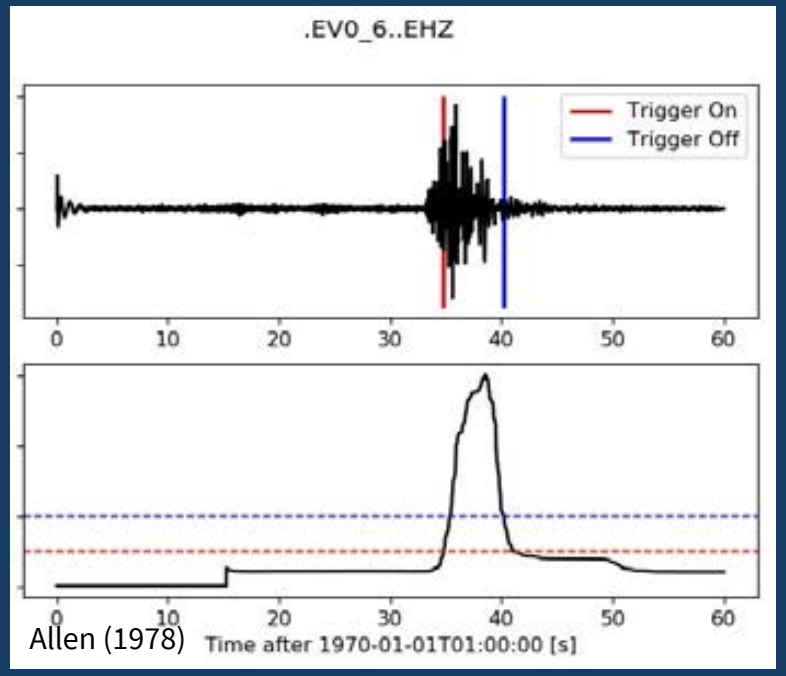
Presenters: Eric Beaucé, Clara Yoon, Ellen Yu, Weiqiang Zhu, Gabrielle Tepp



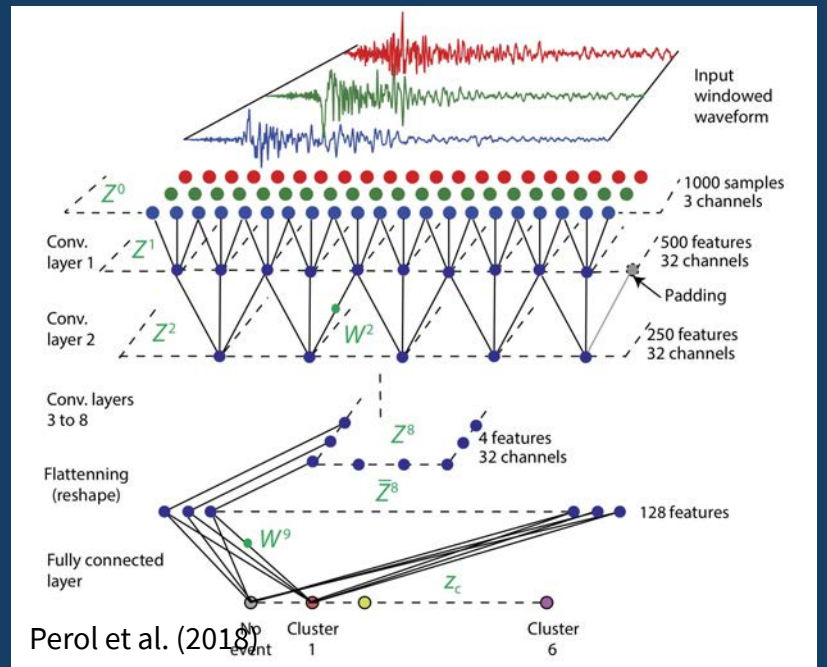
Conclusions

Which catalog should I use?

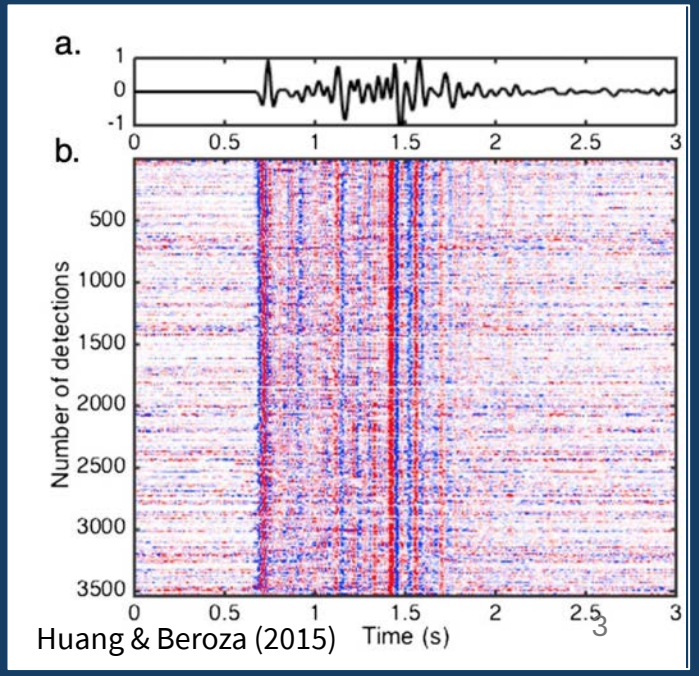
- Routine Catalog (STA/LTA, manual)
- Enhanced Catalog
  - Machine Learning
  - Template Matching



Allen (1978)



Perol et al. (2018)

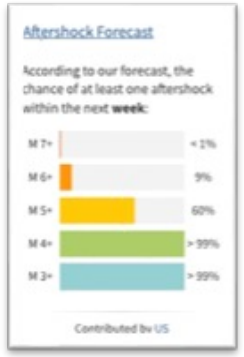


Huang & Beroza (2015)

When should I use a routine network catalog (ComCat)?

Maybe I don't need an enhanced catalog ?

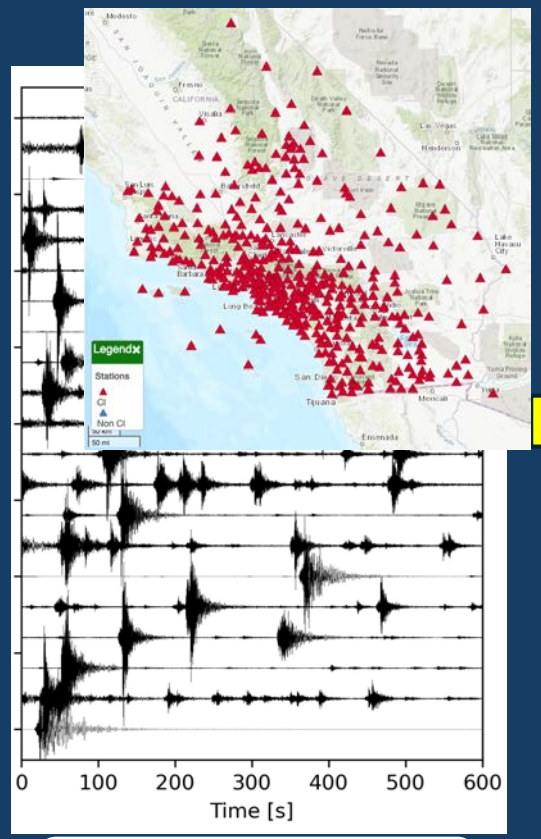
- When catalog quality & accuracy are more important than completeness
  - When you need a reference catalog!
    - Manually reviewed events & picks, no false detections
  - **When you care about big earthquakes, but can safely ignore small earthquakes**
    - e.g. seismic hazard analysis, aftershock forecasting
    - e.g. near-real-time operational earthquake monitoring



- General exploration of past seismicity & tectonics in a region
  - Depths may not be precise or well-constrained

Which method to use at each step in automatic earthquake catalog workflow?

INPUT



Seismic Data

Phase Detection & Picking

Phase Association

Earthquake Location

Earthquake Magnitude

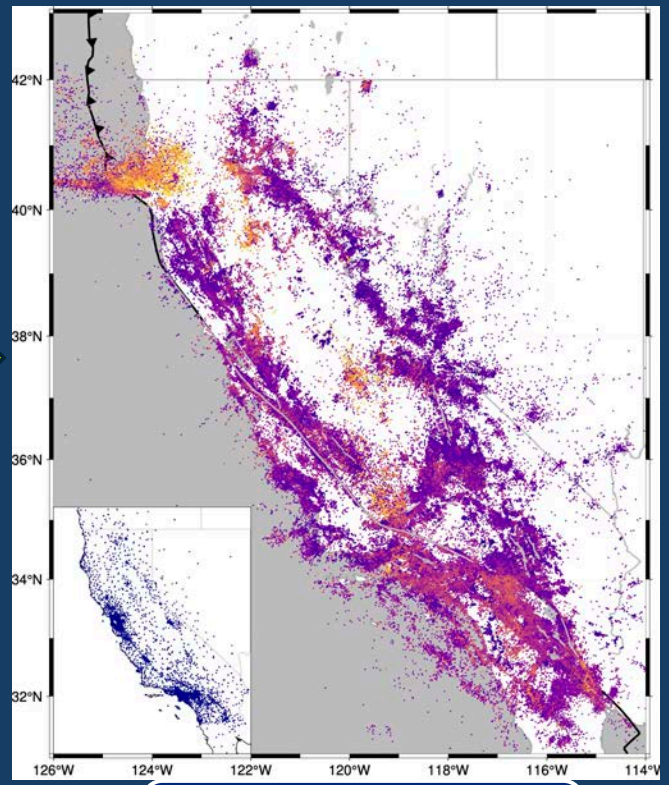
Focal Mechanism / Moment Tensor

First motion polarity

S/P amplitude ratio

Full-waveform detection & location

OUTPUT



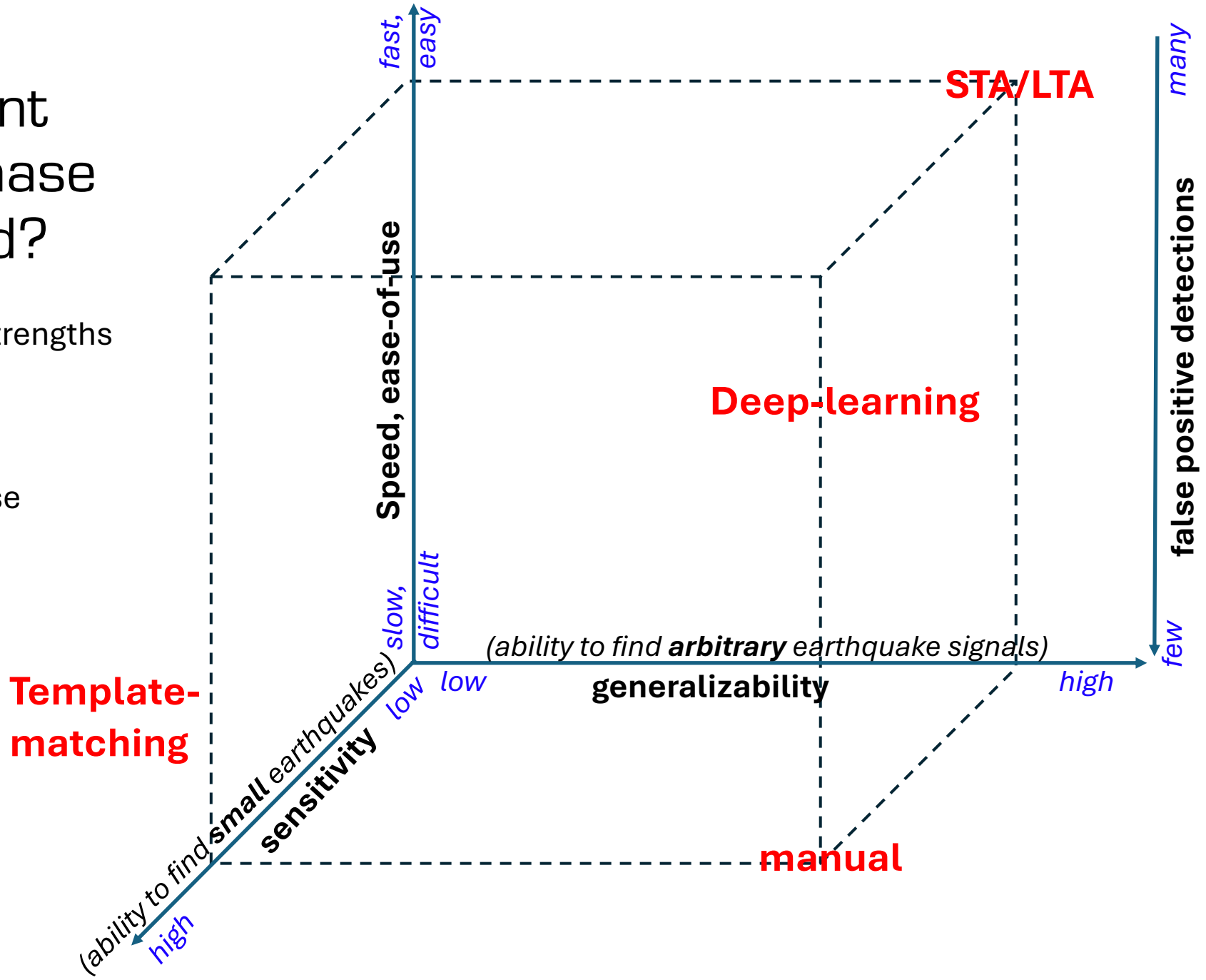
Earthquake Catalog

How should I choose an event detection & phase picking method?

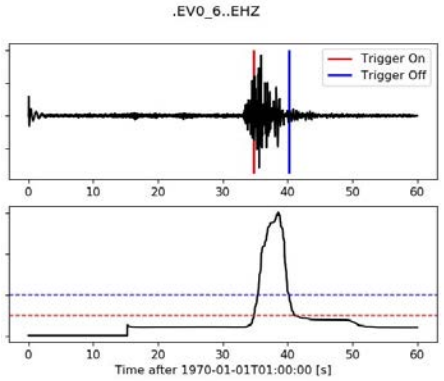
Each method has their strengths & weaknesses

Note **tradeoff**: speed/ease-of-use & false detections

Purpose and resources guide your choice.

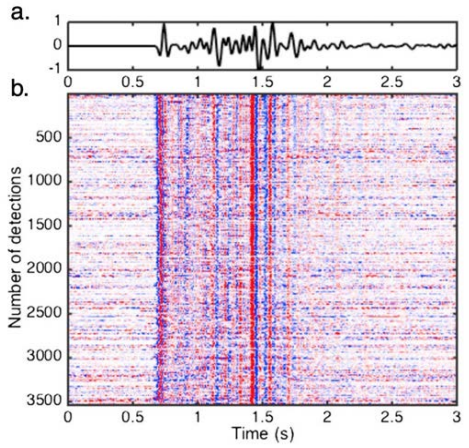


When should I use **STA/LTA**?



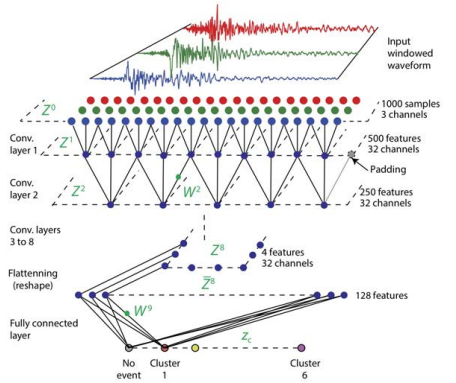
- 😀 Advantages
  - Runs very fast: automatically in real-time
  - Easy to understand & implement; optimize for different window lengths & ratios
  - No prior knowledge needed about earthquake sources or waveforms
  - Amplitude-based detector, so it reliably detects large earthquake signals
- ☹️ Limitations
  - High rate of false detections during active sequences
  - Automatic picks not as precise
  - Need manual review and refinement of picks for a quality catalog

When should I use **template-matching**?



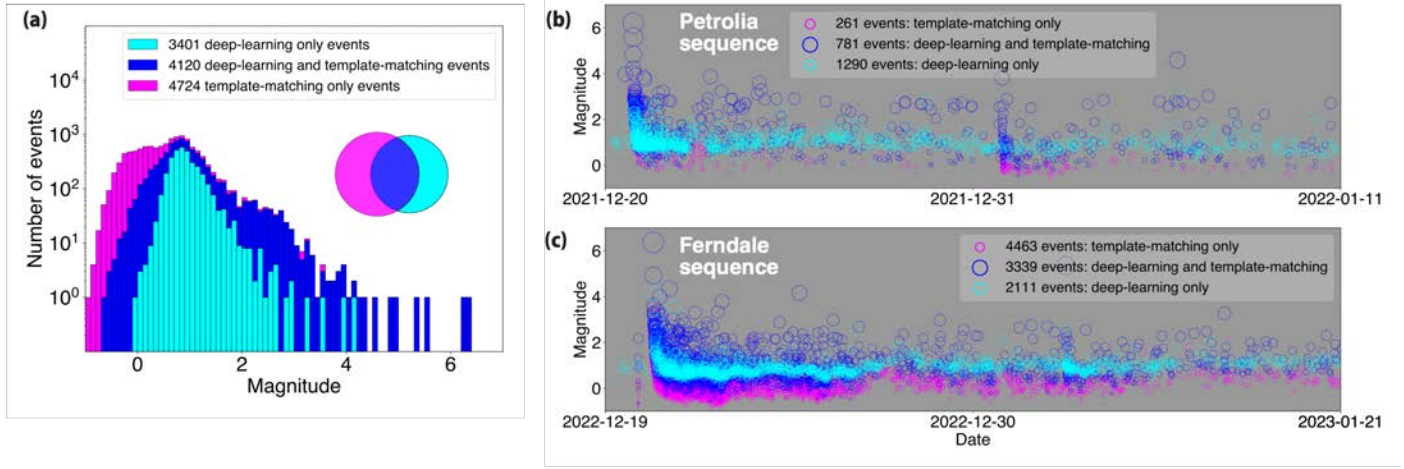
- 😀 Advantages
  - Optimally sensitive detector (more so than deep-learning): find smallest earthquakes buried in noise, if similar enough to template waveform
  - Excellent for improving temporal resolution of earthquake sequence
  - False detections are not as concerning, if high detection threshold
- ☹️ Limitations
  - Need prior knowledge about earthquake sources: **template waveforms** with good picks from preexisting catalog
  - Does not improve spatial resolution: unknown earthquake sources, not similar enough to template, cannot be found
  - Requires some effort to extract template waveforms and set up processing
  - Computationally intensive

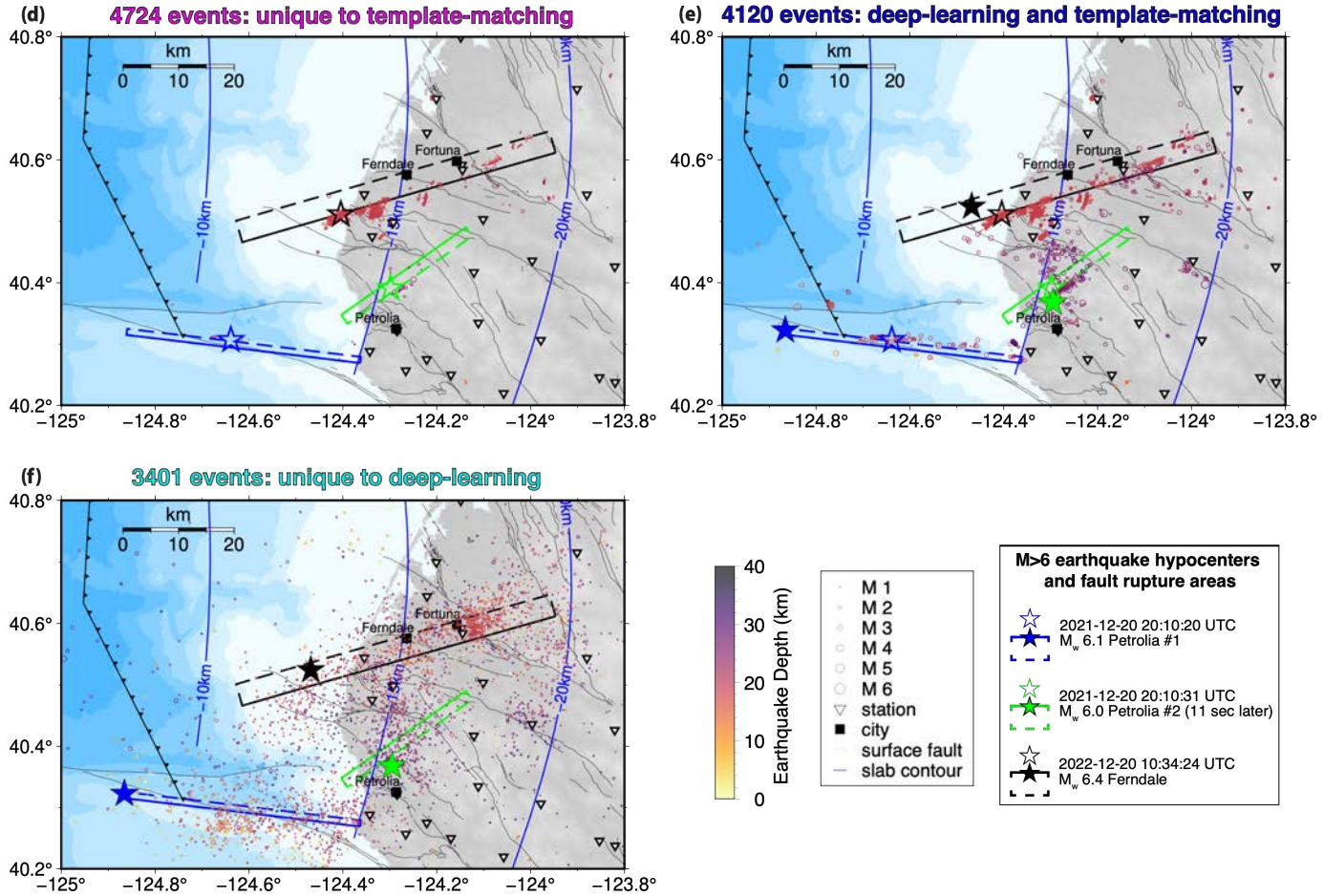
When should I use **deep-learning pickers**?



- Useful when:
  - Adds most value when existing seismic networks are sparse/nonexistent
  - Automatically & rapidly create more complete catalog during active sequences
- Need continuous seismic data
  - Best on broadband stations, but also usable picks on accelerometers, nodals, Raspberry Shakes
  - Use case: temporary deployment of broadband or nodal stations, and you want an automatically generated local earthquake catalog
- 😀 Advantages
  - No prior knowledge needed about earthquake sources or waveforms
  - Finds lots of small local earthquakes (lower Mc), with fewer false detections, than STA/LTA
  - Relatively easy to set up and run; reasonable runtime with parallel processing
- ☹️ Limitations
  - For out-of-distribution data sets (not in training data set): larger automated pick errors (0.1-0.5 s) & missed picks
  - Cannot pick phases completely buried in noise – not quite as sensitive as template-matching
  - Sometimes misses picks from larger earthquakes that are obvious to humans, for unexplained reason

Deep-learning vs. Template-matching?





Complementary methods; ideally use both?

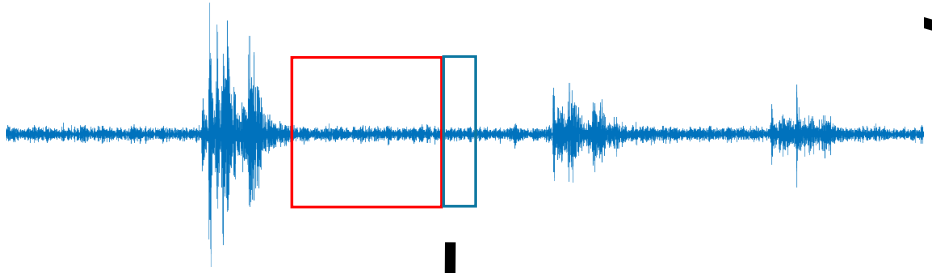
Deep-learning: finds smaller background seismicity

Template-matching: finds smallest events near already known earthquakes

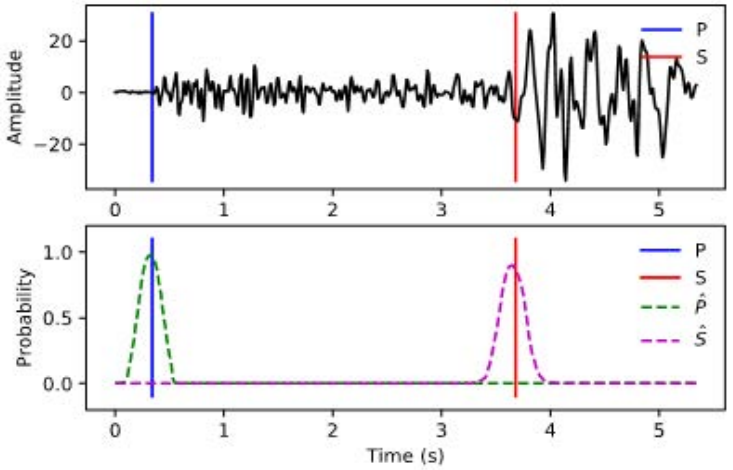
Yoon and Shelly (2024), TSR

Combining methods can mitigate limitations

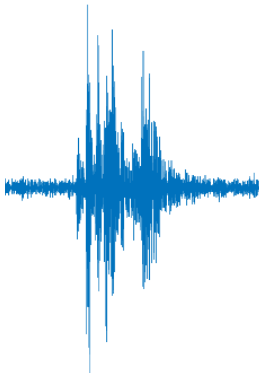
Event Detection (STA/LTA)



Deep-Learning Events, Phase-Picks



Template-Matching/
Cross-Correlation



??

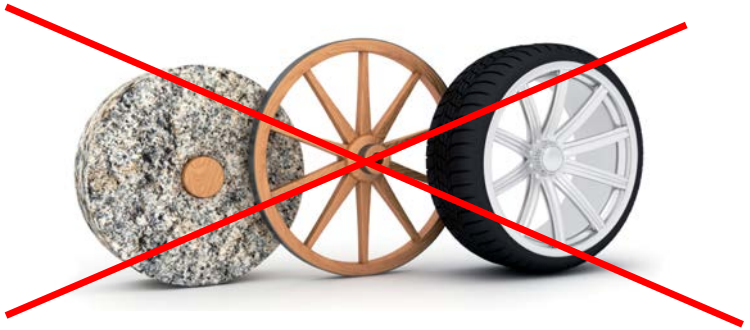
Template 1
Template 2
Template 3
Template 4
....
Template
N

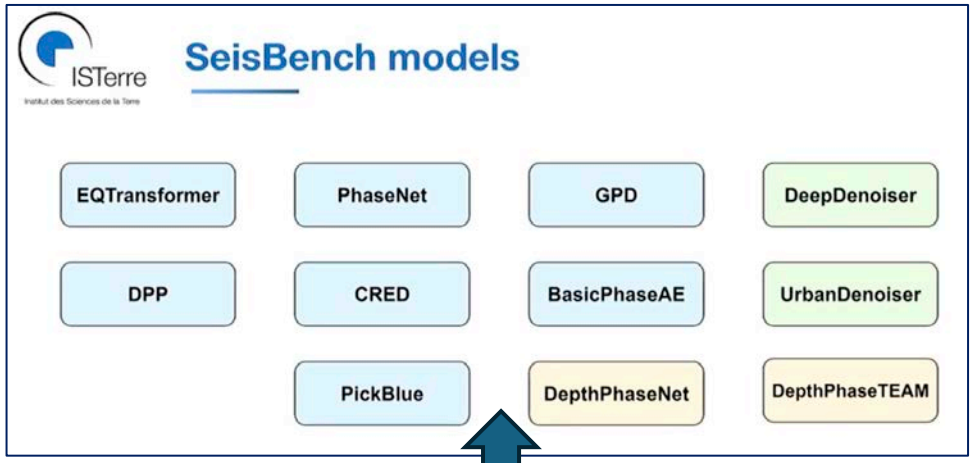
(+ merge correlated templates)

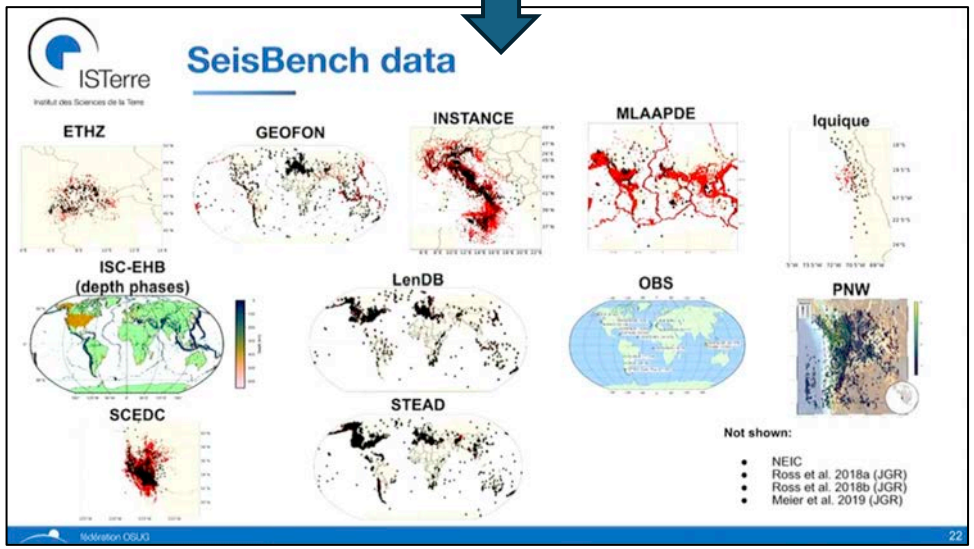
Result: More complete catalog than with any method alone

e.g., see Tepp, SRL, 2018

How can I get started with a deep-learning picker for my catalog?







https://www.nationalacademies.org/event/40915_11-2023_artificial-intelligence-machine-learning-geophysics-beyond-black-box-committee-on-solid-earth-geophysics-fall-2023-meeting#sectionEventMaterials

SeisBench
A toolbox for machine learning in seismology

https://github.com/seisbench/seisbench

Deploy ML pickers

Open in Colab

Rapidly deploy ML models integrated in SeisBench to seismic streams.

Using DeepDenoiser

Open in Colab

Denoising seismic waveform streams with DeepDenoiser in SeisBench.

Picking depth phases and determining earthquake depth

Open in Colab

Determining earthquake hypocentral depth using depth phase in SeisBench.

Training ML models on seismic data

Open in Colab

Get started with training deep learning routines (PhaseNet) on a benchmark seismic dataset in SeisBench.

Creating a dataset

Open in Colab

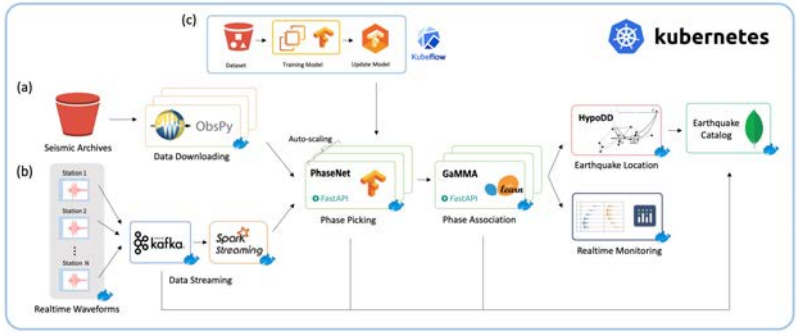
Learn how to create a dataset in SeisBench, using build-in functions and the obspy FDSN client as data source.

Building an event catalog with GaMMA

Open in Colab

blockly
eqtransformer

https://github.com/maihao14/BlocklyEQTransformer



https://github.com/AI4EPS/QuakeFlow

Deep-learning pickers: so many choices for models…

Check training data set! What's represented, what's not?

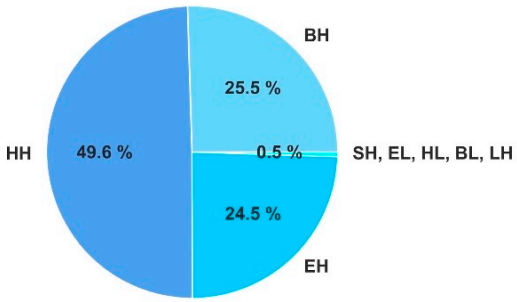
https://seisbench.readthedocs.io/en/stable/pages/benchmark_datasets.html

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

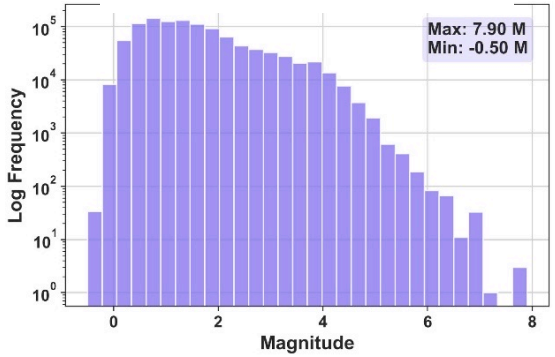
S. MOSTAFA MOUSAVI, YIXIAO SHENG, WEIQIANG ZHU, AND GREGORY C. BEROZA
Geophysics Department, Stanford University, Stanford, CA 94305-2215, USA

Mousavi et al. 2019, IEEE

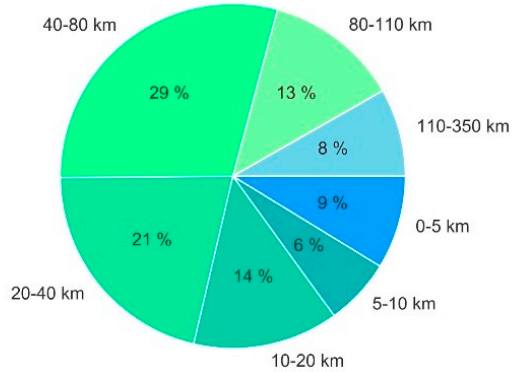
Instrument type?



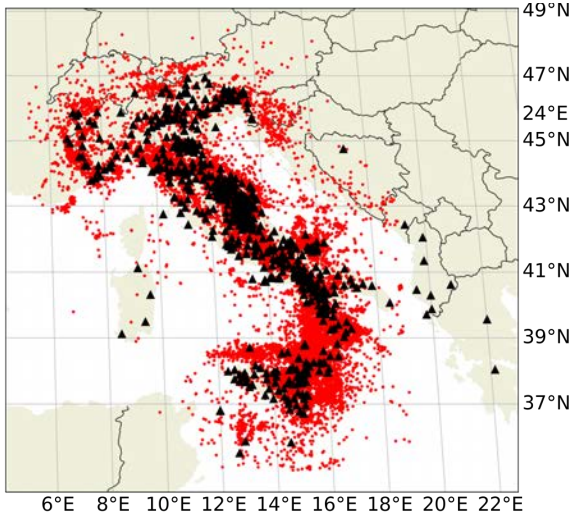
Magnitude range?



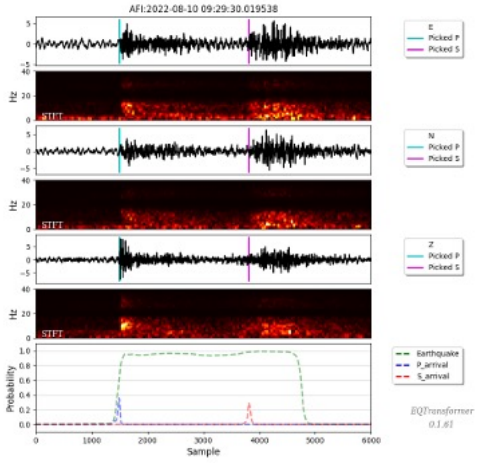
Event-station distances? (local, regional, teleseismic)?



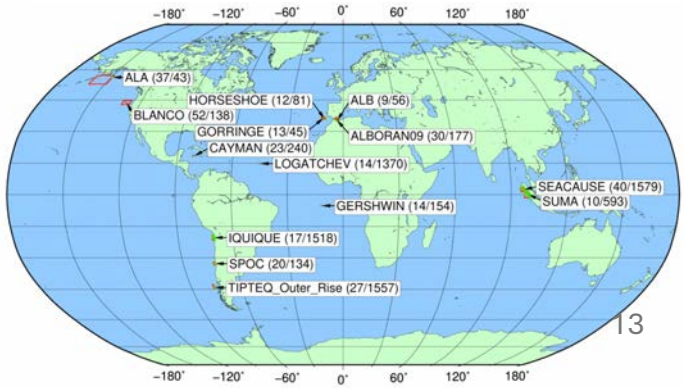
Specific region, or global? Only earthquakes, or noise too?



Input time window length? (PhaseNet: 30 s, EQTransformer: 60 s)



Ocean bottom seismometers?



Read benchmark papers comparing different methods for catalog workflow

Event Detection & Phase Picking

Munchmeyer et al. 2022, JGR

A Mitigation Strategy for the Prediction Inconsistency of Neural Phase Pickers

Yongsoo Park*[1], Gregory C. Beroza[2], and William L. Ellsworth[2]

Park et al. 2023, SRL

Parametric Testing of EQTransformer's Performance against a High-Quality, Manually Picked Catalog for Reliable and Accurate Seismic Phase Picking

Olivia Pita-Sllim*[1], Calum J. Chamberlain[1], John Townend[1], and Emily Warren-Smith[2]

Pita-Sllim et al. 2024, TSR

Event Association

A Comparison of Machine Learning Methods of Association Tested on Dense Nodal Arrays

Colin N. Pennington*[1], Ian W. McBrearty[2], Qingkai Kong[1], and William R. Walter[1]

Pennington et al. 2025, SRL

Benchmarking seismic phase associators: Insights from synthetic scenarios

Puente et al. 2025, arXiv

Jorge Puente[1,2], Christian Sippl[1], Jannes Münchmeyer[3], Ian W. McBrearty[4]

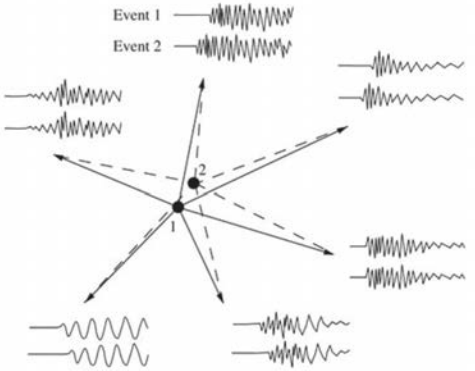
Event Location

Accuracy and Precision of Earthquake Location Programs: Insights from a Synthetic Controlled Experiment

Yifan Yu*[1], William L. Ellsworth[1], and Gregory C. Beroza[1]

Yu et al. 2024, SRL

When should I use **relative location methods** for a **relocated** earthquake catalog?



- Can use triggered event waveforms (also continuous data)
- 😀 Advantages
  - Sharpens up seismicity trends; delineates active fault structures at depth (difficult to get any other way)
  - Relative location uncertainties can be very low (meter-scale)
- ☹️ Limitations
  - Absolute locations are still uncertain
  - Larger and/or isolated earthquakes: waveforms less likely to be similar to those of other earthquakes, so they are lost (not relocated): **NOT a complete catalog**!
  - Pair-wise waveform cross-correlation: requires effort to extract waveforms around picks, computationally intensive

Catalog workflow steps & tips

- Decide that an existing catalog is not enough for your needs
- On short-duration subset of your data (1 day) with desired target (e.g. aftershock sequence), create initial catalog with all workflow steps
  - Select existing deep-learning picker model, pre-trained on appropriate dataset (SeisBench)
  - Select event association & location method with appropriate velocity model
  - Select magnitude equation for your region
  - Quality control on initial catalog, post-process to eliminate false detections & unwanted signals
- Happy with catalog now?
  - **If not, iterate.**
    - Transfer learning, or even re-training picker model: can be worth it (if you have labeled data/picks), but much more effort!
  - If so, run on entire dataset (decades? 100's-1000's of stations?)

Final thoughts

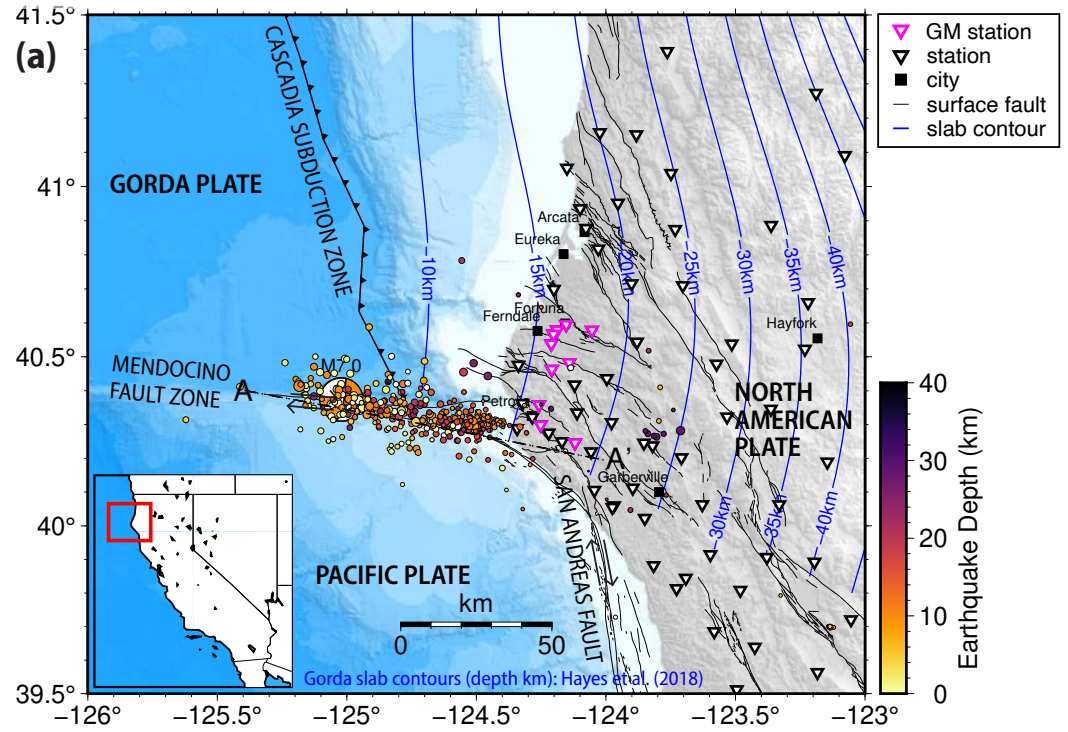
- This workshop is meant to get you started on the journey to
  - understand fundamentals of earthquake catalogs, their uses, their quality
  - develop & evaluate enhanced high-resolution earthquake catalogs.
- Have fun exploring!
  - We look forward to learning about your earthquake catalogs & insights gained from them in future conferences and publications
  - Share your successes, but even more your failures – we often learn more from things that didn't work.
  - Methods (especially machine-learning) for enhancing catalogs are changing rapidly
- Thank you for attending our workshop!!!

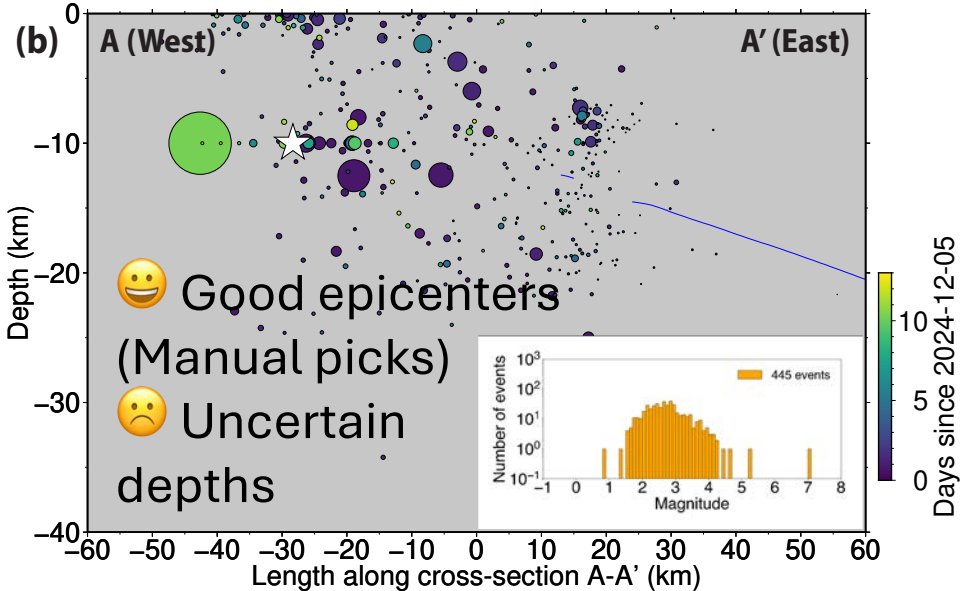
Further reading

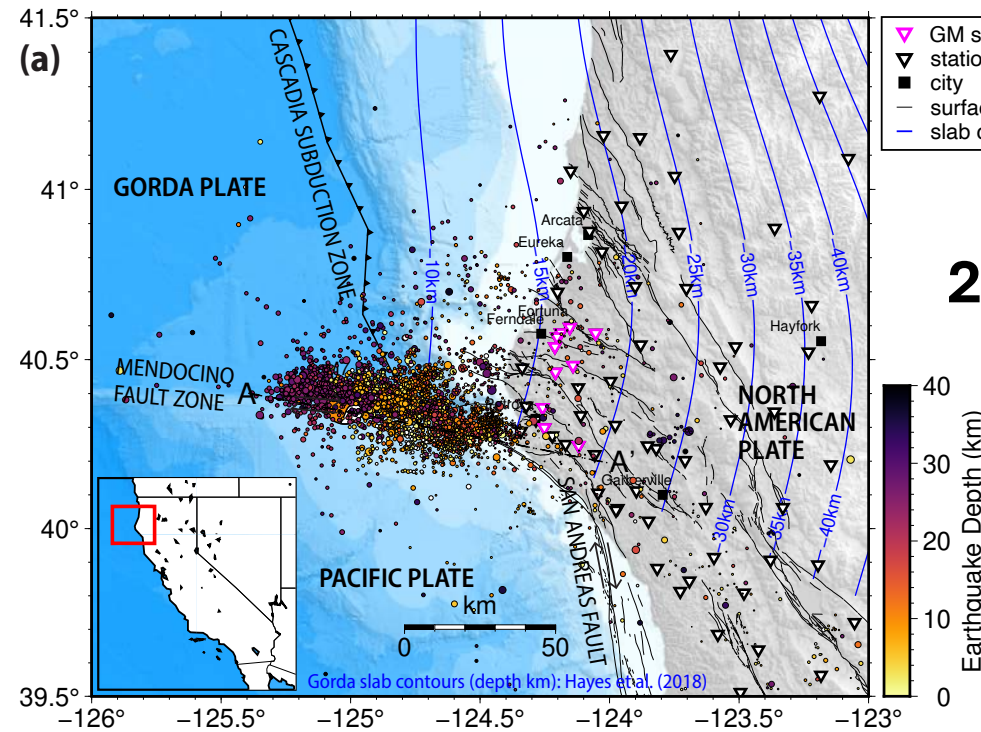
Supplementary Slides

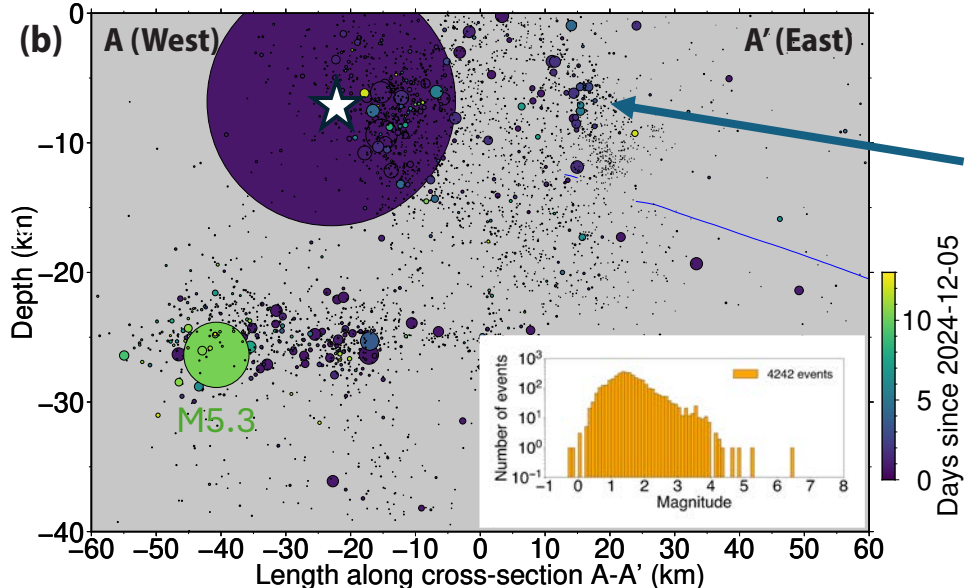
Reference catalog: NCSN/ComCat

Enhanced catalog: absolute locations









2024-12-05 M_w 7.0 Mendocino Aftershock Sequence

2024-12-05 00:00:00 to 2024-12-18 00:00:00 UTC (13 days)

☹ Uncertain locations (from pick errors)

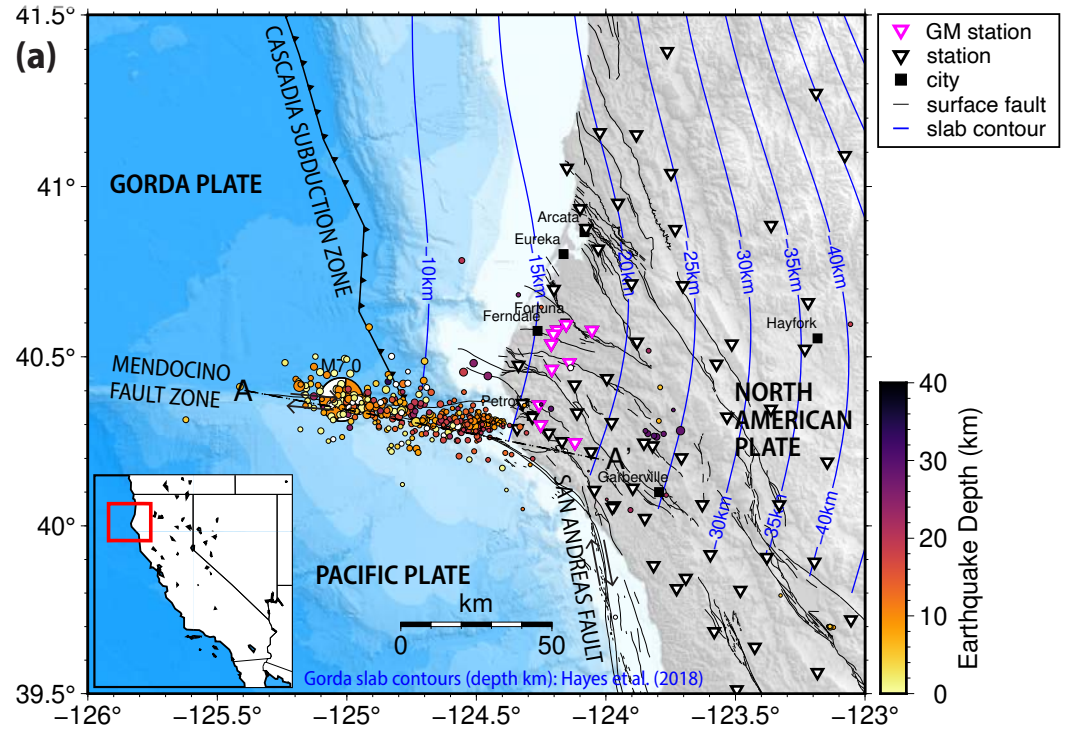
😀 Better depths

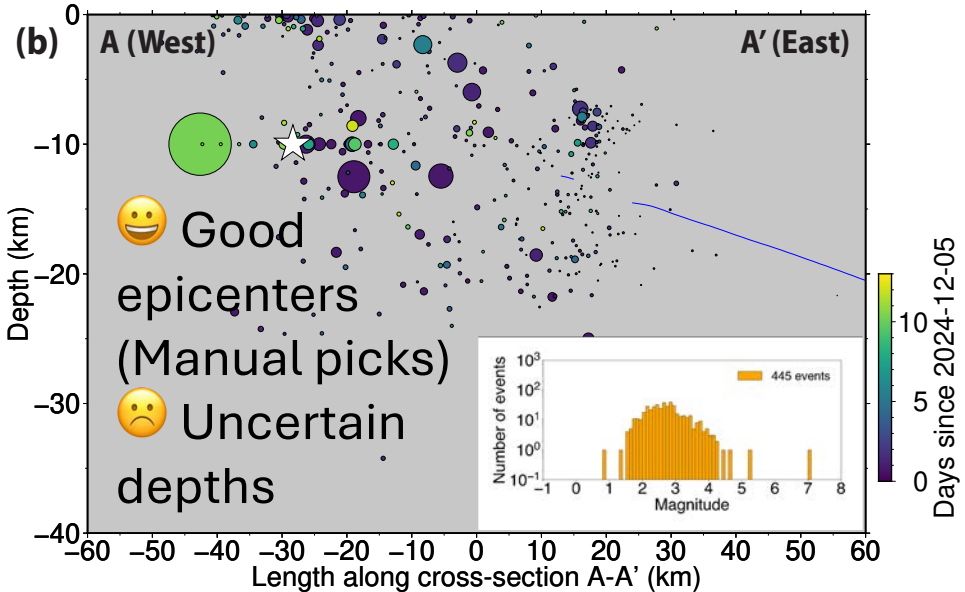
😀 More complete to lower magnitude

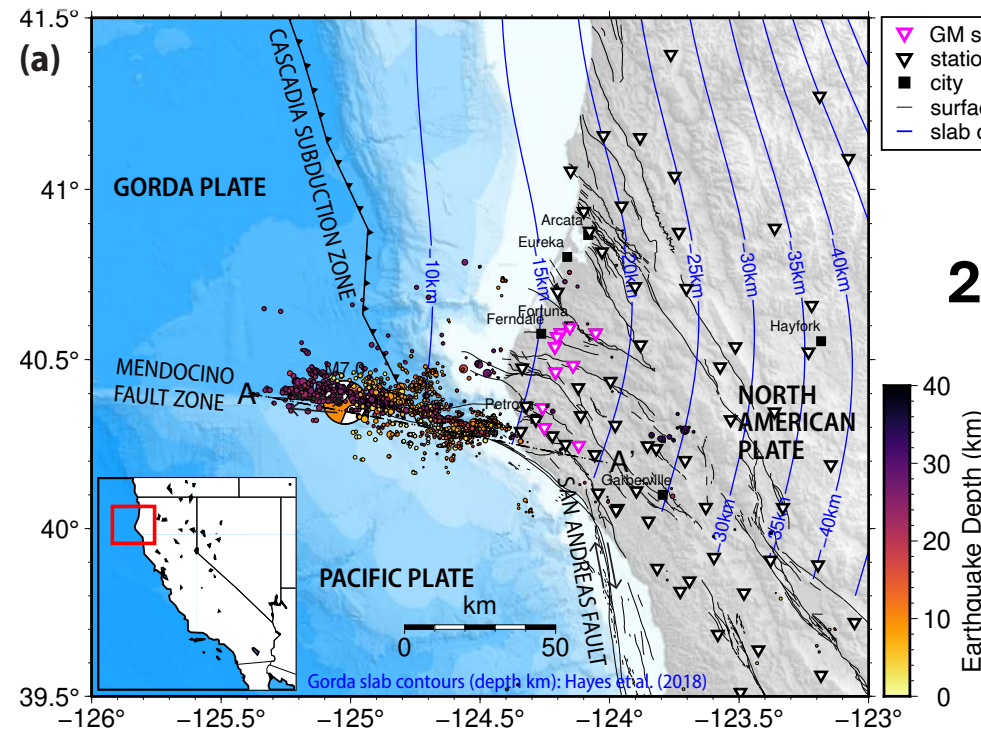
Pollitz et al. (2025), submitted to GRL

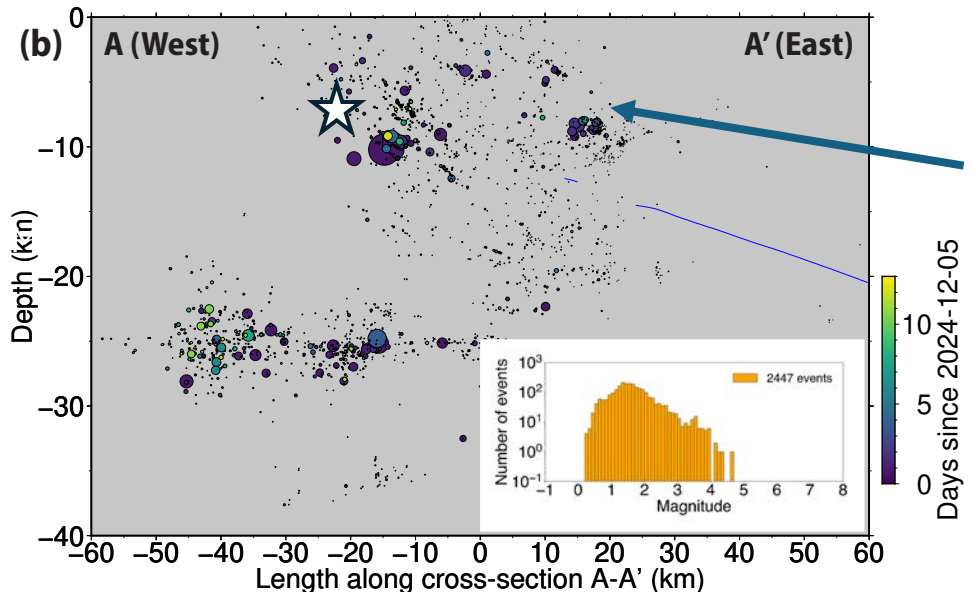
Reference catalog: NCSN/ComCat

Enhanced + relocated catalog: precise relative locations









2024-12-05 M_w 7.0 Mendocino Aftershock Sequence

2024-12-05 00:00:00 to 2024-12-18 00:00:00 UTC (13 days)

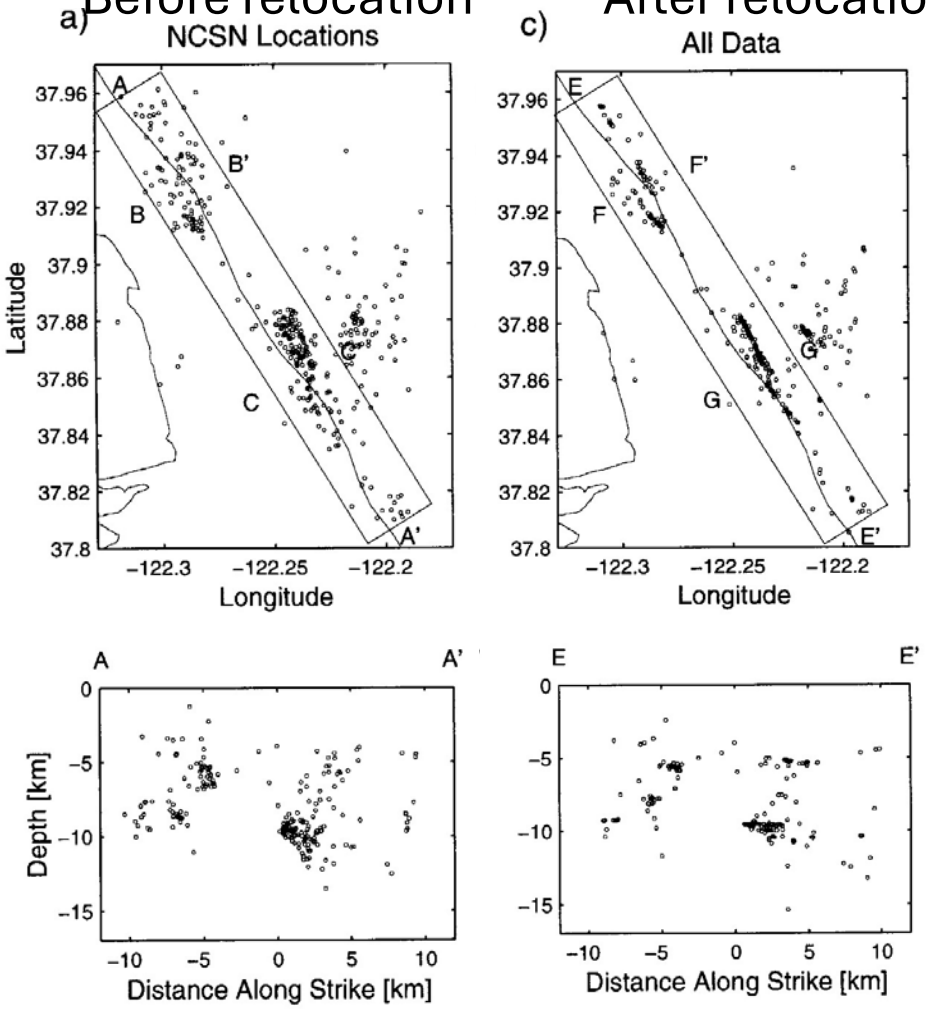
😀 Sharper locations
😀 Better depths
☹ Missing larger or isolated events; not complete catalog!

Pollitz et al. (2025), submitted to GRL

HypoDD – Double Difference

Before relocation

After relocation



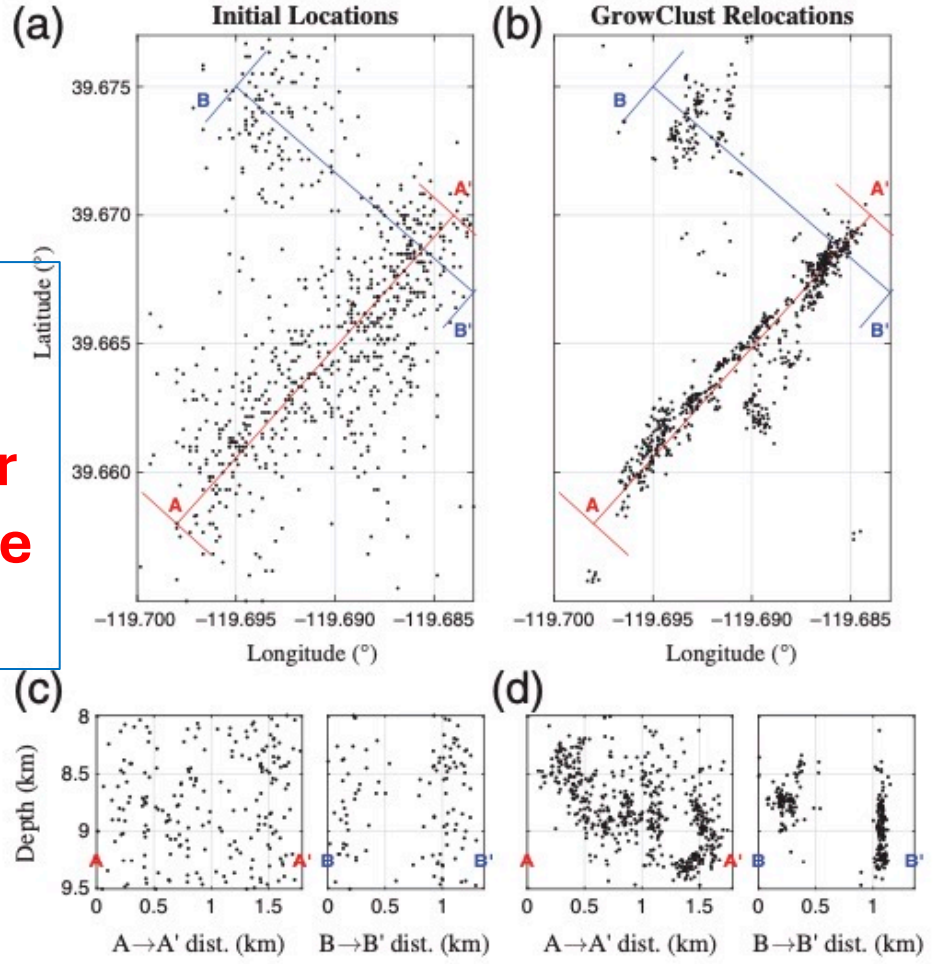
Waldhauser and Ellsworth (2000)

Relocation results are overall similar between these 2 methods

GrowClust

Before relocation

After relocation



Trugman and Shearer (2017); Trugman et al. (2022)