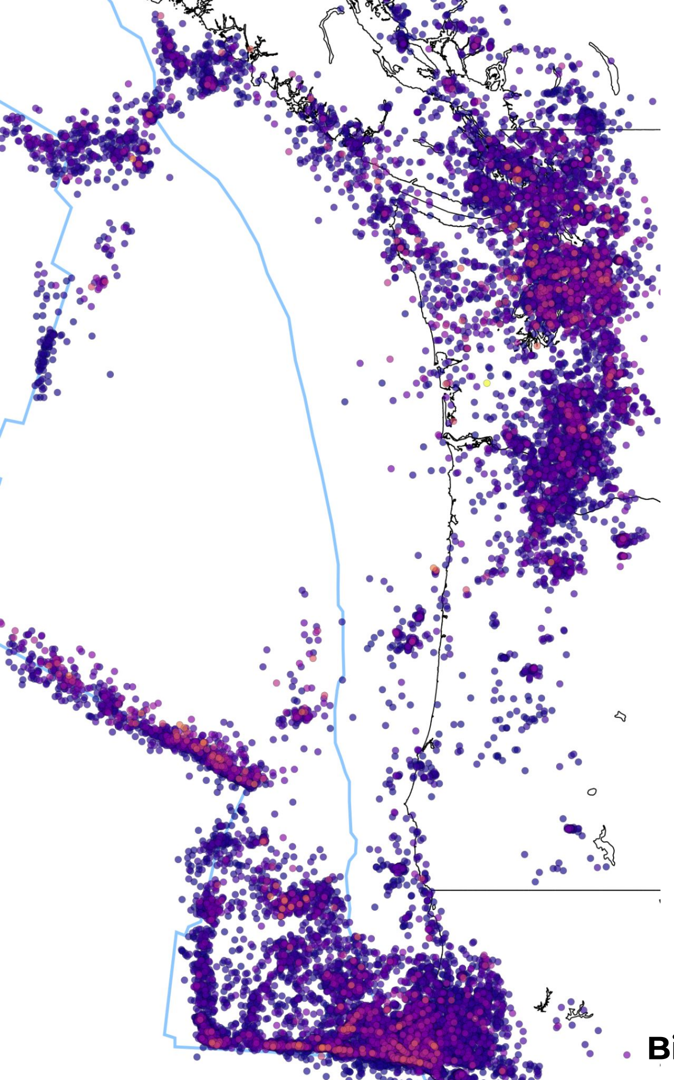
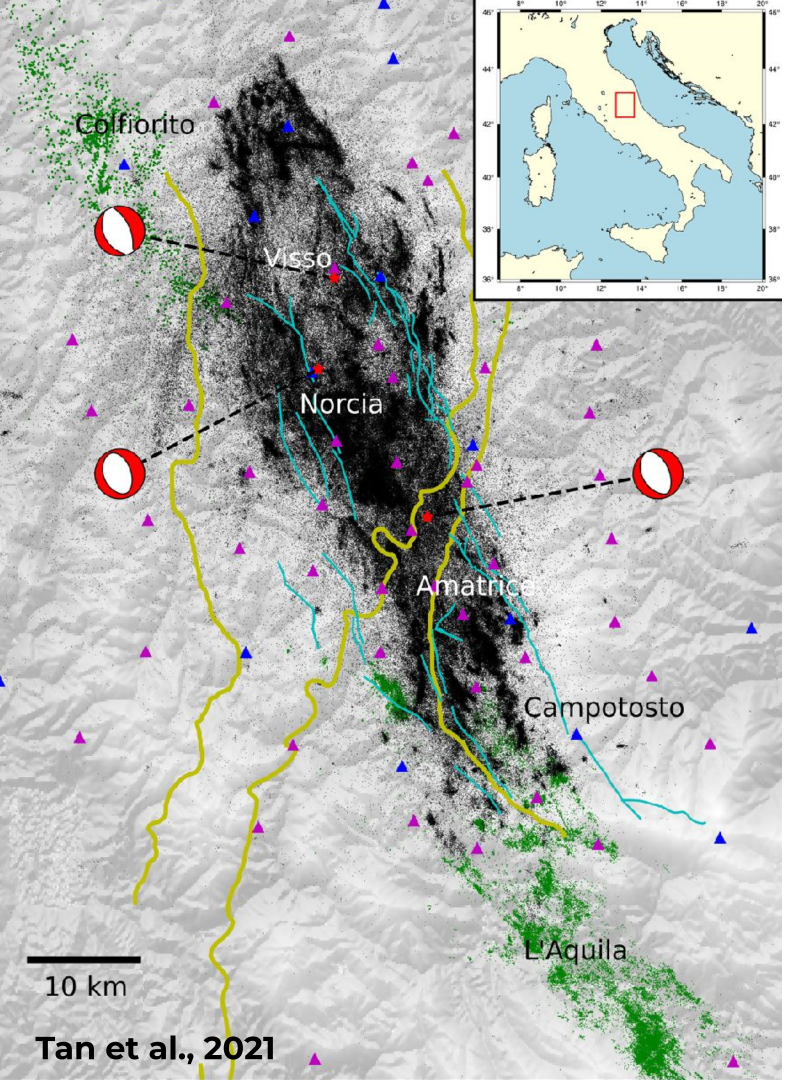
Earthquake Catalog Building aided by AI

Marine Denolle, Yiyu Ni, Jannes Munchmeyer, Ian Wang
Ian McBreatry, Greg Beroza
Amanda Thomas, Loic Bachelot
Alex Hamilton, Robert Weekly, Chad Trabant



Discovering *earthquakes* and *active faults* in remote, noisy environments

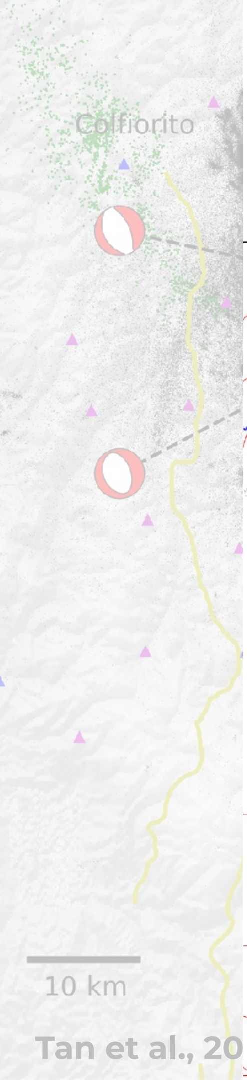
Bito, Denolle, et al (in prep)

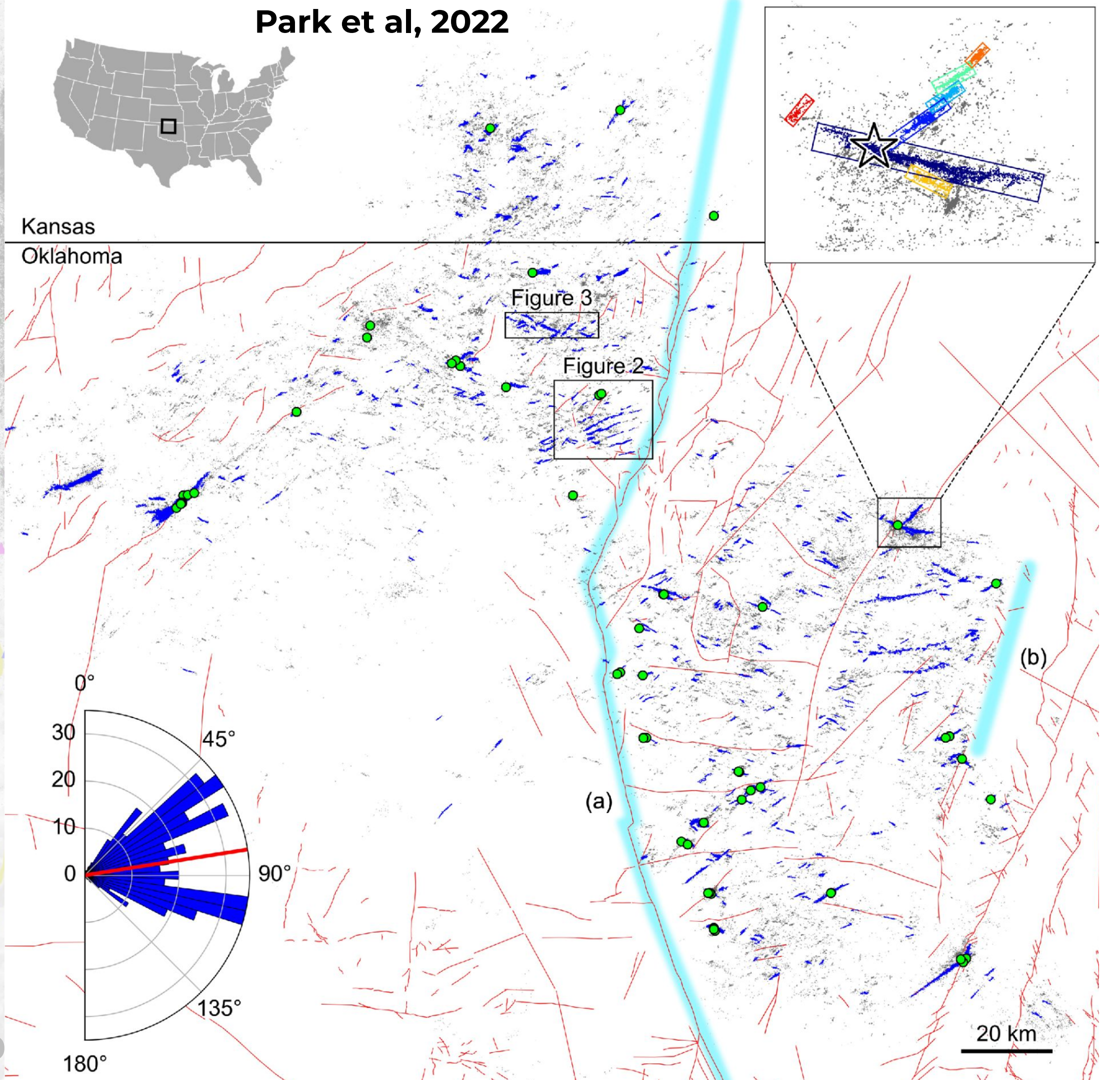


Tan et al., 2021

Tracing fault networks **to understand the complex earthquake sequences**

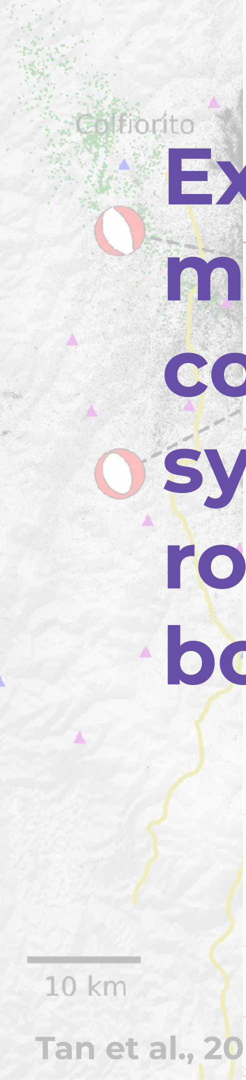
~ 1M earthquakes in 1 year

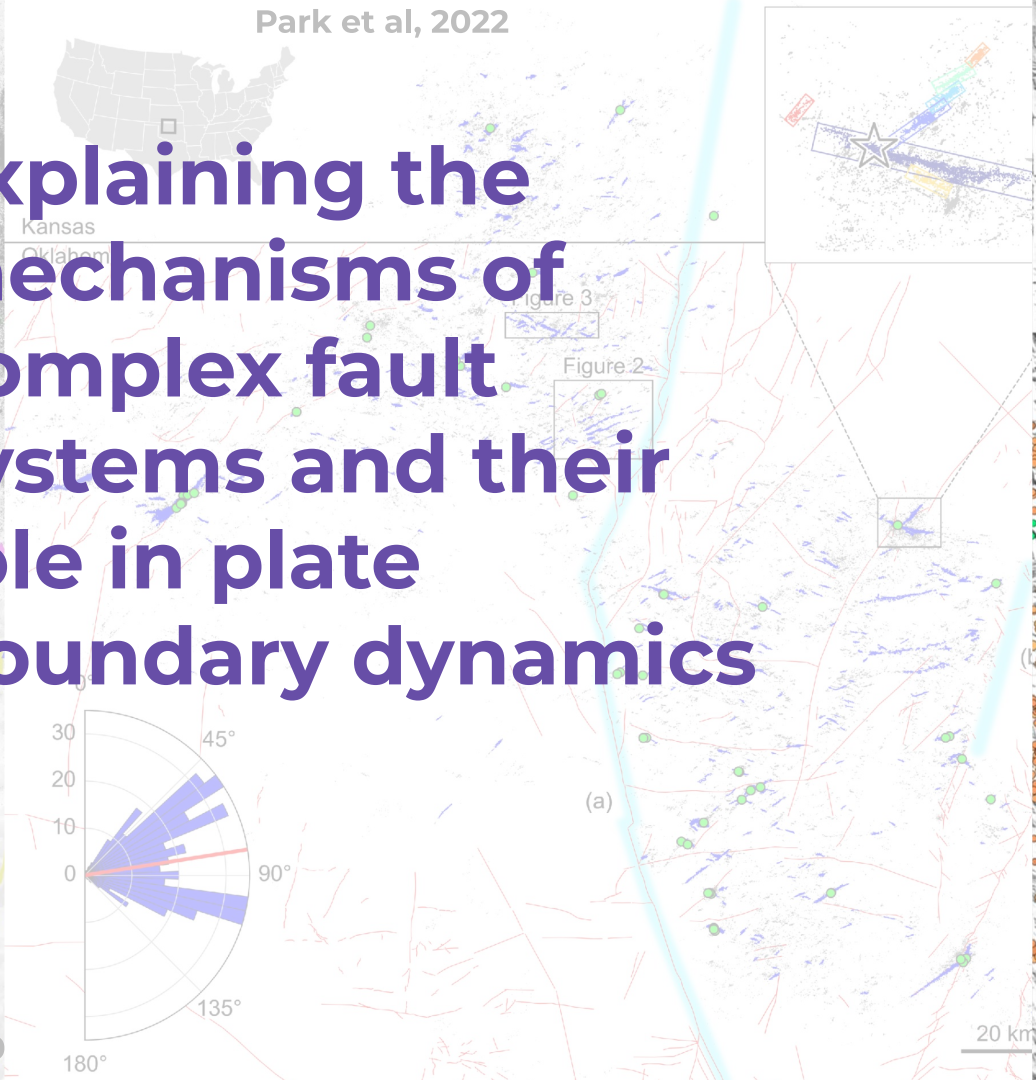


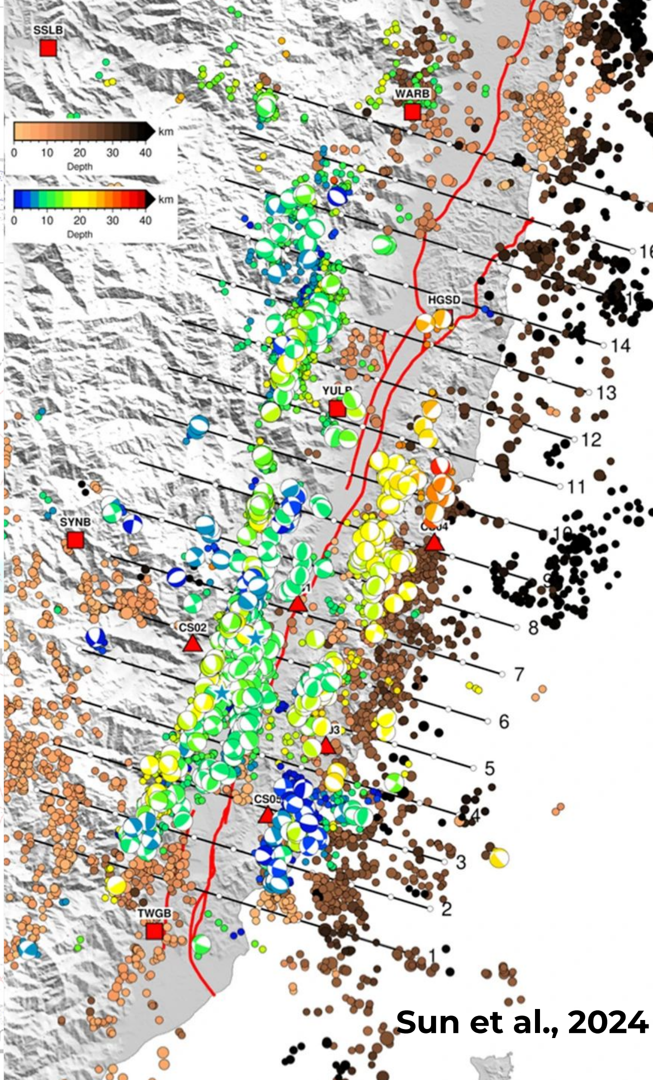


Unearthing potential for M5+ intraplate earthquakes activated by anthropogenic activities

Explaining the mechanisms of complex fault systems and their role in plate boundary dynamics



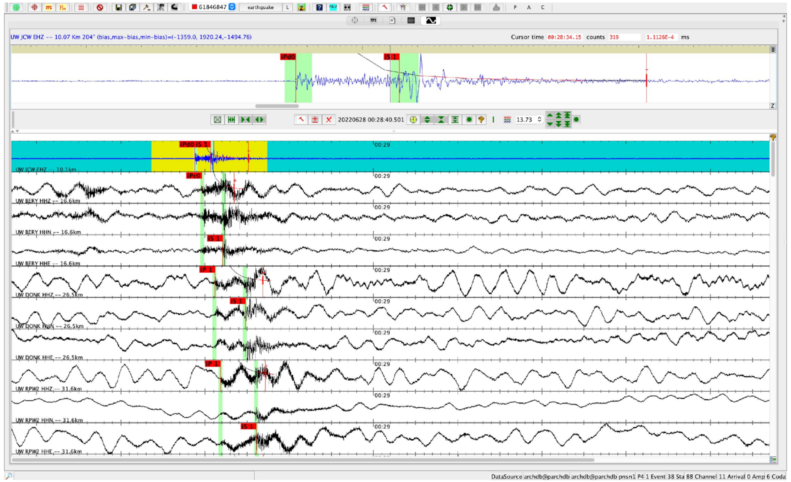




Machine Learning for building earthquake catalogs

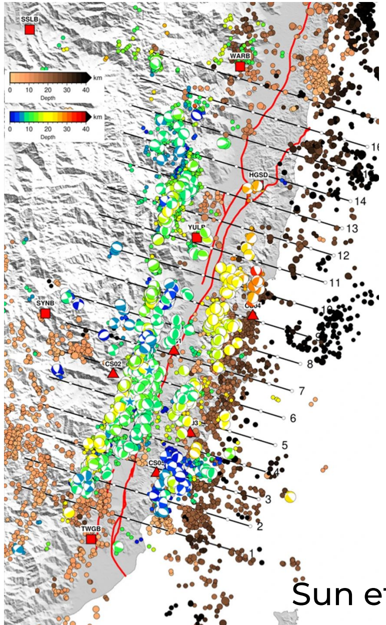
Data Preparation → Event Discrimination → Phase Picking → Phase Association → Location & *Relocation* → Magnitude (Md, Ml, Mw) → Focal Mechanism

Data received at seismic network



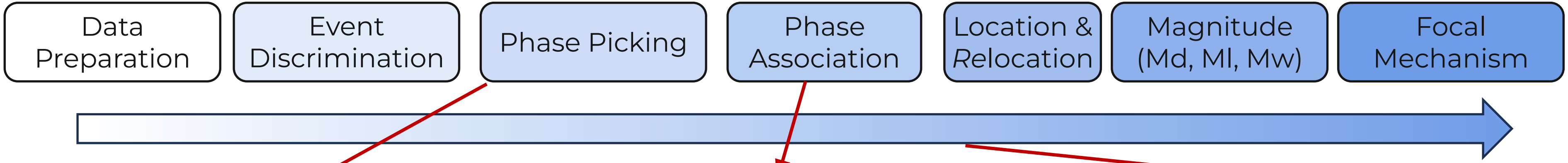
PNSN - Jiggle

Precise Earthquake Catalog



Sun et al., 2024

Machine Learning for building earthquake catalogs

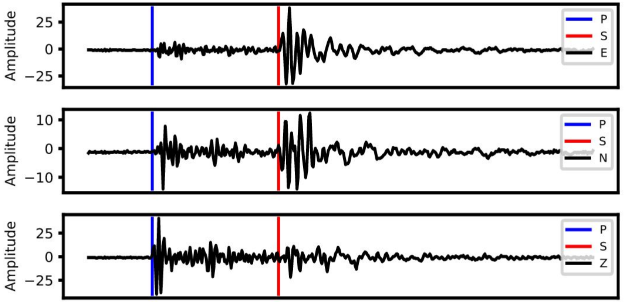


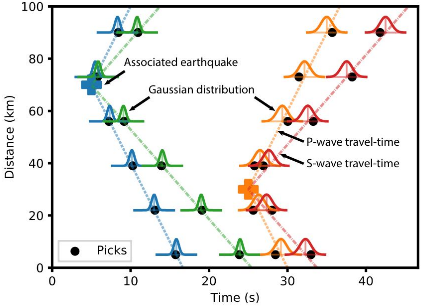
- GPD (Ross et al., 2018)
- PhaseNet (Zhu et al., 2018)
- EqTransformer (Mousavi et al., 2020)
- PhaseNO (Sun et al., 2022)
- ELEP (Yuan et al., 2023)

- PhaseLink (Ross et al., 2019)
- GaMMA (Zhu et al, 2021)
- Neuma (Ross et al., 2023)
- PyOcto (Münchmeyer, 2023)
- GENIE (McBreatry&Beroza, 2023)

Workflows

- easyQuake (Walter et al., 2021)
- SeisBench (Woollam et al., 2022)
- QuakeFlow (Zhu et al., 2022)
- Loc-Flow (Zhang et al., 2022)
- QuakeScope (Ni et al, *in prep*)

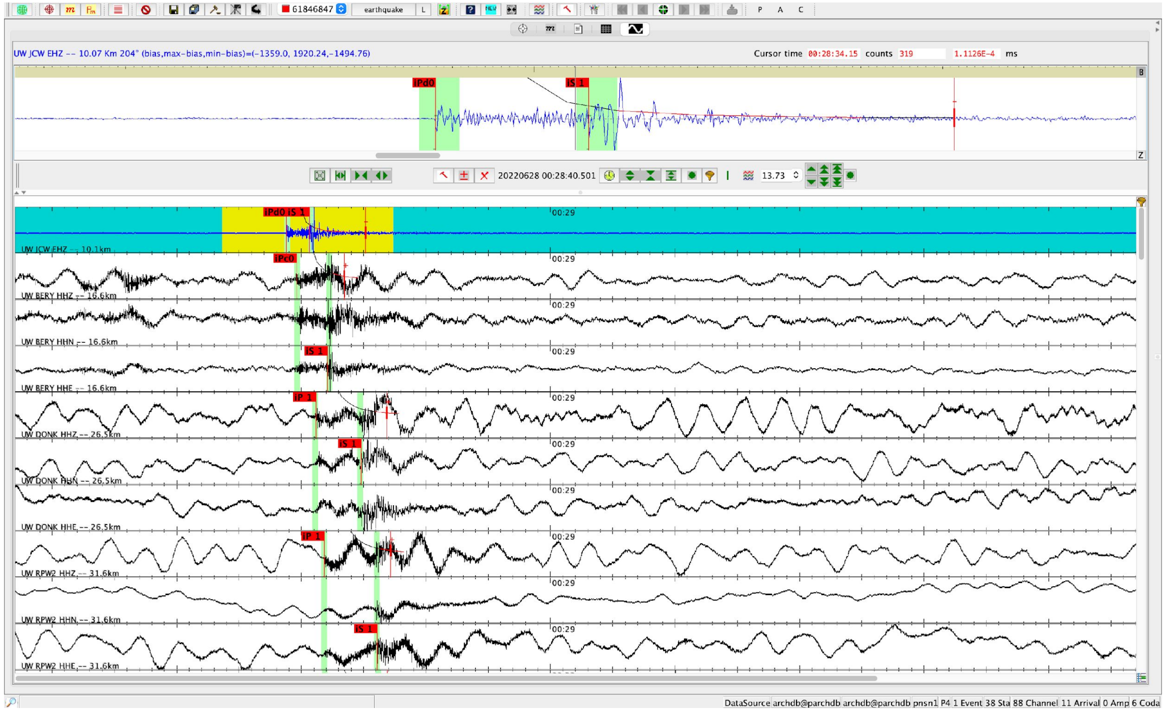




How a seismic network (PNSN) processes an event

PNSN analysts use **Jiggle** to pick phases, locate earthquakes, and calculate magnitude.

0 min	Trigger alarm
+0.5 min	Duty Seismologist notified
+1 min	Jiggling
+3-10 min	Finalize event and send to ANSS/ComCat



- False triggered events
- Limited stations processed
- Extra training time for new analysts to pick with quality and consistency

Seismic Curated Data Sets

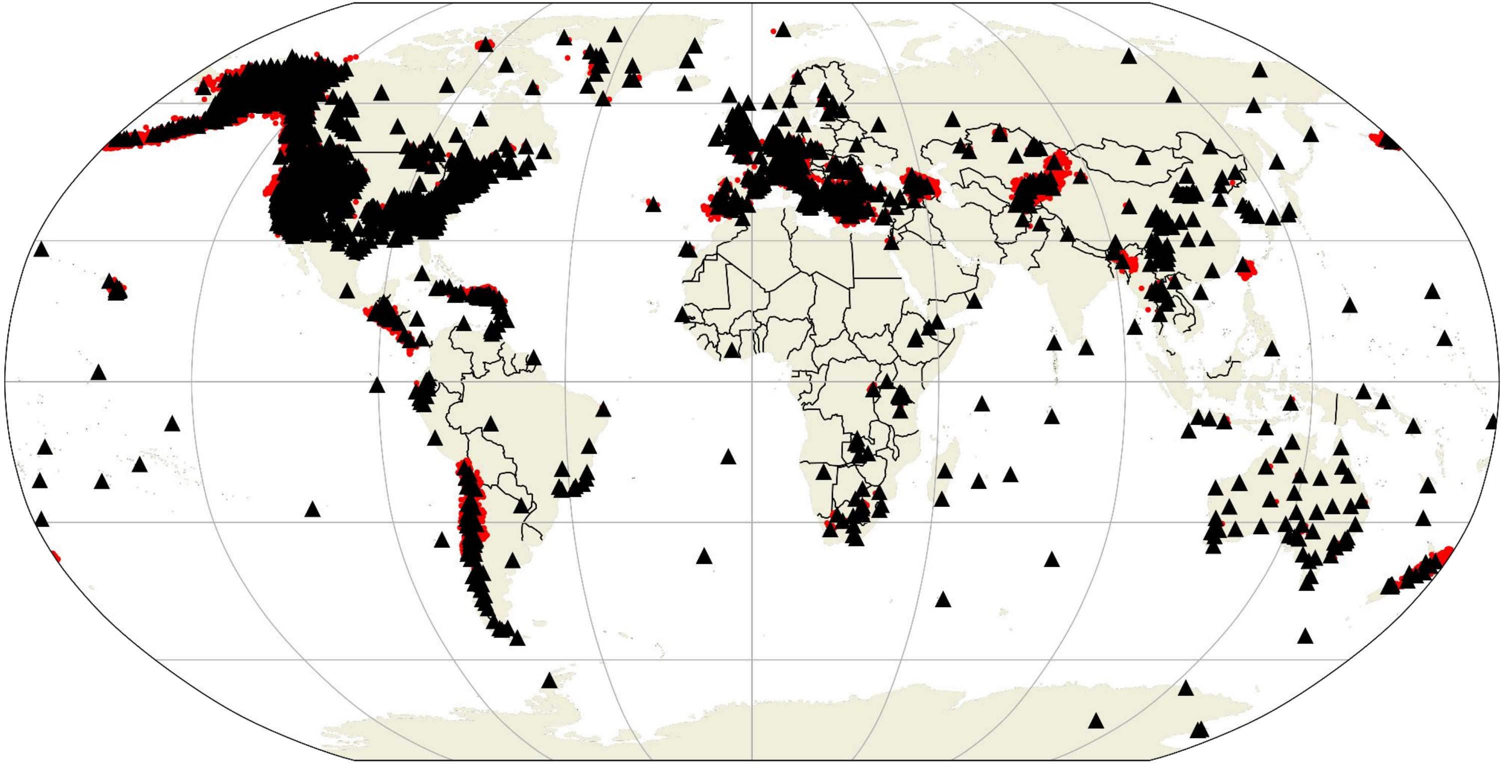
STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI

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

Corresponding author: S. Mostafa Mousavi (mmousavi@stanford.edu)

The work of S. M. Mousavi was partially supported by Stanford Center for Induced and Triggered Seismicity (SCITS). The work of G. C. Beroza was supported by AFRL under the contract number FA9453-19-C-0073.

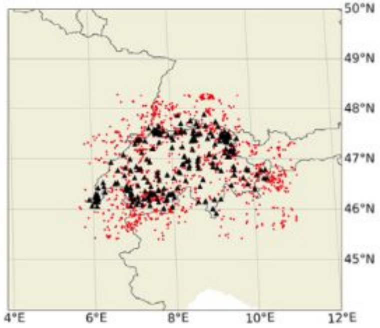
- 1.2M waveforms and attributes
- Earthquakes < 350 km



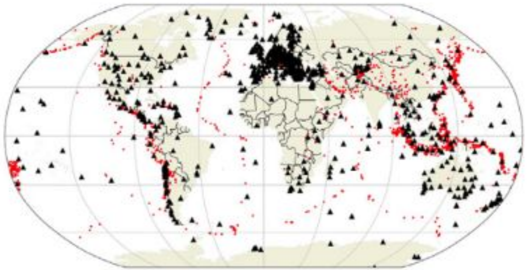
Seismic Curated Data Sets

Michelini et al. (2021) Woollam et al. (2019)

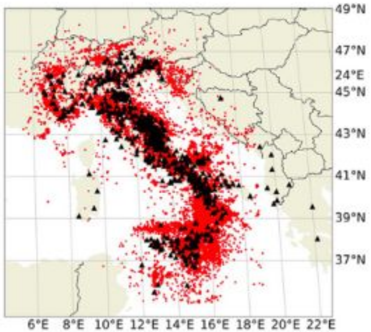
ETHZ



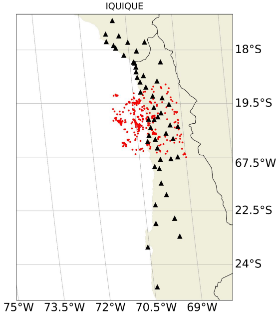
GEOFON



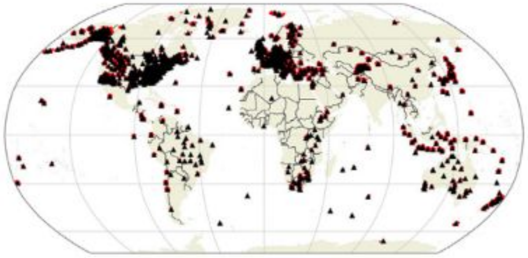
INSTANCE



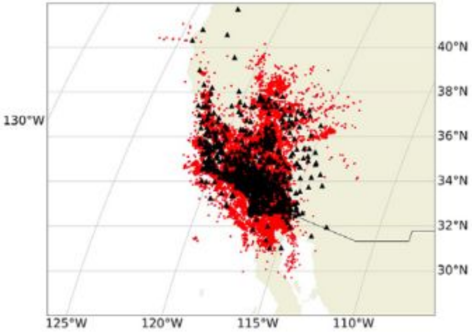
IQUIQUE



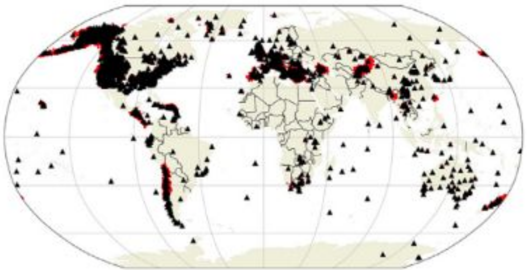
LenDB



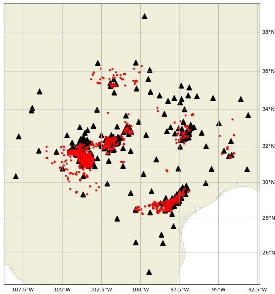
SCEDC



STEAD



TEXNED



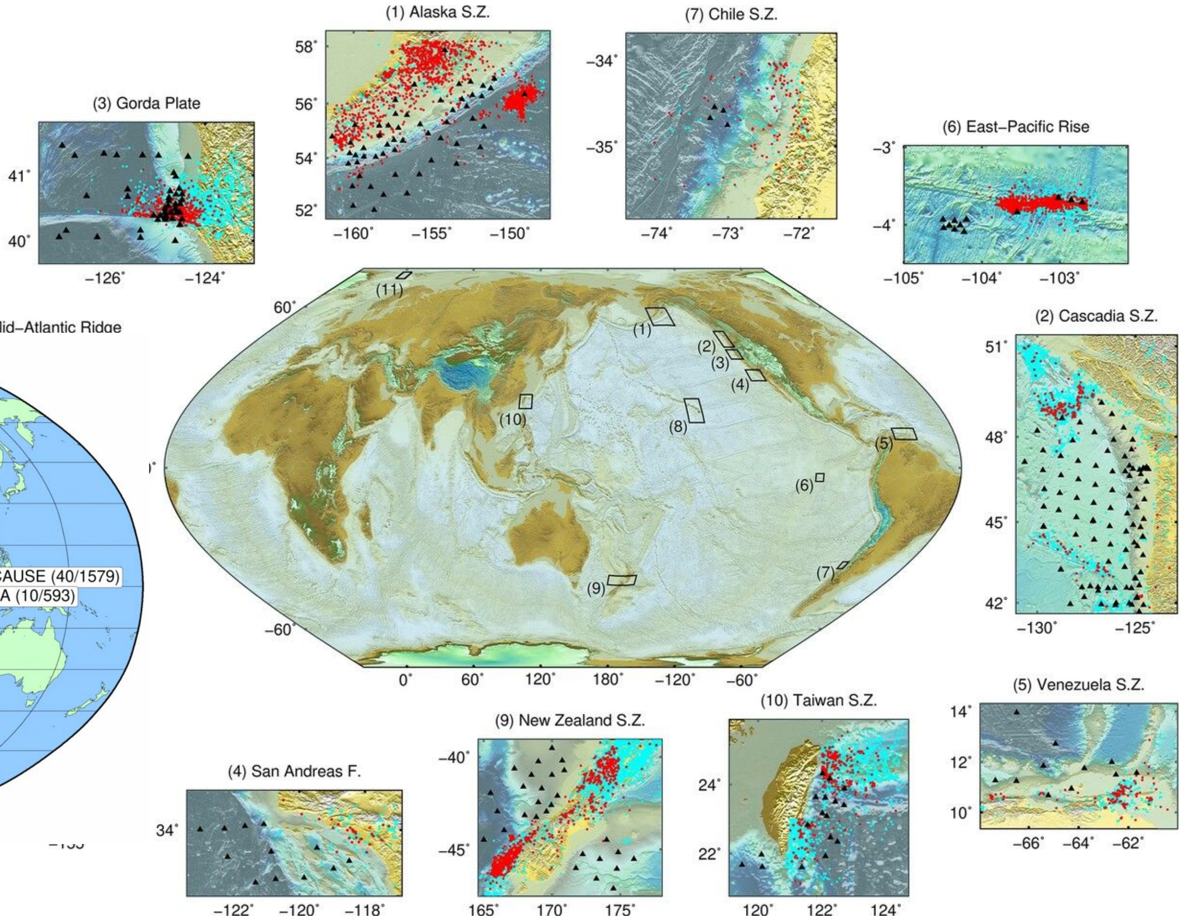
Magrini et al. (2020)

Mousavi et al. (2019)

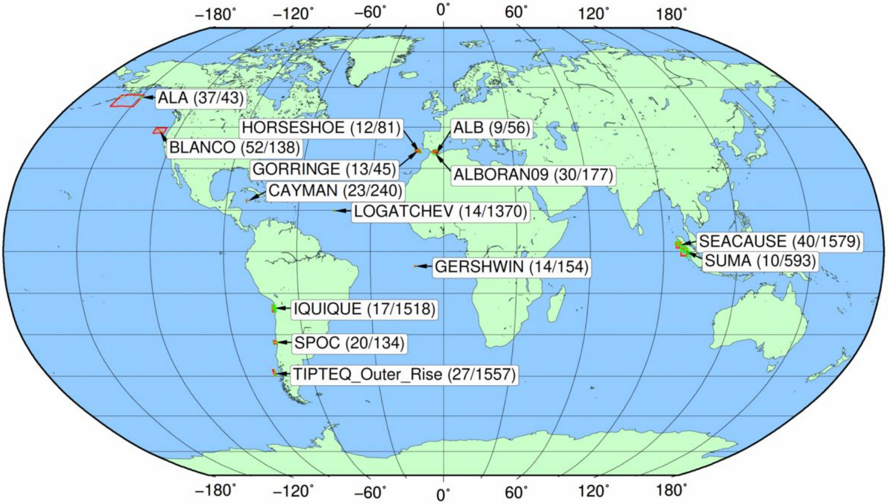
Chen et al. (2024)

Offshore Bottom Seismometer

OBST2024: Niksejel et al. (2024)



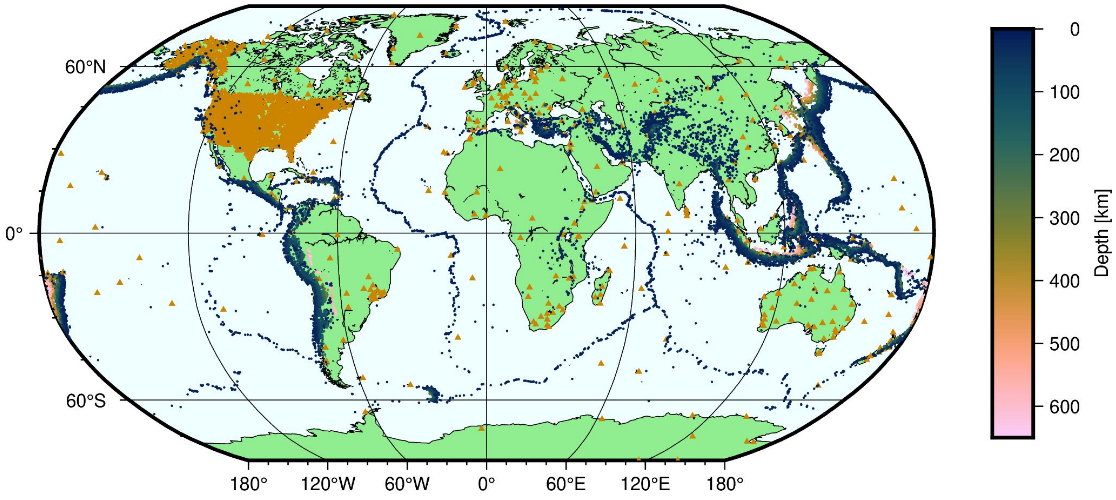
OBS: Bornstein et al. (2023)



Beyond direct P and S arrivals

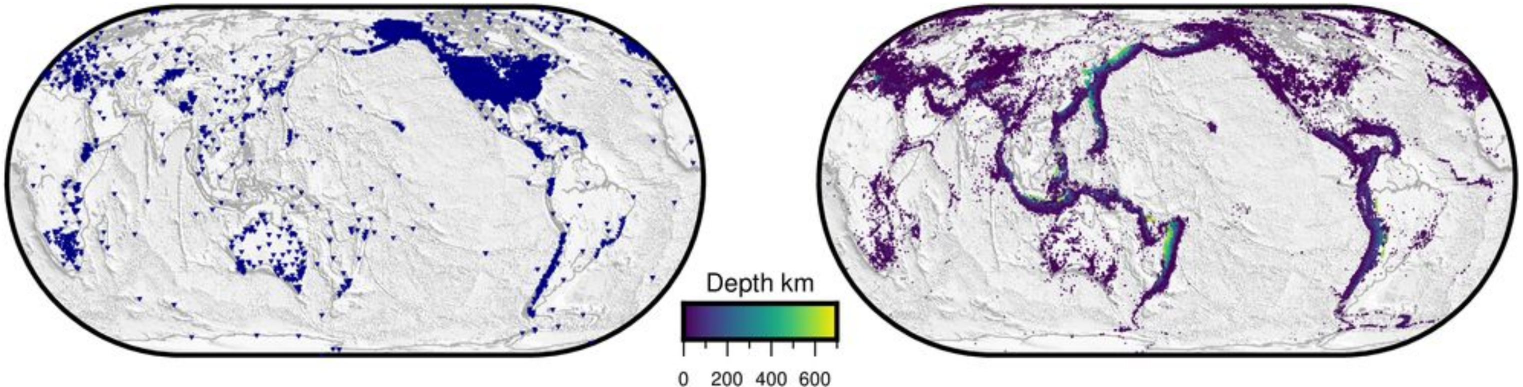
Depth Phases: pP, sP, pwP

Munchmeyer et al. (2023)



CREW: P, Pn, Pg, S, Sn, Sg

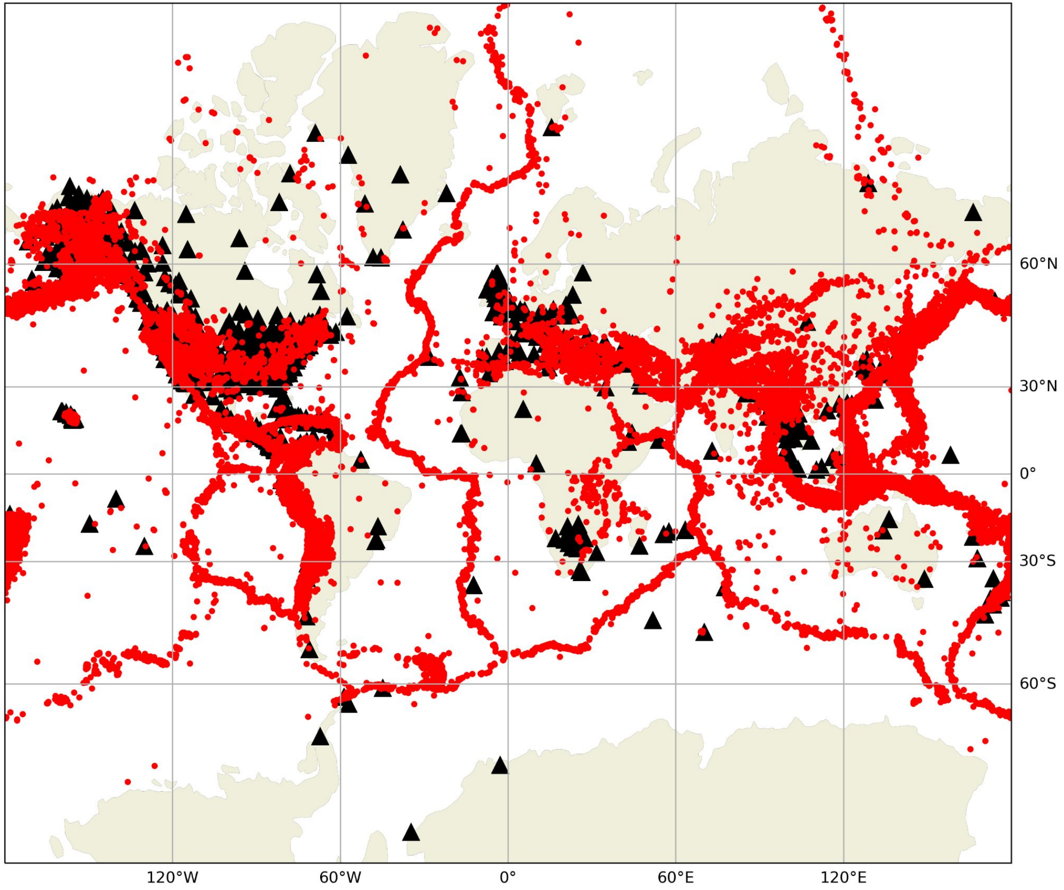
Aguilar-Suarez and Beroza (2024)



Beyond direct P and S arrivals

MLAADPE
P, Pn, Pg, S, Sn, Sg
(USGS)

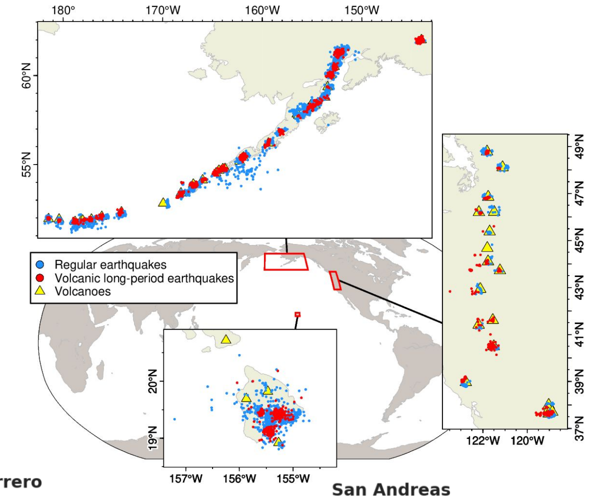
Cole et al. (2023)



Beyond earthquakes

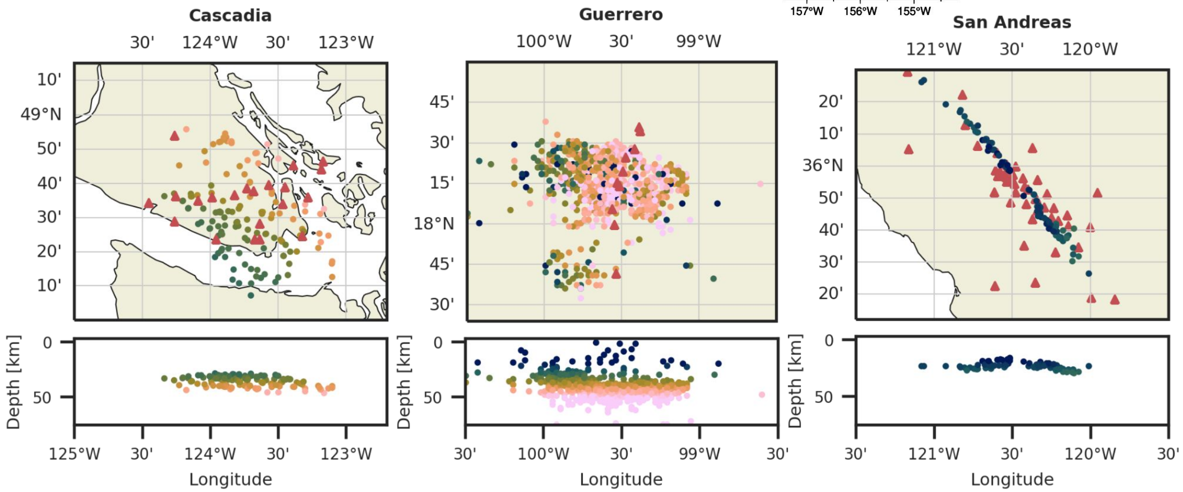
Volcanic LP events

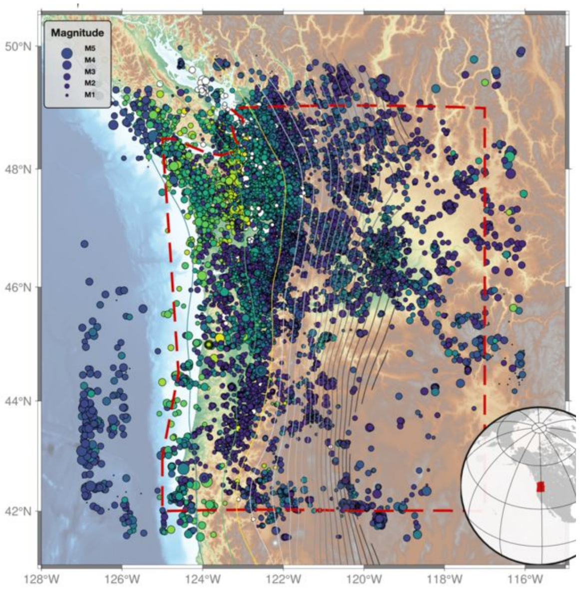
Zhong et al. (2024)



Quakes + Explosions + Surface Events

Ni et al. (2023)





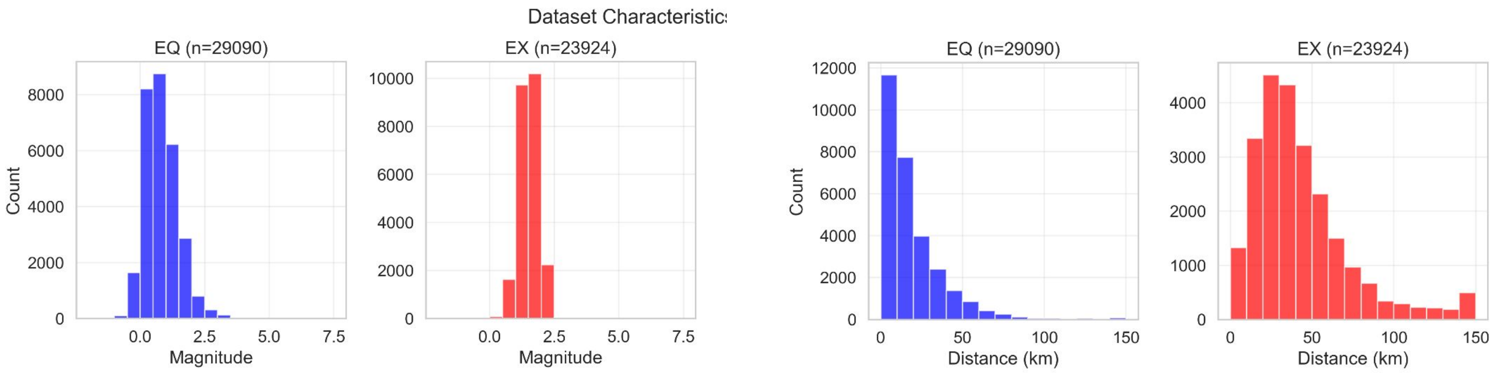
Stacked LFEs

Munchmeyer et al. (2024)

How diverse or balanced are these data sets?

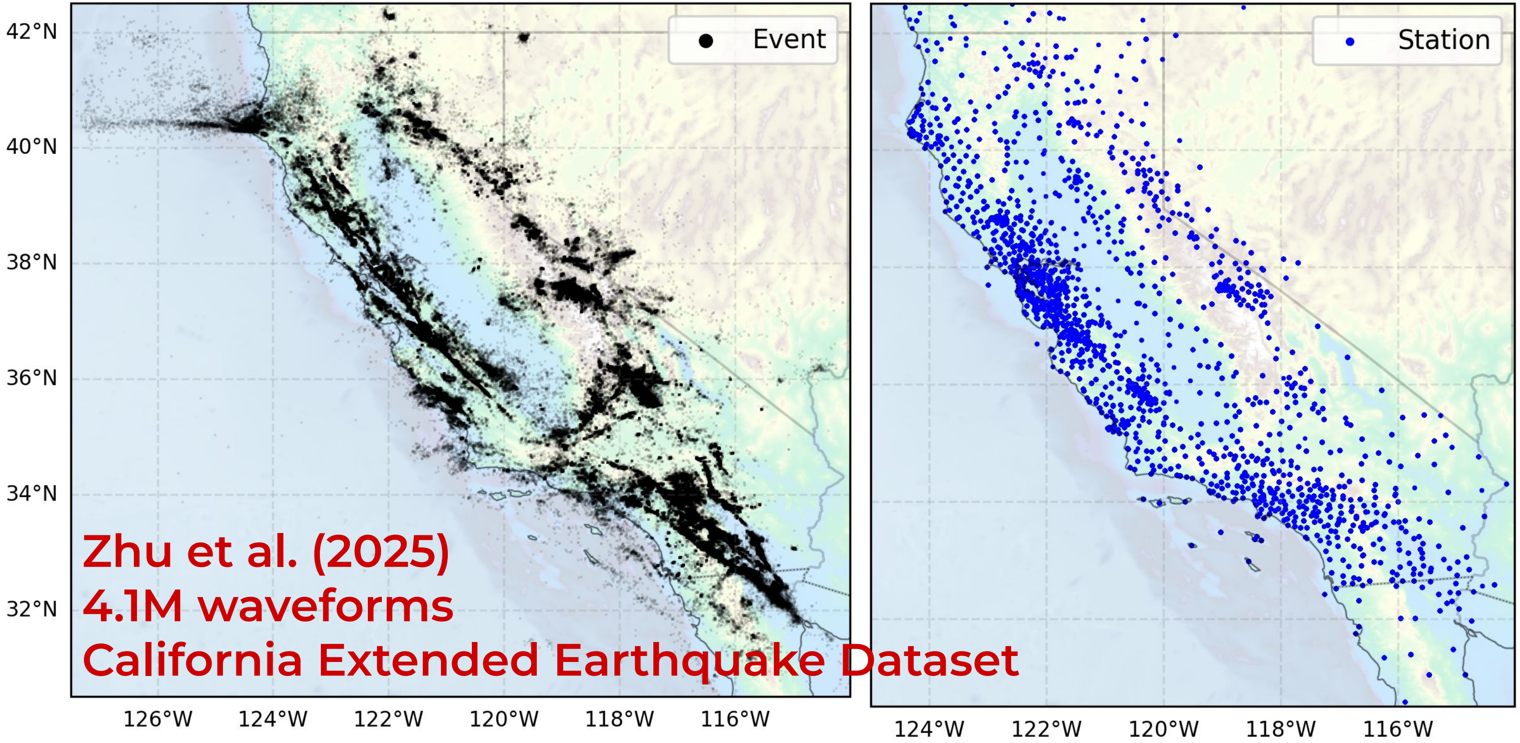
Very unbalanced

- **Very skewed distribution in magnitude** (fewer low M < 3 due to noise level, few M>5 due to their scarcity).
- **Distance and SNR covary** (selecting by SNR threshold removes long distance records)
- **Noise matters** (especially to remove false picks)
- **Event-type diversity is growing** (driven by increase in interdisciplinary and environmental seismology)



Benchmark data sets can be large

metadata.csv: data attributes used for information + labeling (output)
waveforms.h5: time series data (input) **~GBs -> TBs**

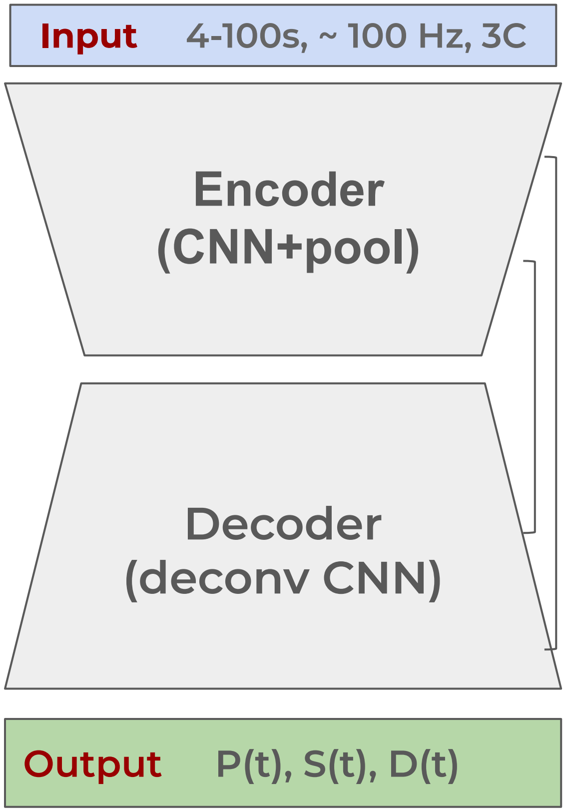


Zhu et al. (2025)
4.1M waveforms
California Extended Earthquake Dataset

Practicum: Explore Data Sets

Deep Learning for Phase Picking

Model Architectures - Many U-Nets



Taper, bandpass, normalize

- bi-LSTM (eqT)
- Dilated Conv (BasicPhaseAE)
- Self-attention/transformer (eqT)

Typically < 5M parameters

Peak-picking or threshold crossing

Pure U-nets
(PhaseNet, BasicPhaseAE, ARRU)

Hybrid CNN-RNN-Attenti on
(EqT)

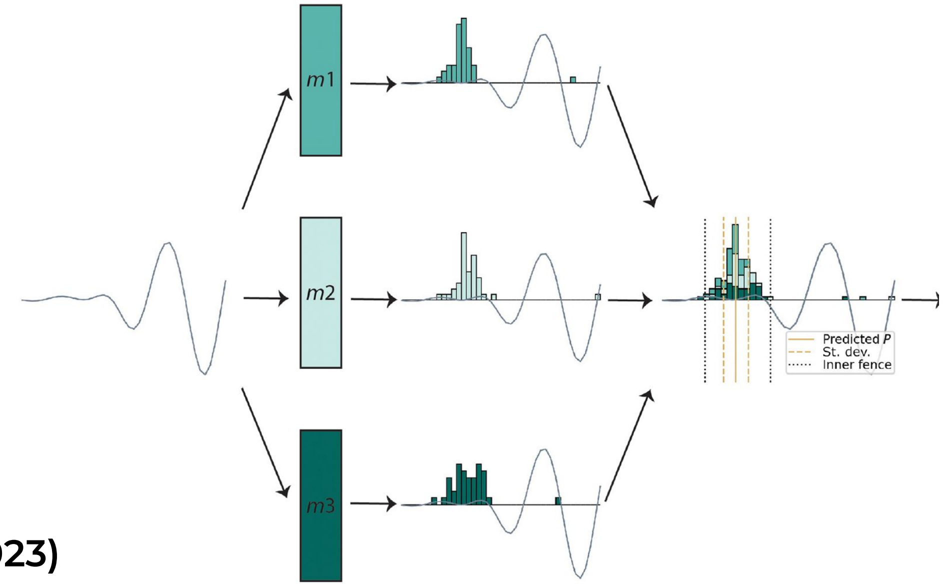
Ultra-light CNN
(GPD, PickNet) great for IoT

Domain-specific
Add hydrophone channel as input (PickBlue)

What is the uncertainty of picks?

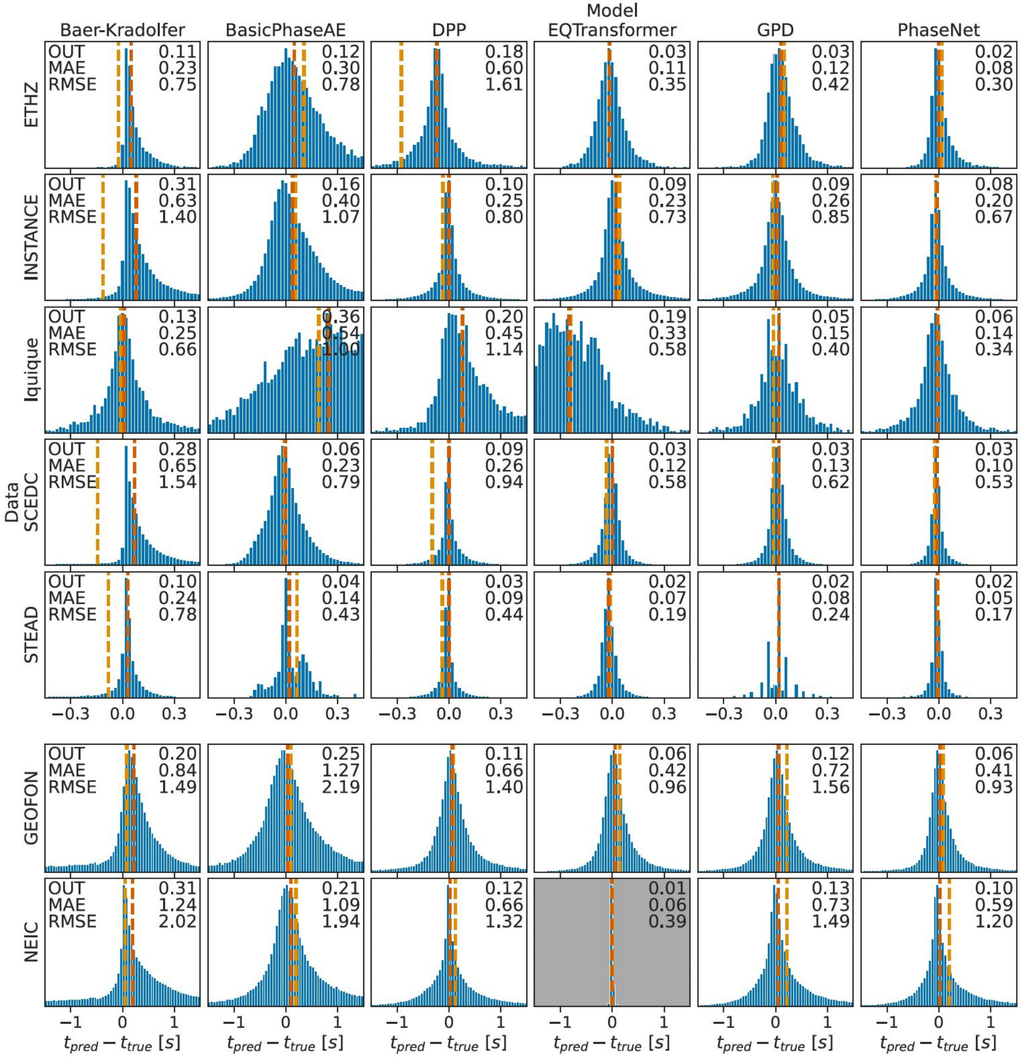
- Uncertainties from seismic analyst are typically not included in the training.
- Most labeling assumes a fixed gaussian/triangle uncertainty

- DL model uncertainties can be estimated using **Dropout** in inference.



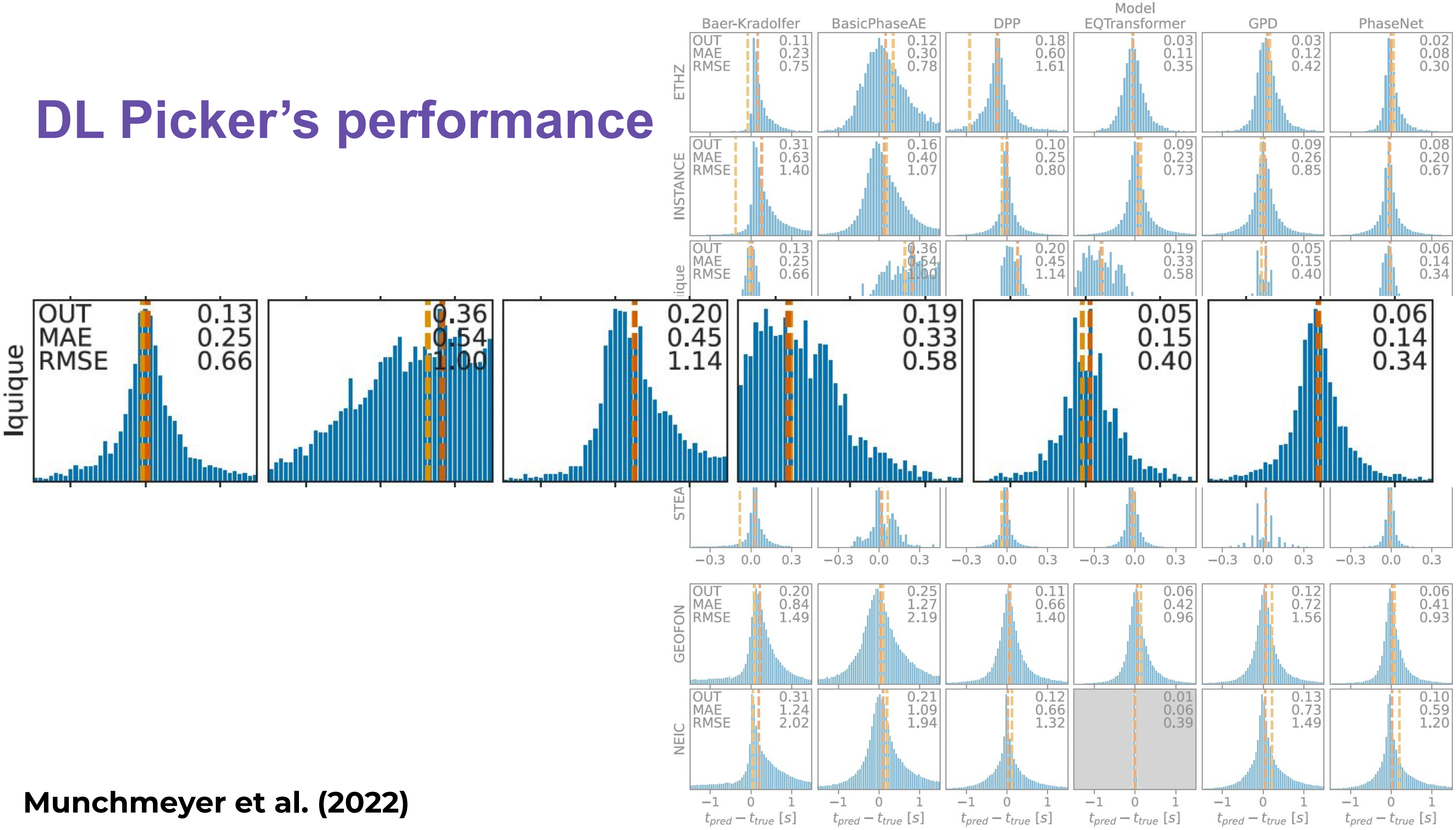
Armstrong et al. (2023)

DL Picker’s performance



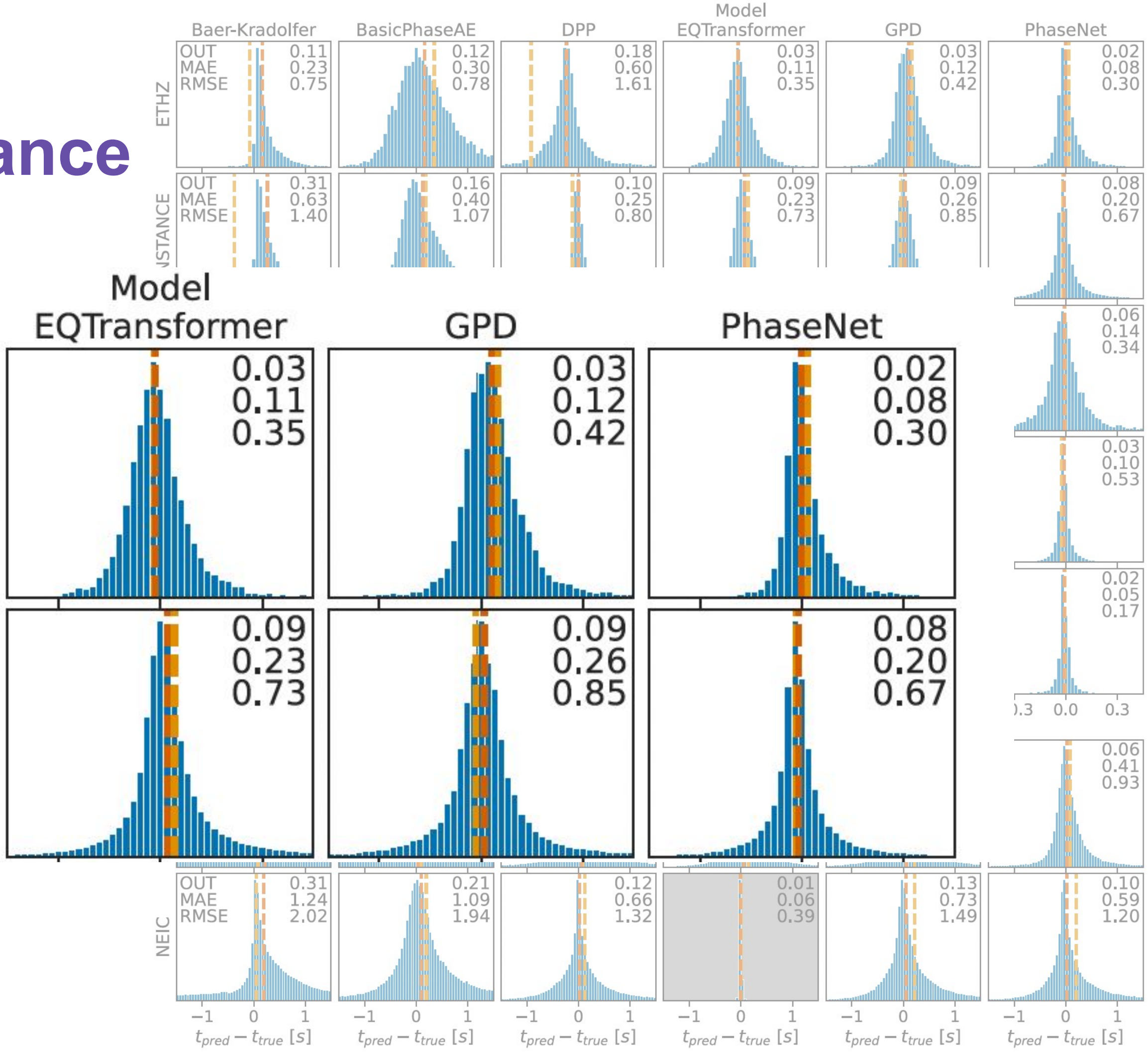
Munchmeyer et al. (2022)

DL Picker’s performance



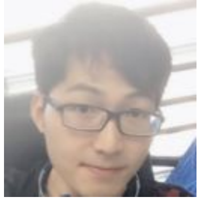
Munchmeyer et al. (2022)

DL Picker's performance



Munchmeyer et al. (2022)

Ensembling over datasets and model architecture: ELEP (Congong Yuan & Yiyu Ni)



YUAN et al.: BETTER TOGETHER: ENSEMBLE LEARNING FOR EARTHQUAKE DETECTION AND PHASE PICKING 5920217

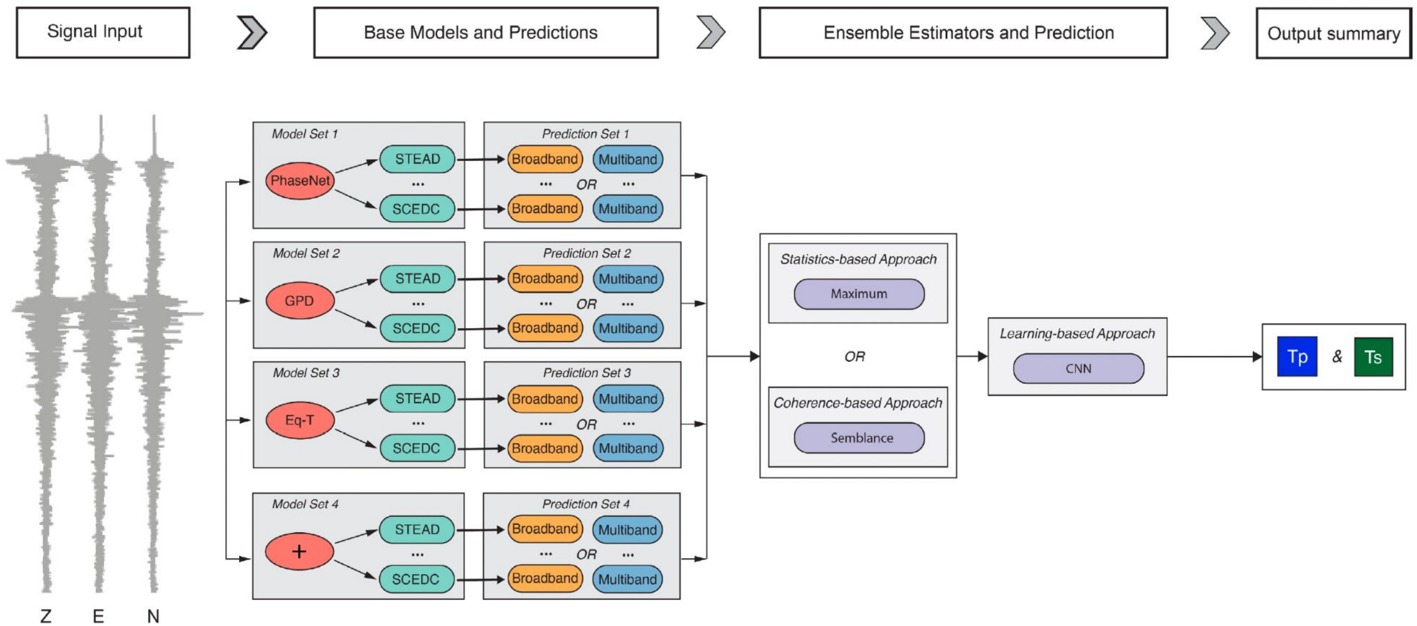
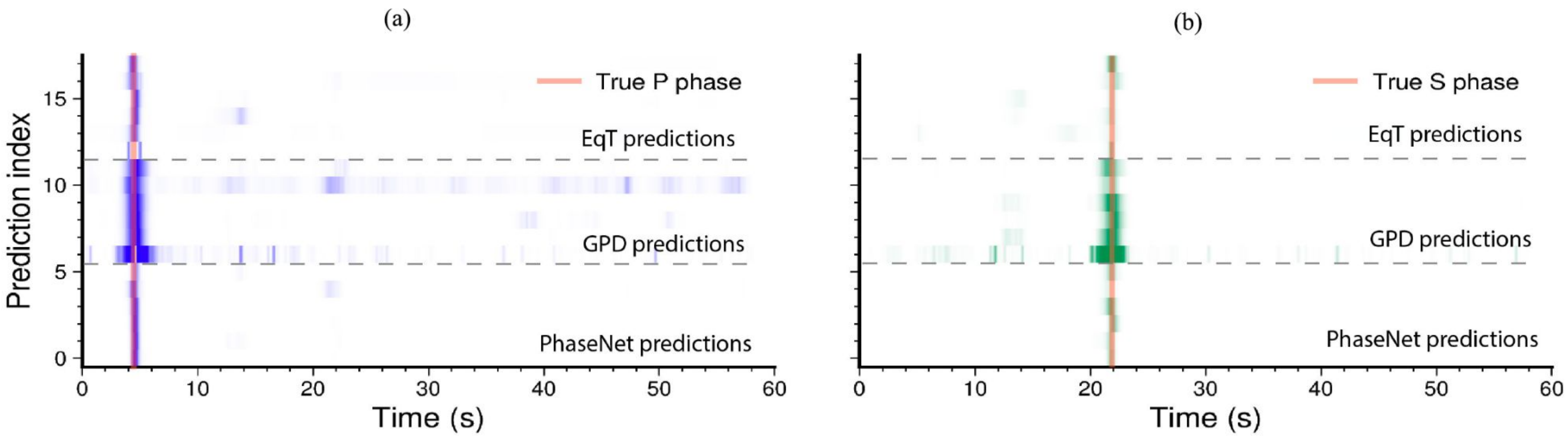
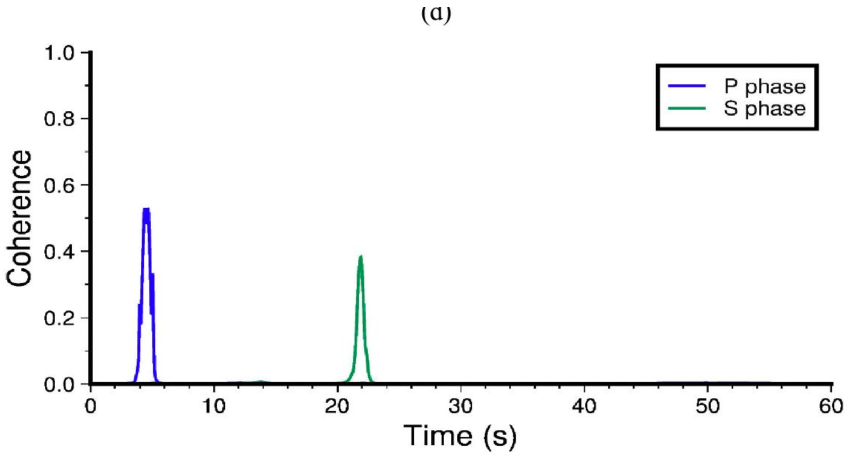


Fig. 1. Ensemble estimation-based framework for earthquake detection and phase picking. The main components include base predictions at BB or multiple-frequency bands (e.g., filtered data) and ensemble estimation by either statistics-, coherence-, or a learning-based approach. Note that only EqT-based pretrained models are tested.

Yuan et al. (2023)



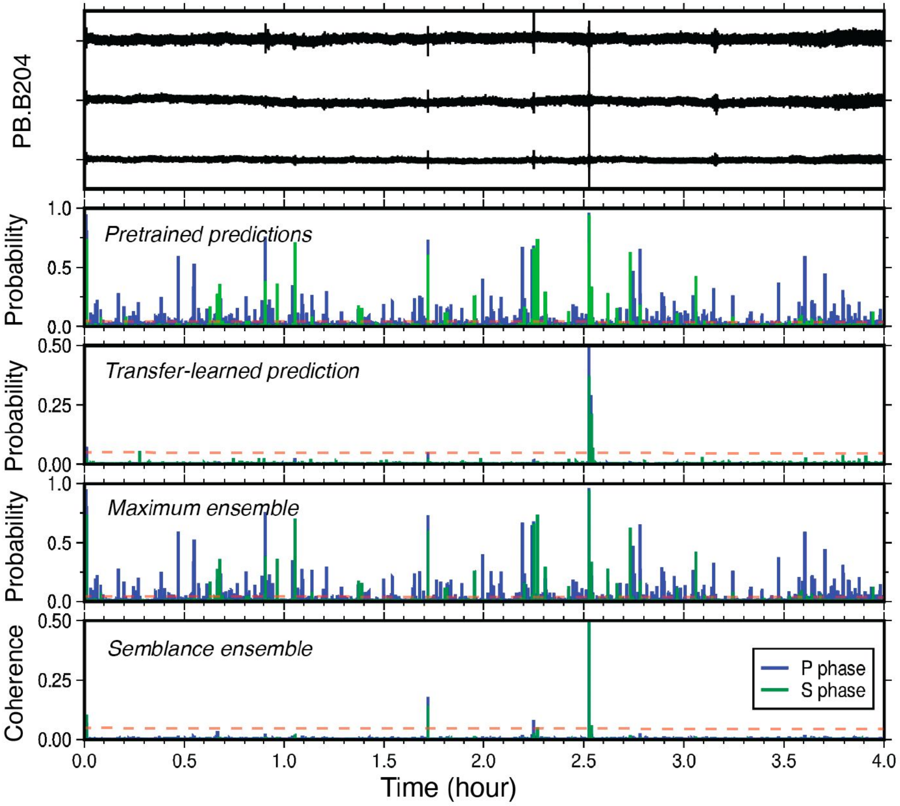


While individual models may be uncertain, the semblance/coherence of their prediction is robust.

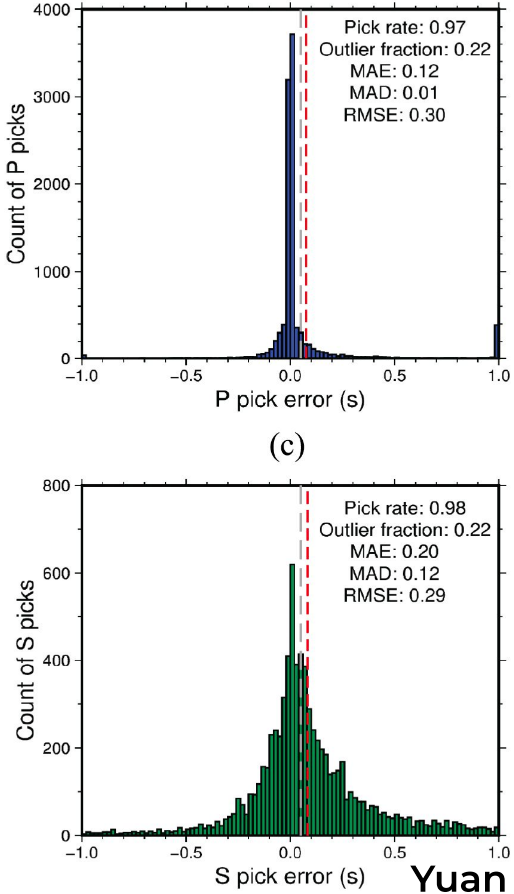
Yuan et al. (2023)

Reduced false positives. As accurate as transfer learned models. No training required.

Mt St Helens



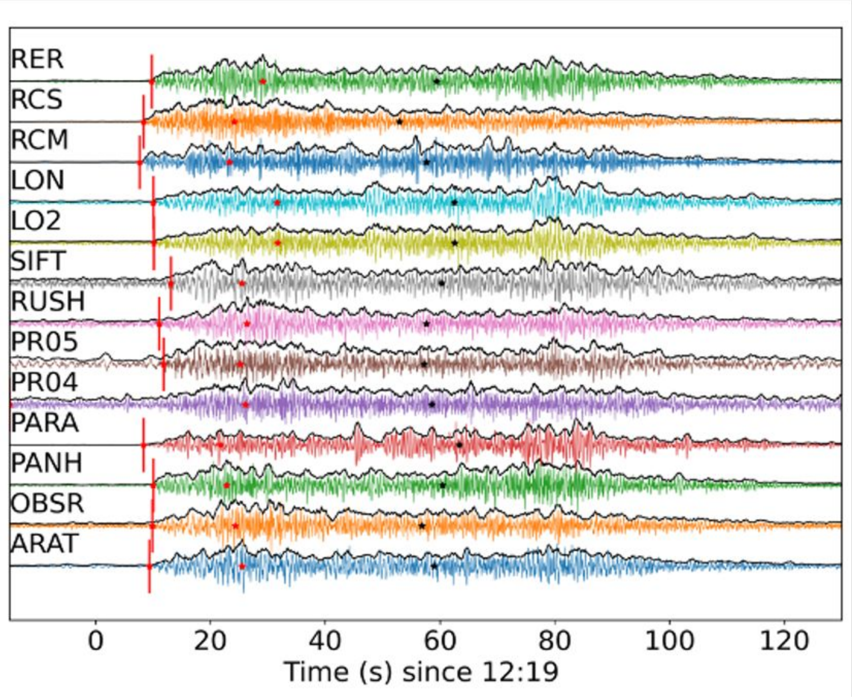
OBS benchmark dataset

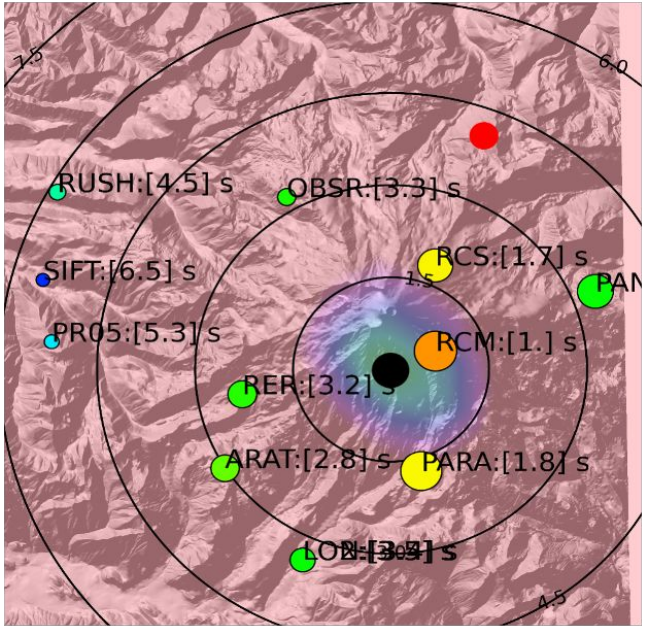


Yuan et al. (2023

Ensemble deep learning generalizes to detected other types of seismic events

Avalanches at Mt Rainier

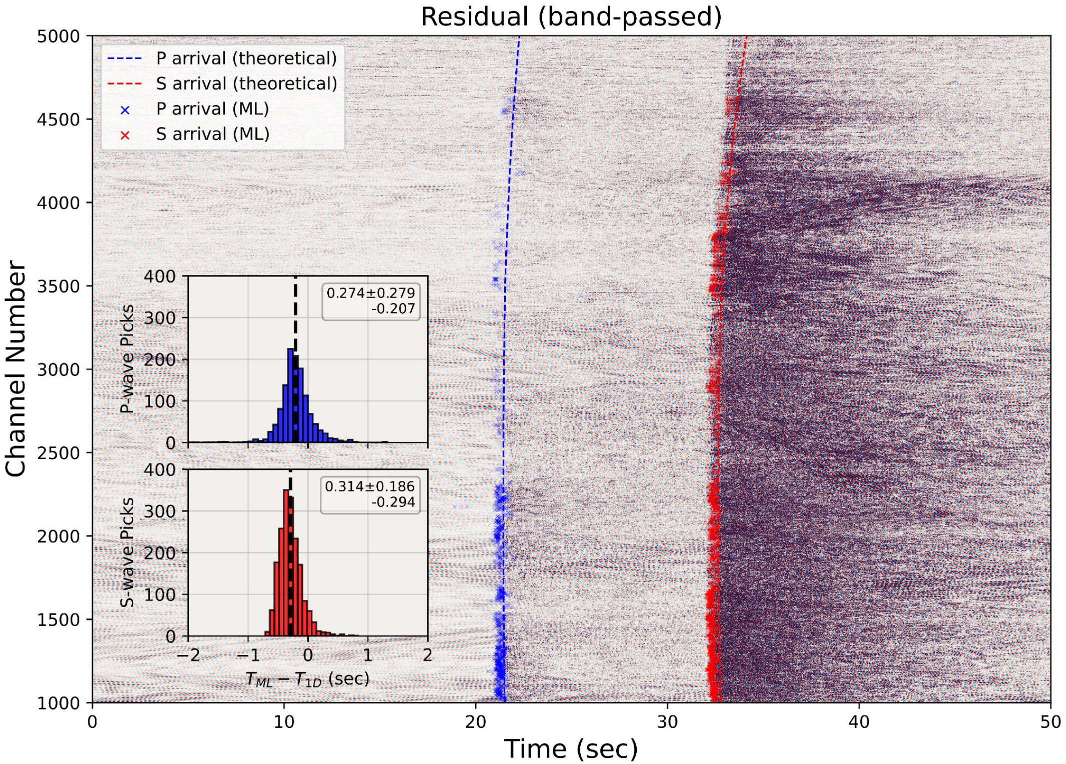




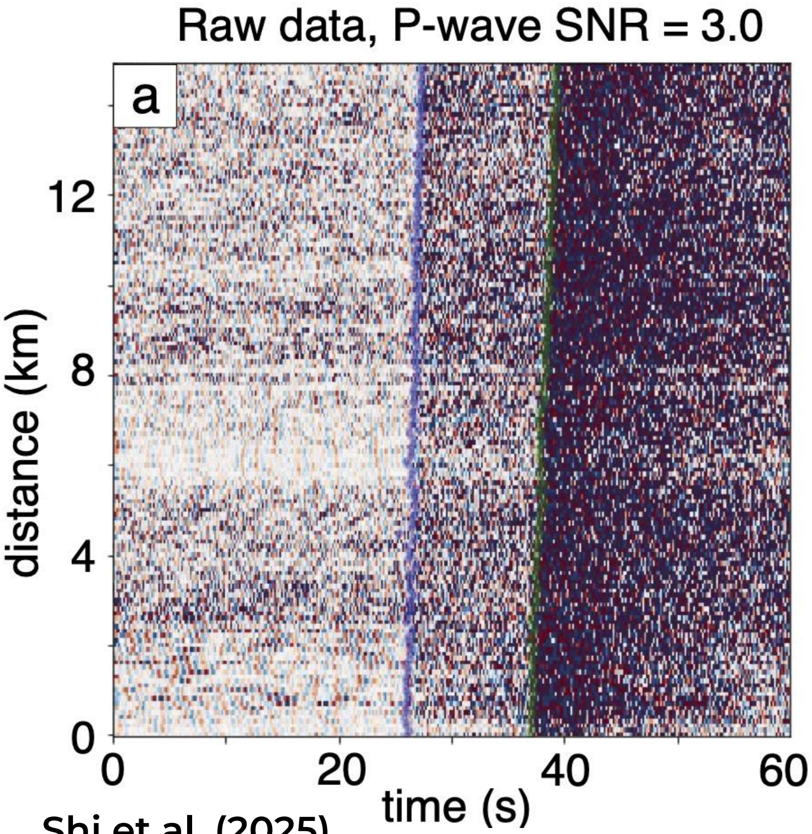
Denolle et al. (in prep)

Ensemble deep learning generalizes to detected other types of seismic events

Detecting Earthquakes on DAS



Ni et al. (2024)



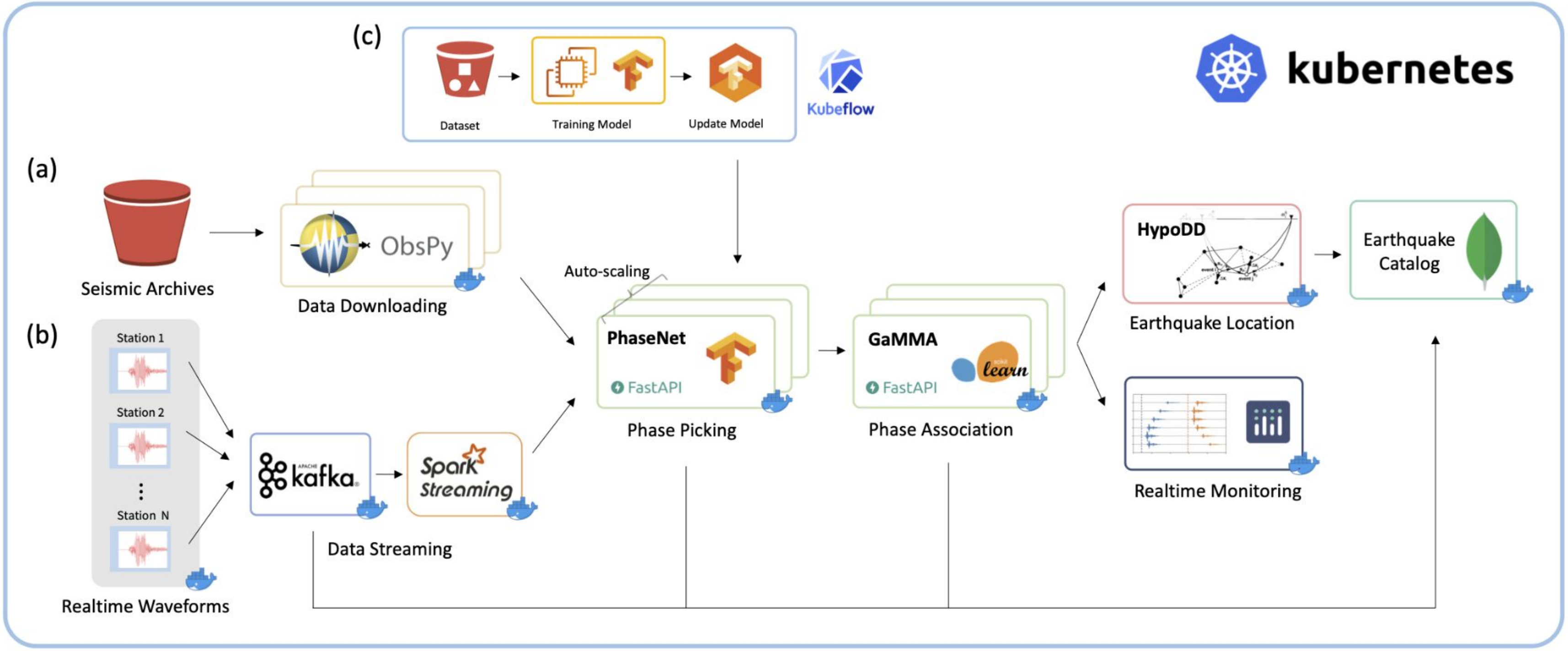
Shi et al. (2025)

Practicum: Train a Picker From Scratch

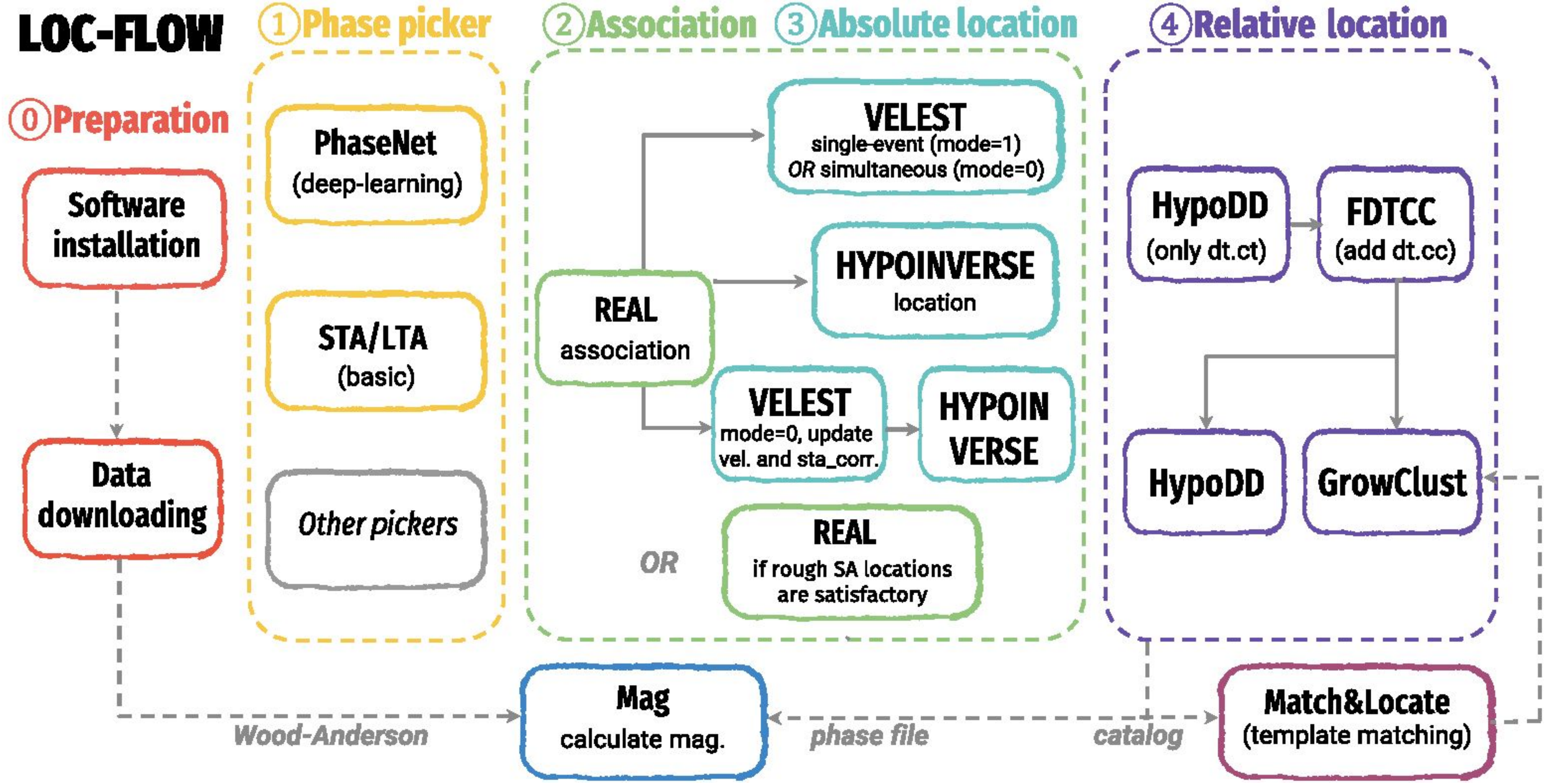
Practicum: Use Seisbench Ecosystem

Deploying Phase Pickers on Continuous Data

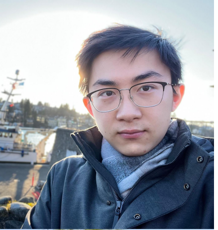
Cloud-based earthquake workflows



Zhu et al. (2023)



Zhang et al, (2022)



REPORT
doi:10.26443/seismica.v2i2.979

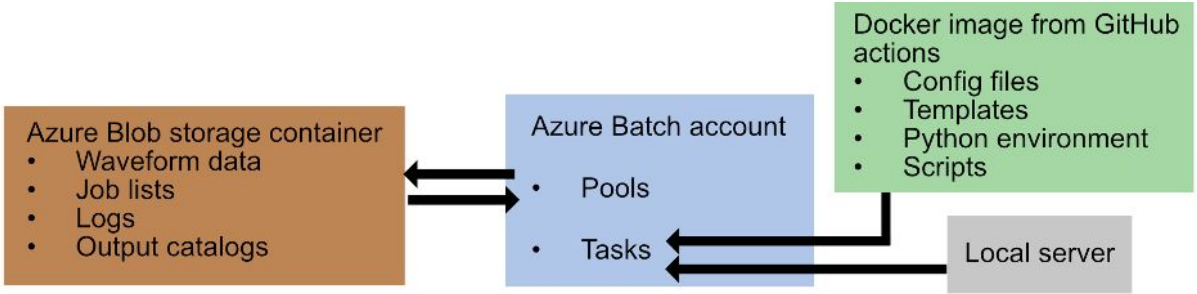
SEISMICA

https://github.com/Denolle-Lab/seismicloud/

Seismology in the cloud: guidance for the individual researcher

Z. Krauss *[1], **Y. Ni** [2], **S. Henderson** [2,3], **M. Denolle** [2]

[1] School of Oceanography, University of Washington, Seattle, WA, USA, [2] Department of Earth and Space Sciences, University of Washington, Seattle, WA, USA, [3] eScience Institute, University of Washington, Seattle, WA, USA



storage

compute

software

Krauss et al (2023)

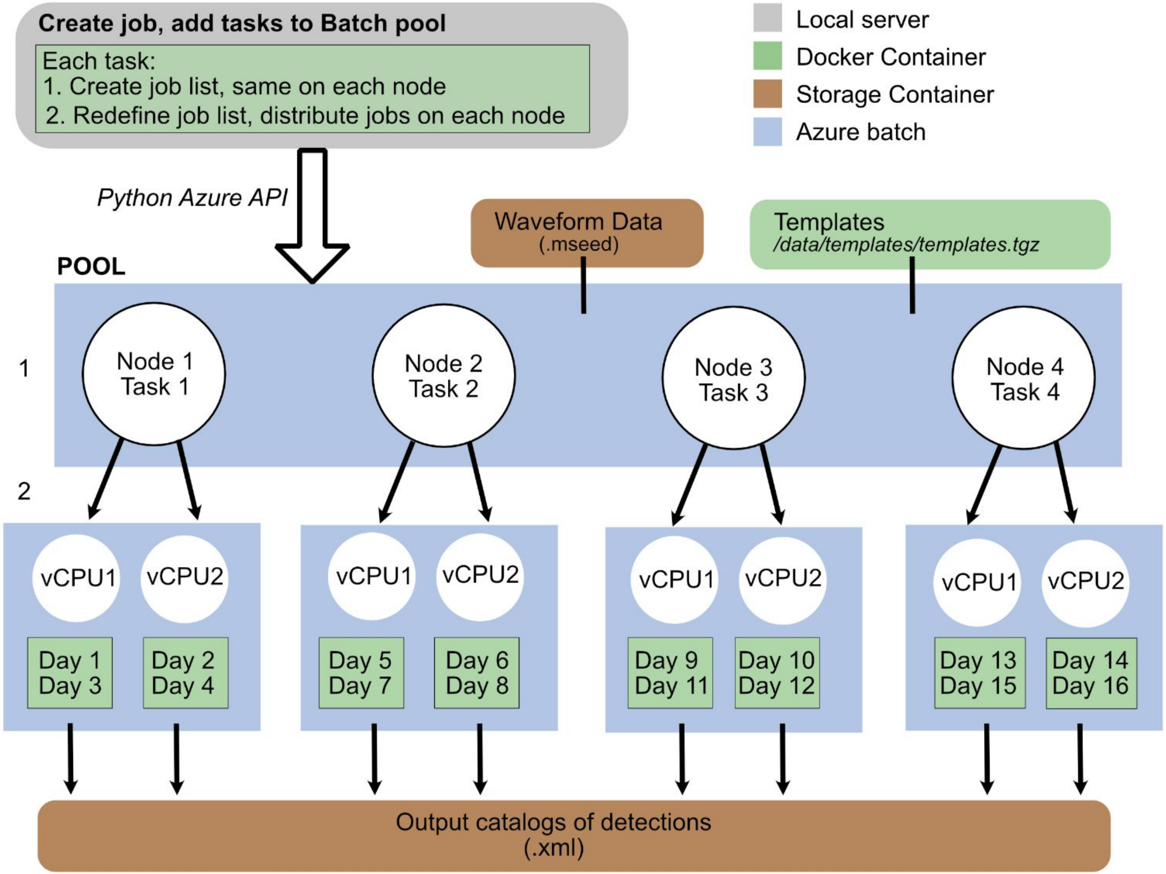


Figure 3 Example flowchart of how Azure cloud Pool-based Parallelization works for the Template Matching workflow, following the color scheme of Figure 2. The EQTransformer workflow is identical, except that the data paths are specific to both the day of the year and the seismic station.

Krauss et al (2023)

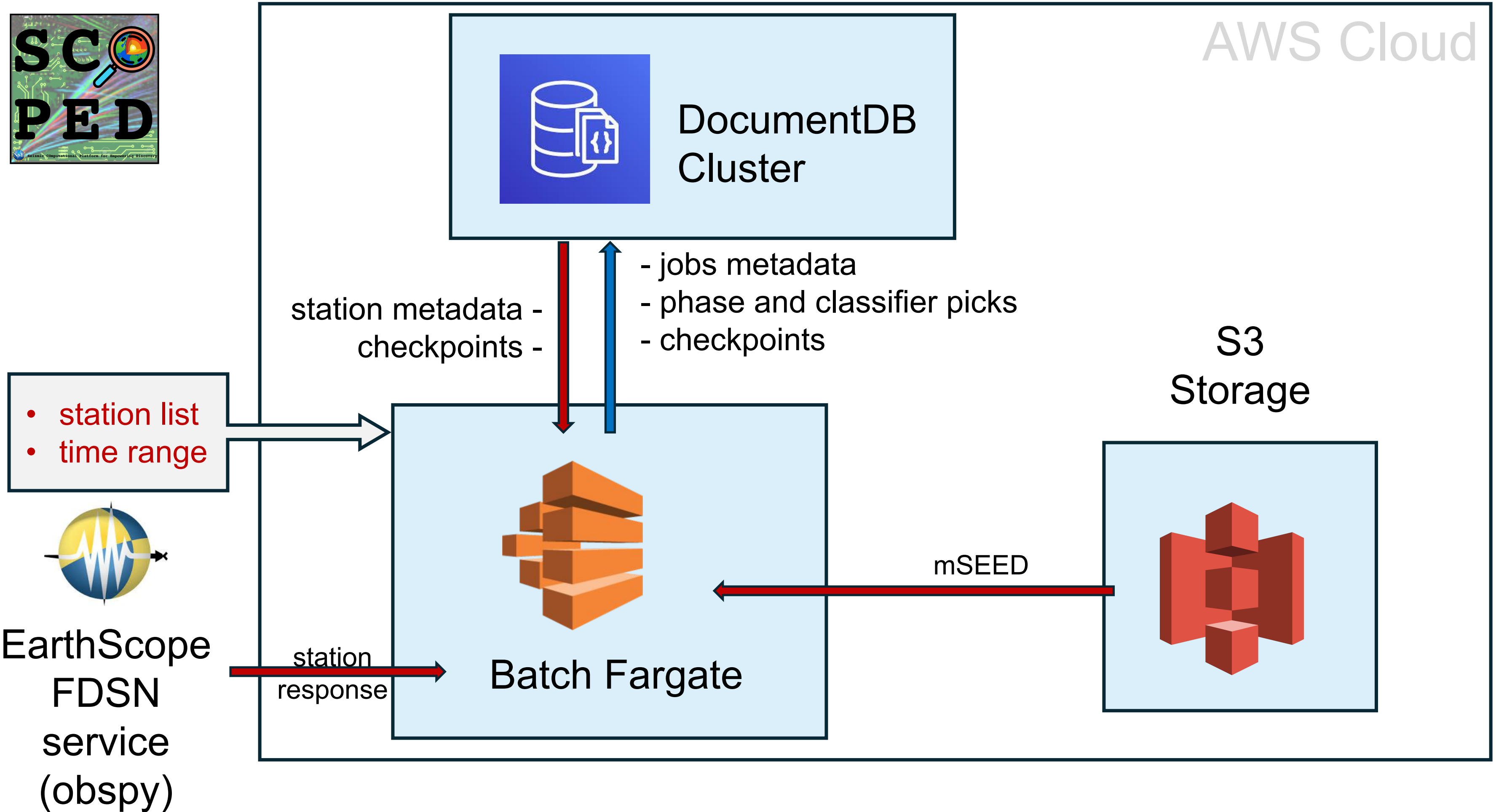
First <u>Community-Concerted</u> *AI-aided Earthquake Catalog*.

1. Detection, Discrimination (DL)
2. Phase picking (DL)
3. Association (DL for large scale)
4. Absolute location (conventional but 1D-3D vel models)
5. Relative location (xcorr based, or DL)
6. Magnitudes
7. Moment Tensors

QC is a challenge!

=> many modules, many choices
=> promoting v***ersion-controlled***
research-grade earthquake catalogs

Jannes Munchmeyer, Yiyu Ni, Ian Wang, Marine Denolle



QuakeScope Workflow

A. Storage

AWS Cloud

EarthScope Consortium (1PB)

- Day-long mSEED in us-east-2.
- Credential required due to restricted data
- Prefix format:
  S3_ACCESS_POINT/miniseed/NET/YEAR/DOY/
- Key format: STA.NET.YEAR.DOY#V

SCEDC (150TB)

- Day-long mSEED in us-west-2. Single channel.
- Prefix format:
  s3://scedc-pds/continuous_waveforms/YEAR/YEAR_DOY/
- Key format: NETSTA_CHA_LOC_YEARDOY.ms

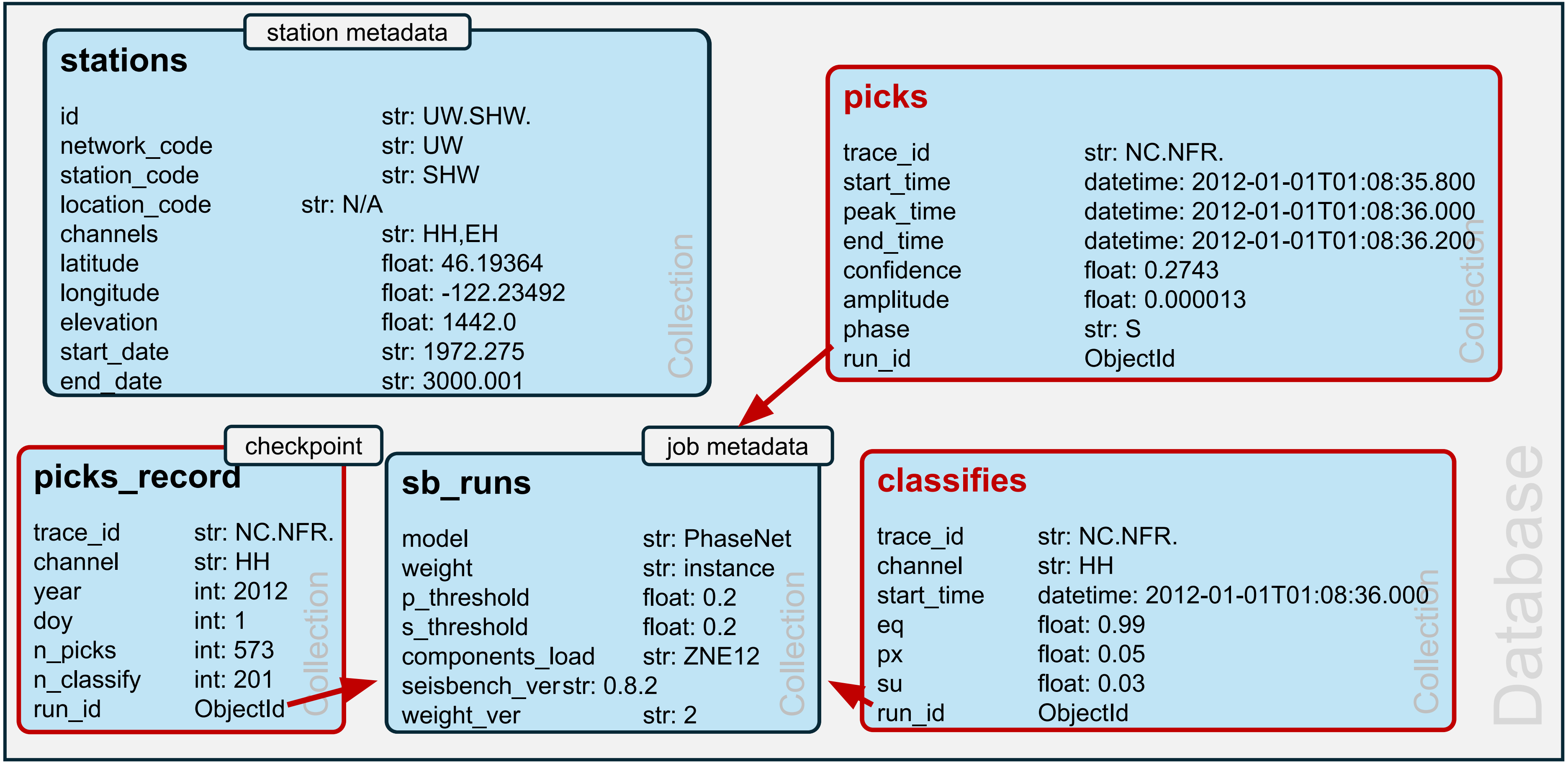
NCEDC (150TB)

- Day-long mSEED in us-east-2. Single channel.
- Prefix format:
  s3://ncedc-pds/continuous_waveforms/NET/YEAR/YEAR.DOY/
- Key format: STA.NET.CHA.LOC.D.YEAR.DOY

S3 Storage

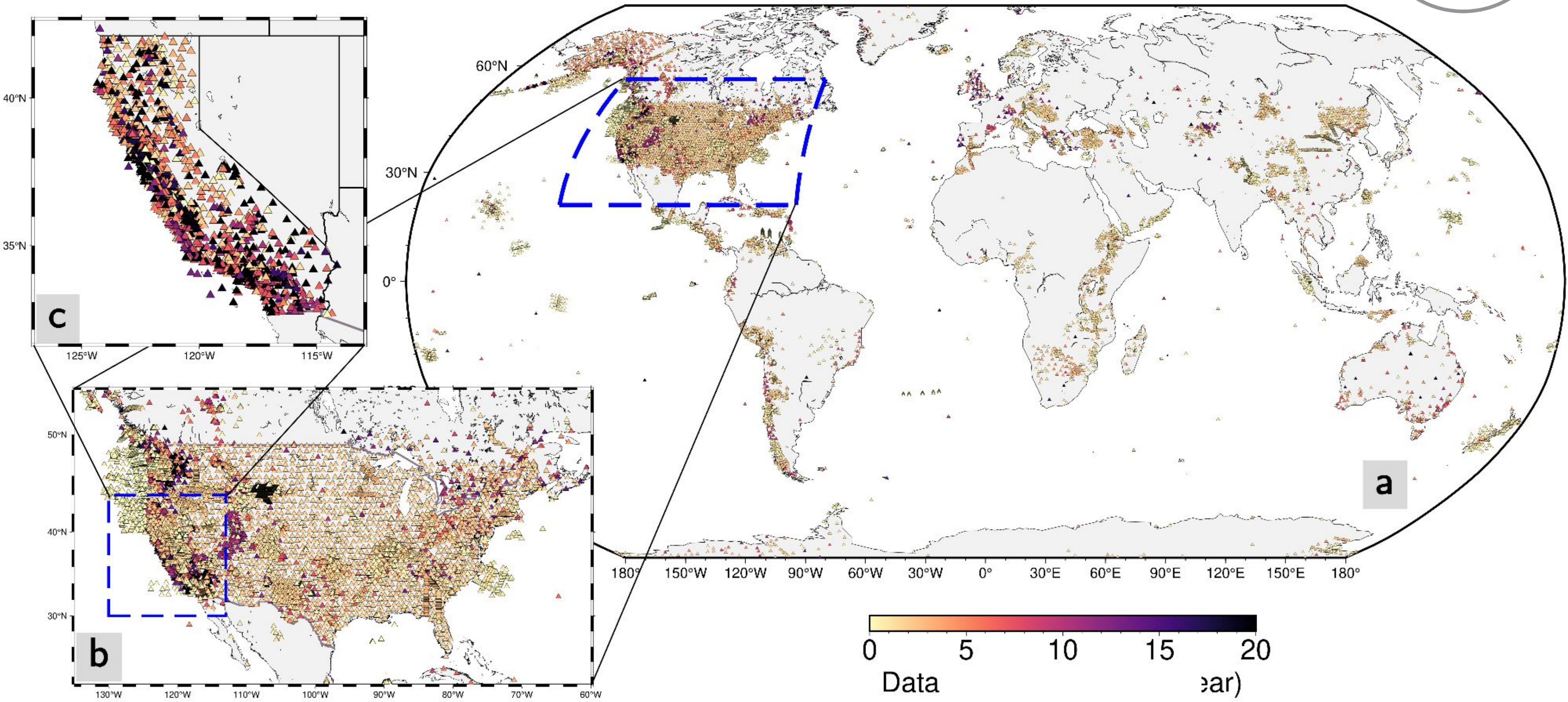
QuakeScope Workflow

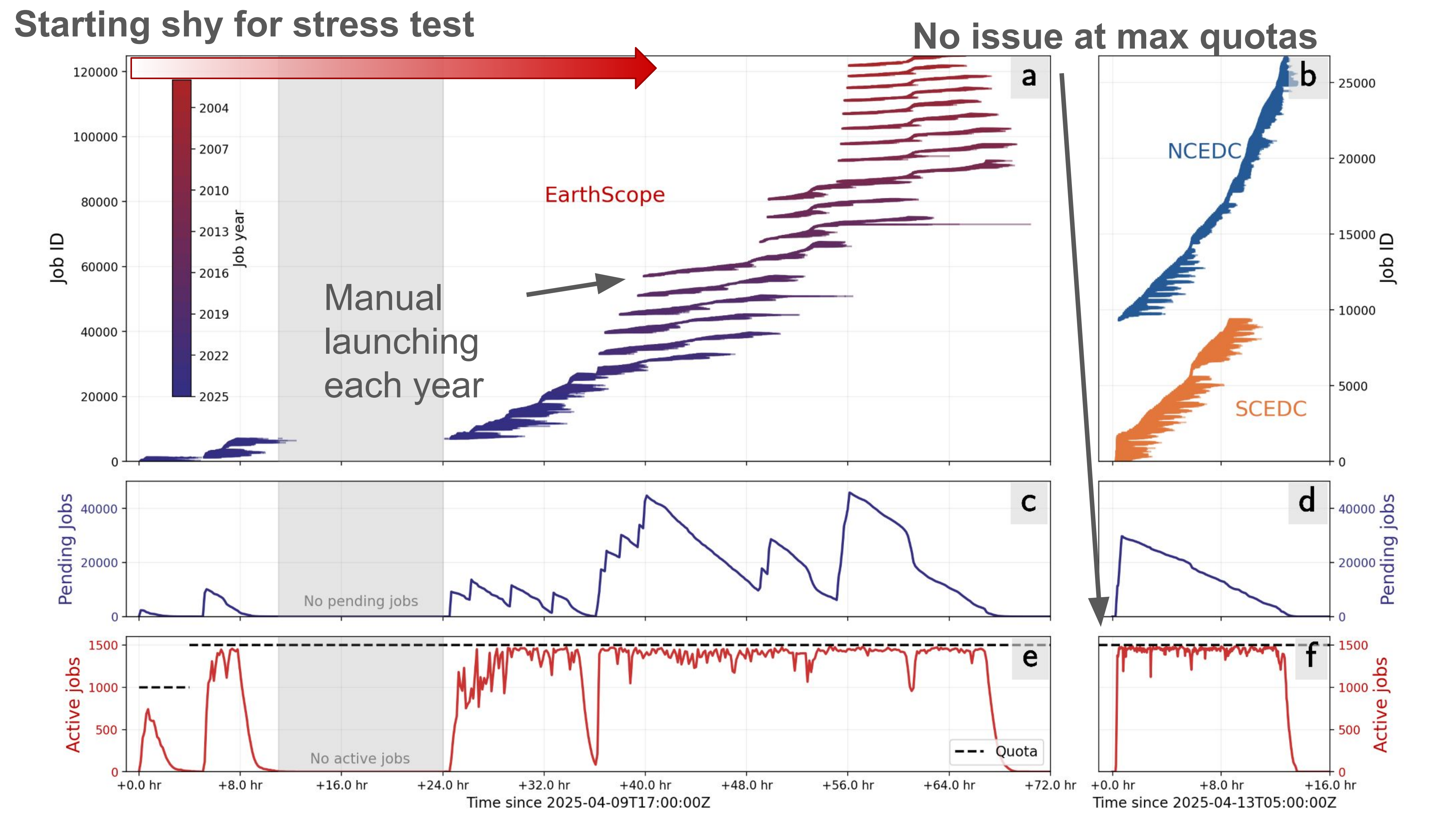
B. DocumentDB



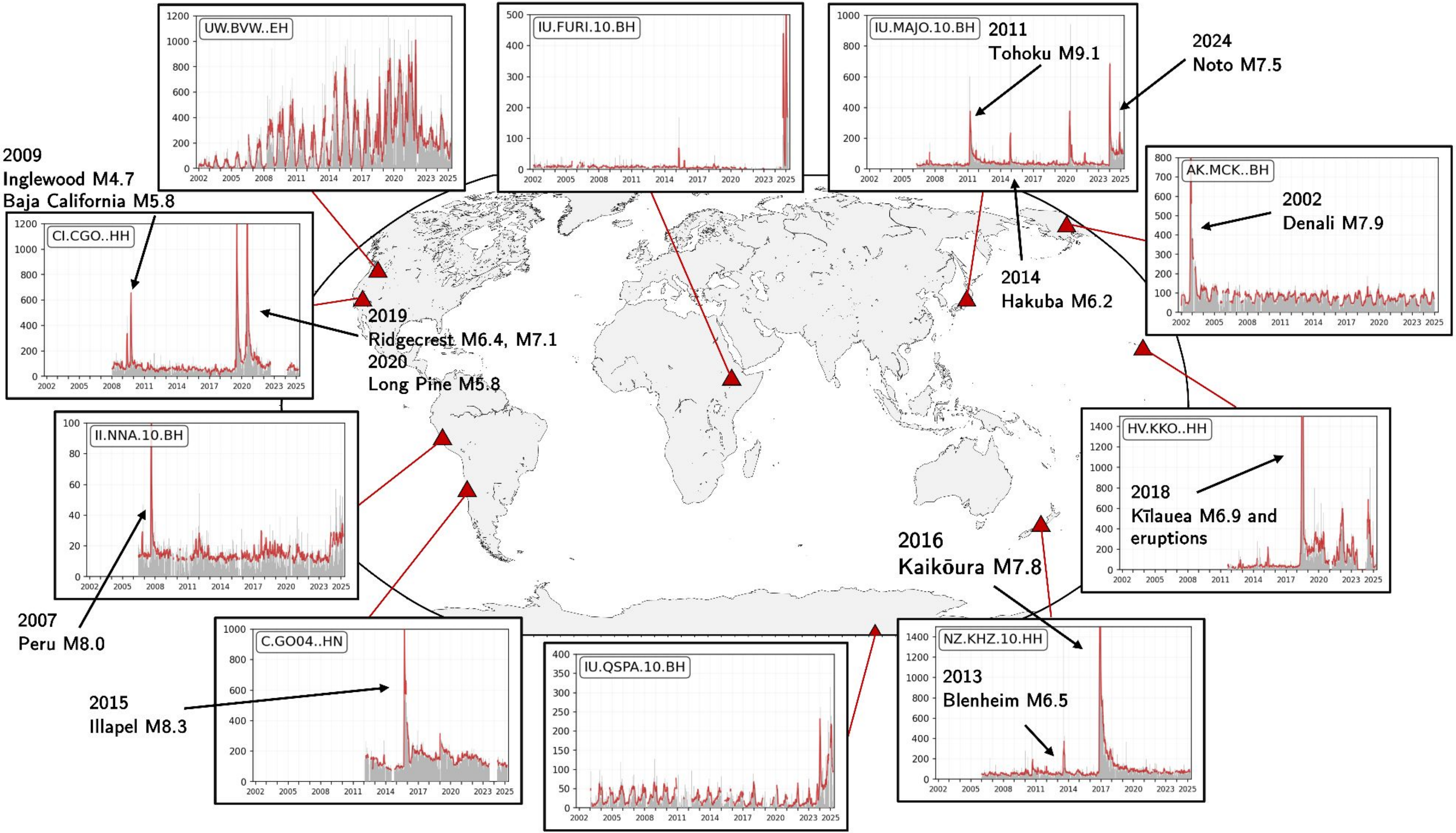
Quite. some. data.

NCEDC+SCEDC

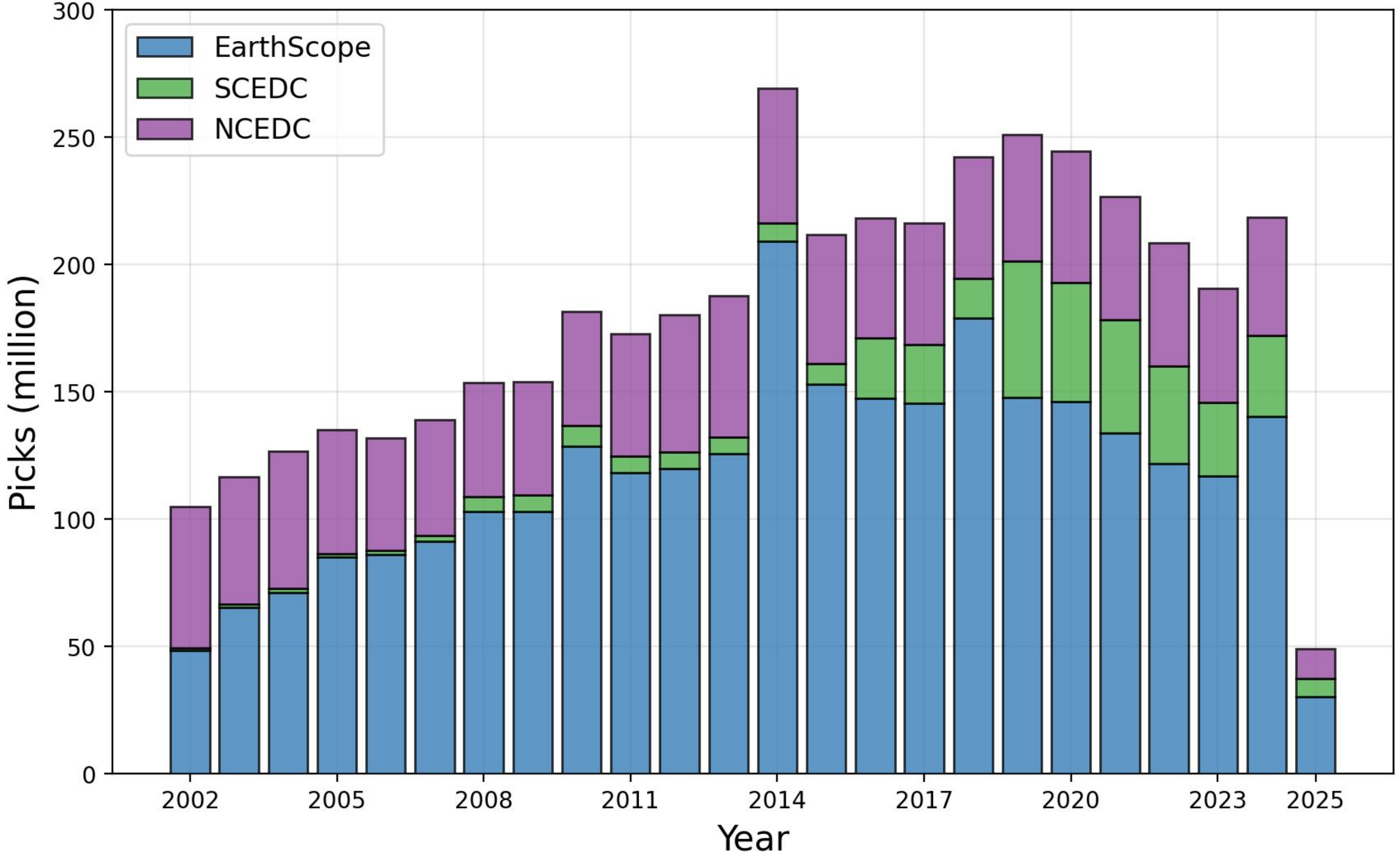




Starting shy for stress test
No issue at max quotas
a
b
c
d
e
f
Job ID
Job year
EarthScope
Manual launching each year
NCEDC
SCEDC
Pending Jobs
Pending jobs
No pending jobs
Active jobs
No active jobs
Quota
Time since 2025-04-09T17:00:00Z
Time since 2025-04-13T05:00:00Z


UW.BVW..EH
IU.FURI.10.BH
IU.MAJO.10.BH
2011
Tohoku M9.1
2024
Noto M7.5
2014
Hakuba M6.2
AK.MCK..BH
2002
Denali M7.9
2009
Inglewood M4.7
Baja California M5.8
CI.CGO..HH
2019
Ridgecrest M6.4, M7.1
2020
Long Pine M5.8
II.NNA.10.BH
2007
Peru M8.0
HV.KKO..HH
2018
Kīlauea M6.9 and
eruptions
2016
Kaikōura M7.8
C.GO04..HN
2015
Illapel M8.3
IU.QSPA.10.BH
NZ.KHZ.10.HH
2013
Blenheim M6.5

PhaseNet Picks Results



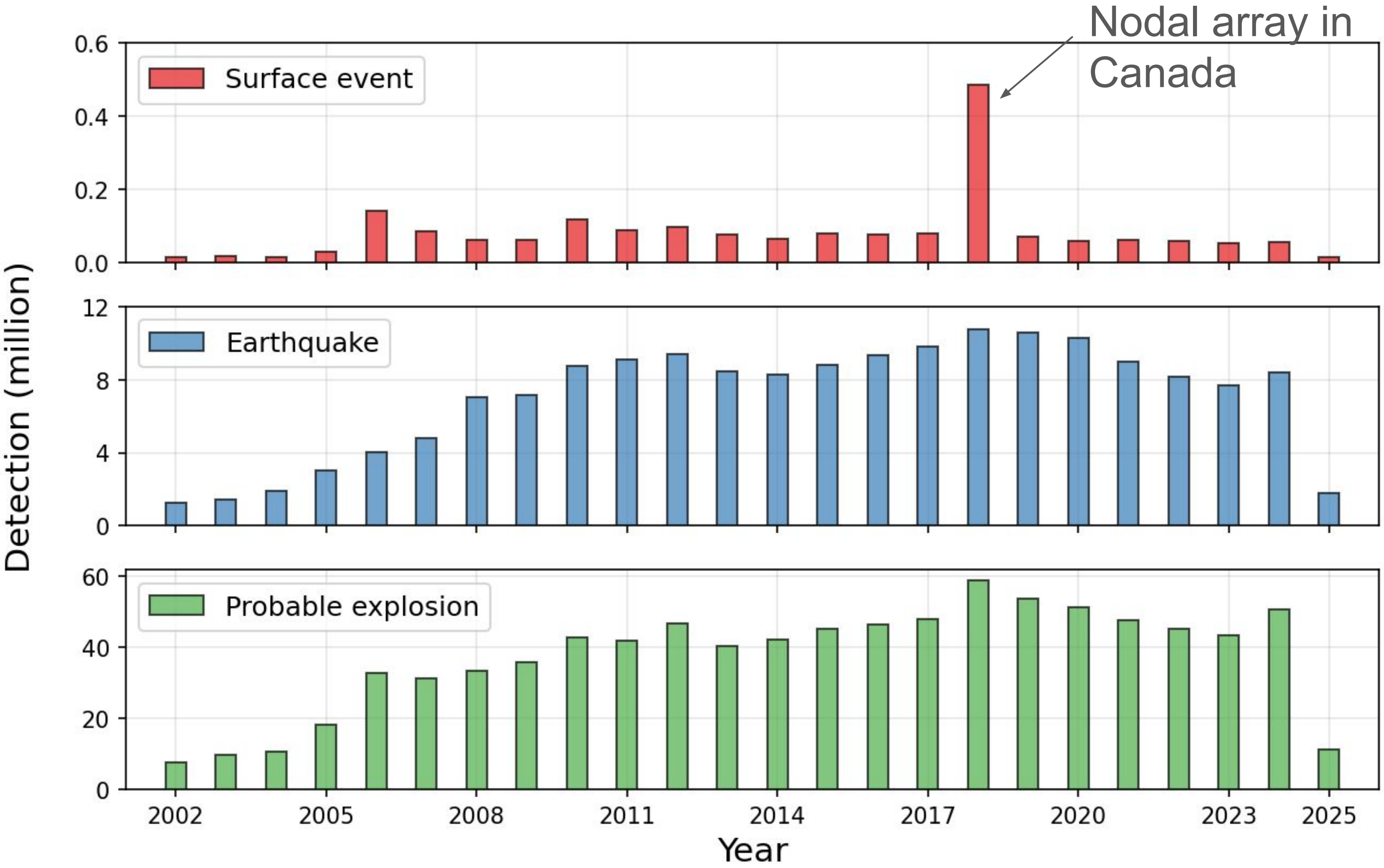
EarthScope**: 2.8 billions**

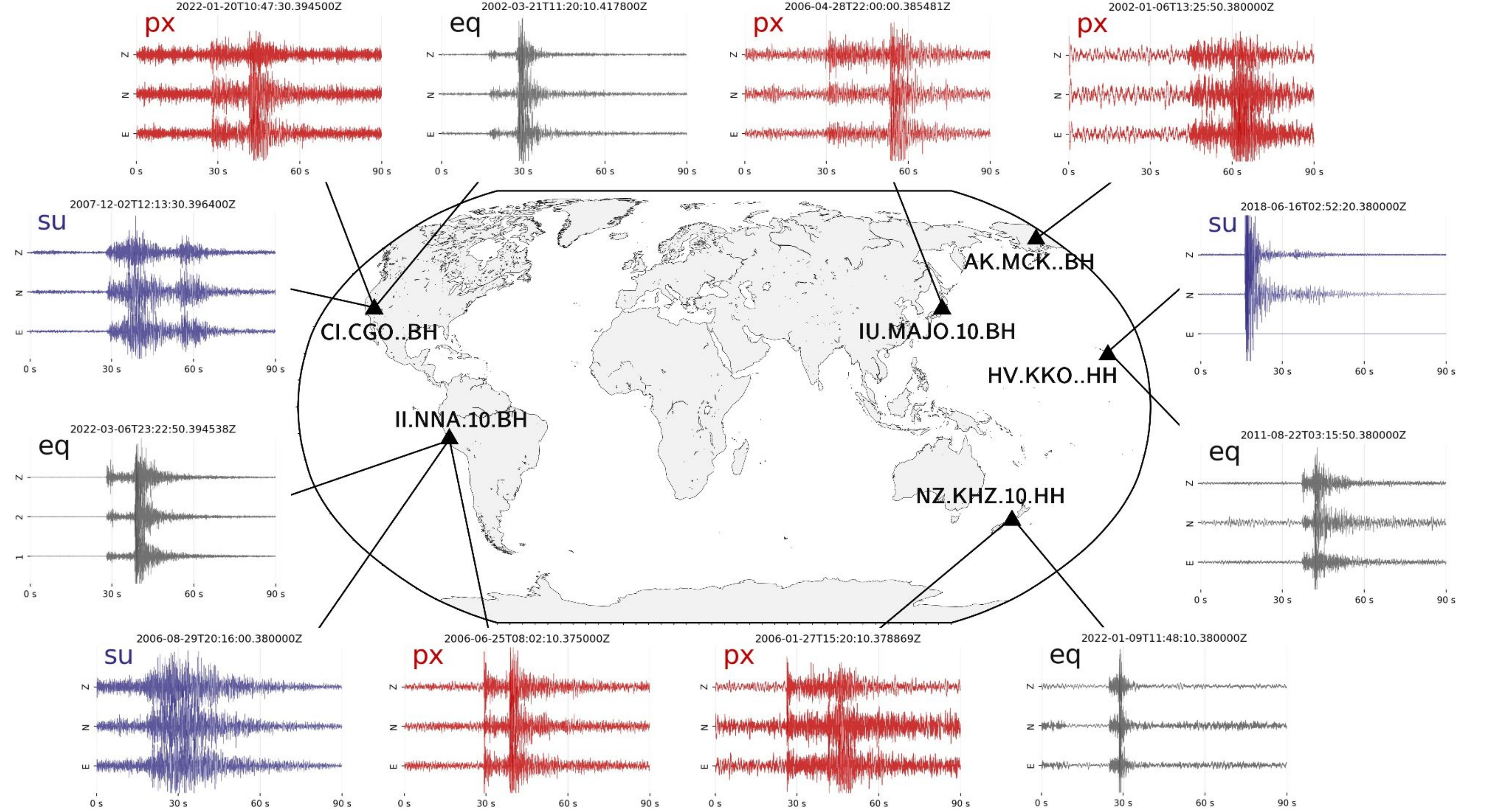
NCEDC**: 1.1 billion**

SCEDC**: 0.4 billion**

Total**: 4.3 billions**

Discrimination Results

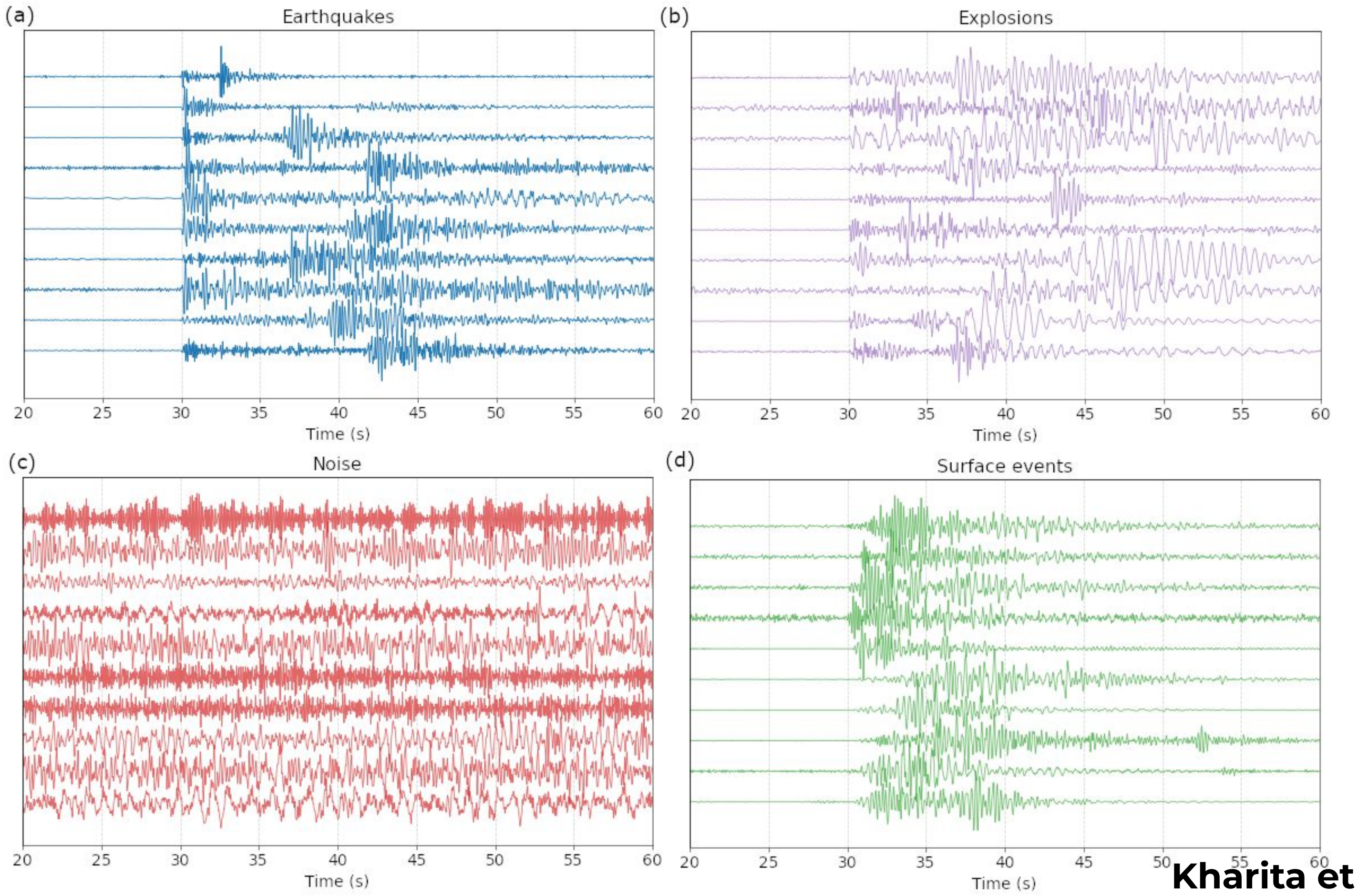



2022-01-20T10:47:30.394500Z
px
2002-03-21T11:20:10.417800Z
eq
2006-04-28T22:00:00.385481Z
px
2002-01-06T13:25:50.380000Z
px
2007-12-02T12:13:30.396400Z
su
2018-06-16T02:52:20.380000Z
su
AK.MCK..BH
CI.CGO..BH
IU.MAJO.10.BH
HV.KKO..HH
II.NNA.10.BH
2022-03-06T23:22:50.394538Z
eq
2011-08-22T03:15:50.380000Z
eq
NZ.KHZ.10.HH
2006-08-29T20:16:00.380000Z
su
2006-06-25T08:02:10.375000Z
px
2006-01-27T15:20:10.378869Z
px
2022-01-09T11:48:10.380000Z
eq
Z
N
E
1
2
0 s
30 s
60 s
90 s

Event Discrimination for multi-geohazard monitoring

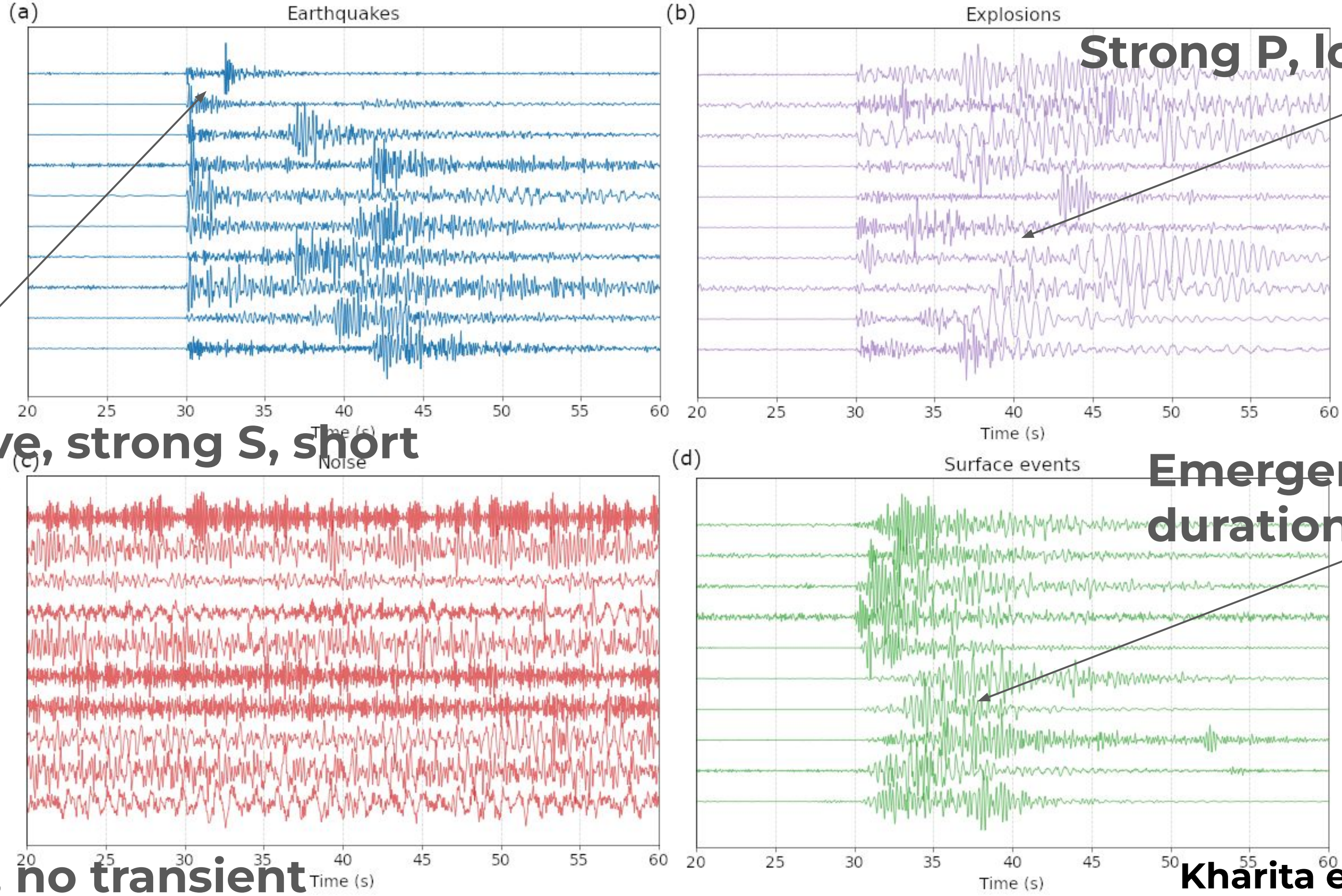
With Akash Kharita and Alex Hutko

The PNW curated data set has diverse source types, with mainly Earthquakes, Explosions, and Surface events



Kharita et al. (in prep)

The PNW curated data set has diverse source types, with mainly Earthquakes, Explosions, and Surface events



Impulsive, strong S, short window

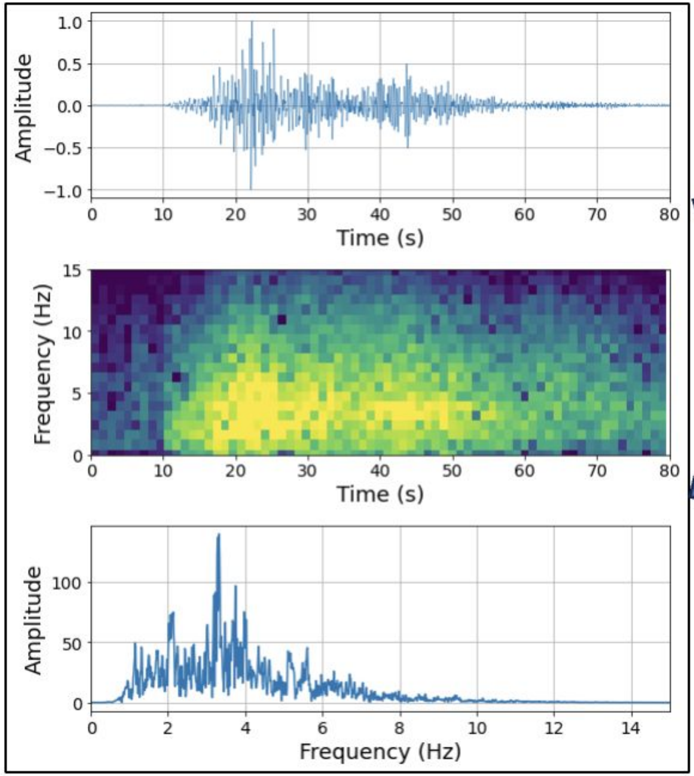
Strong P, long coda

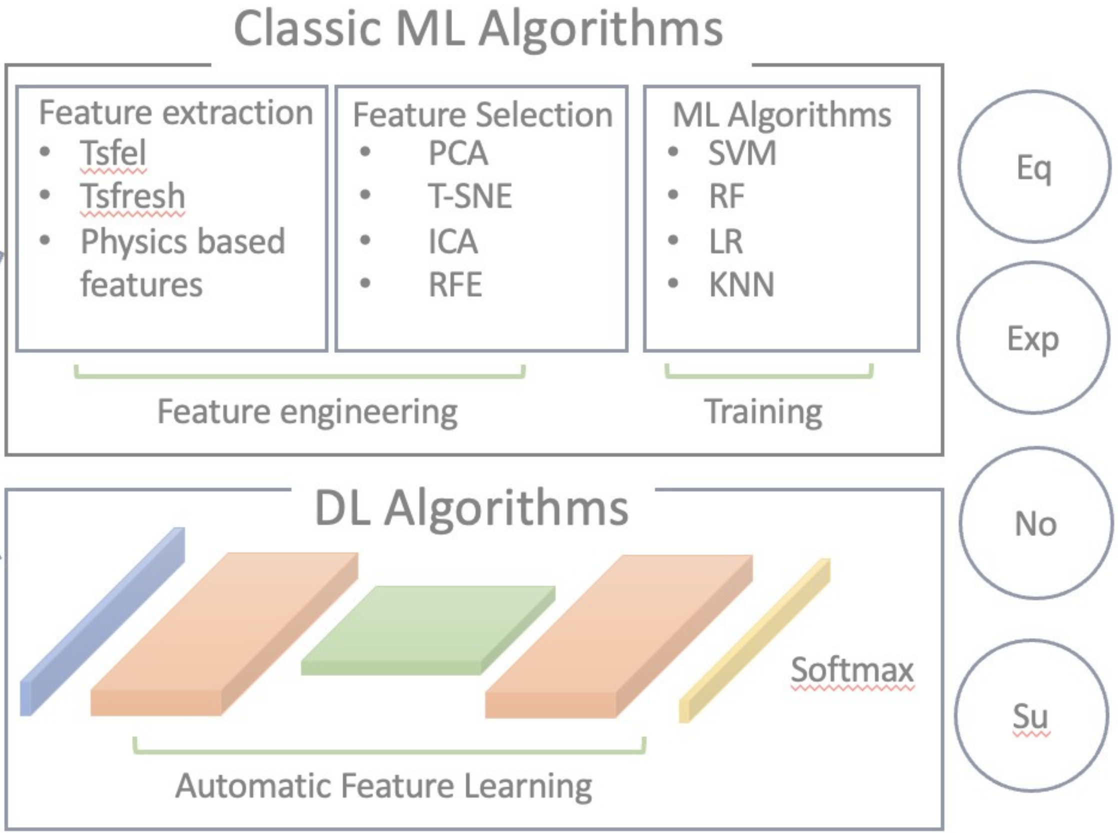
Emergent, long duration

Diffuse, no transient

Kharita et al. (in prep)

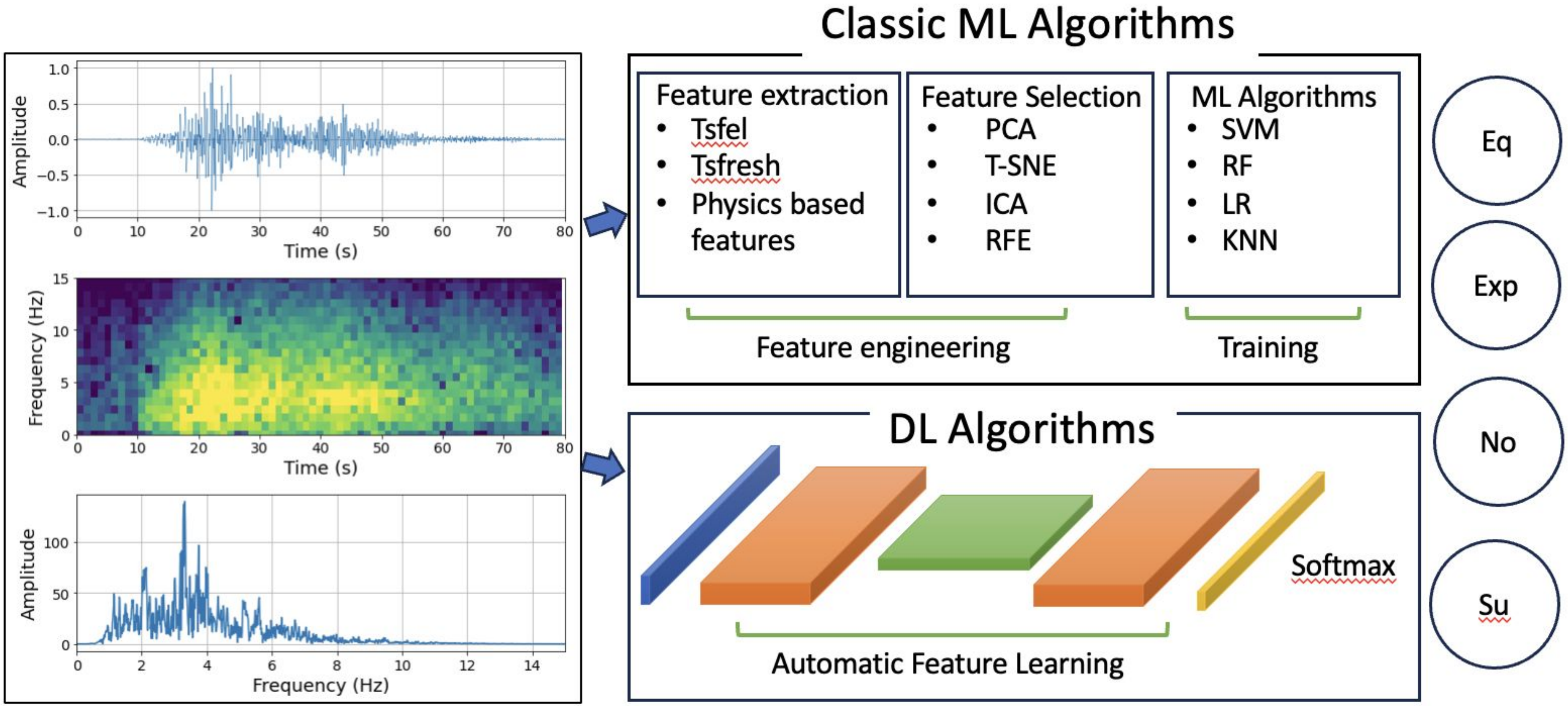
Event Discrimination using Machine Learning





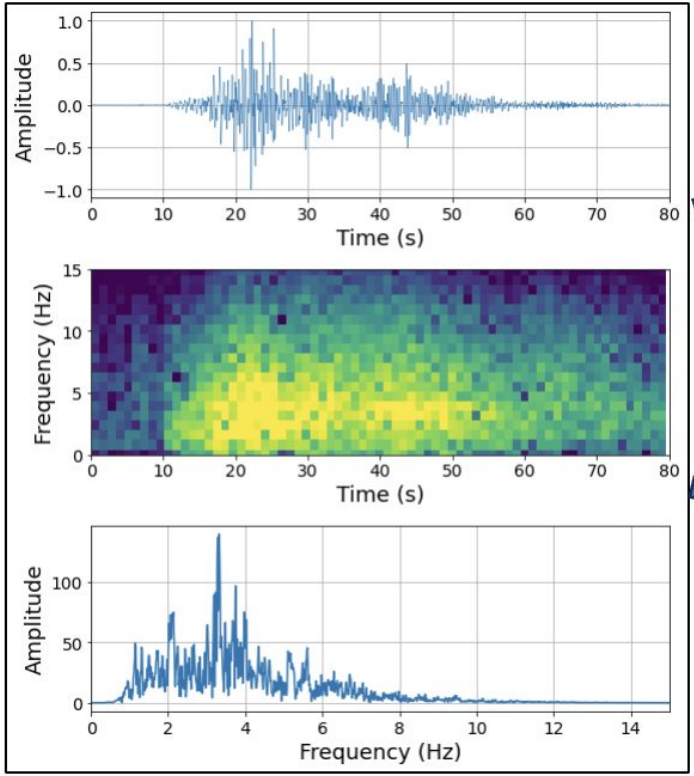
Kharita et al. (in prep)

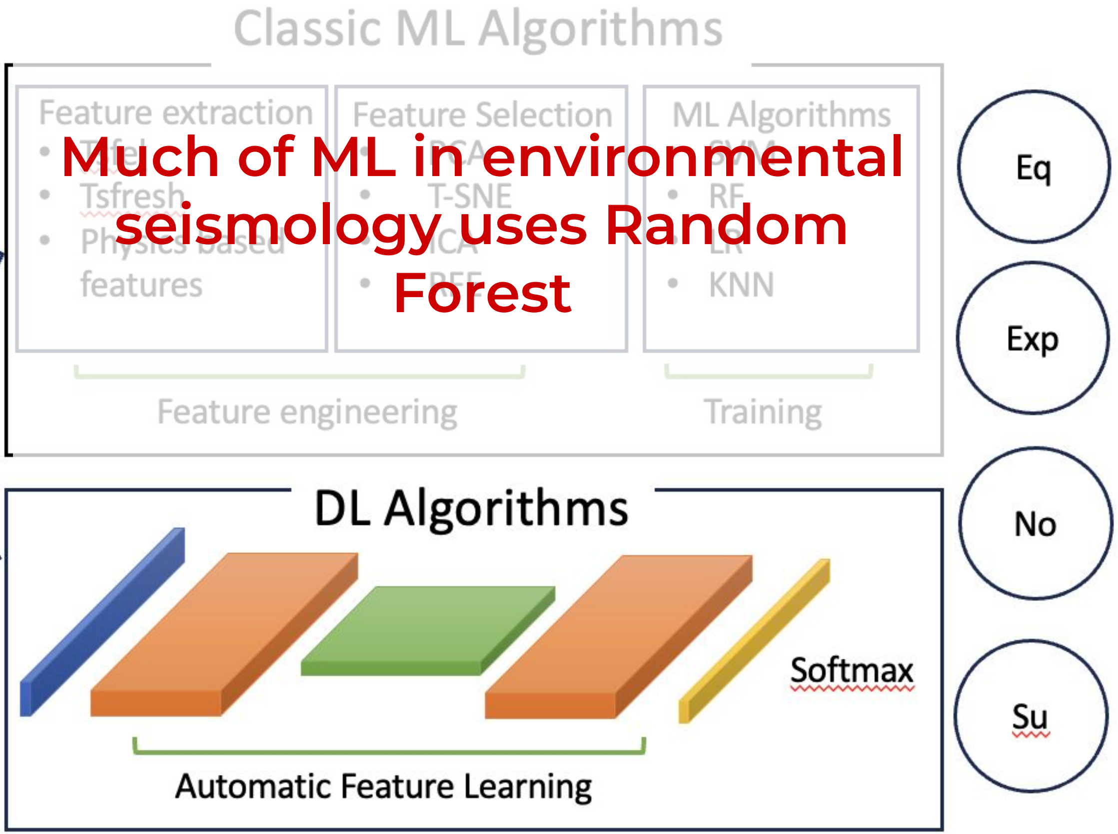
Event Discrimination using Machine Learning



Kharita et al. (in prep)

Event Discrimination using Machine Learning

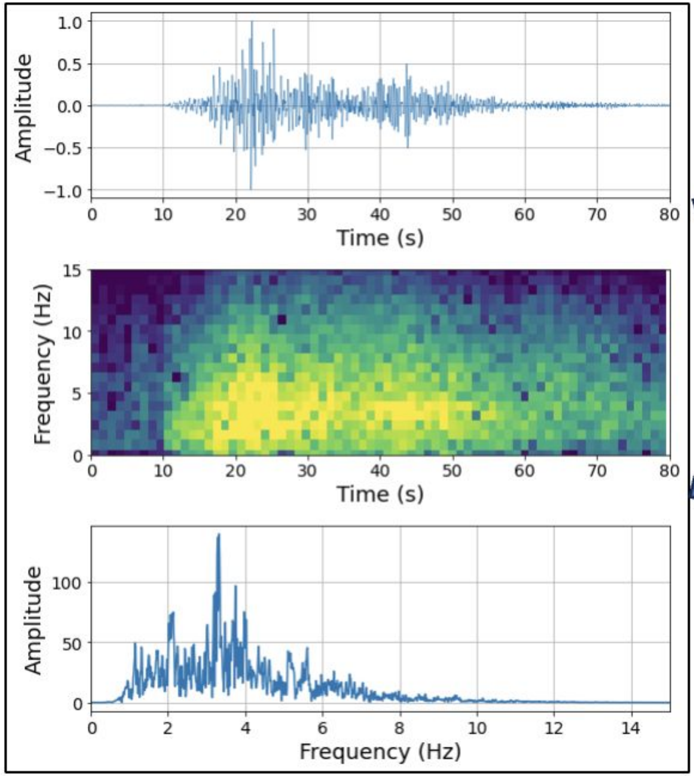


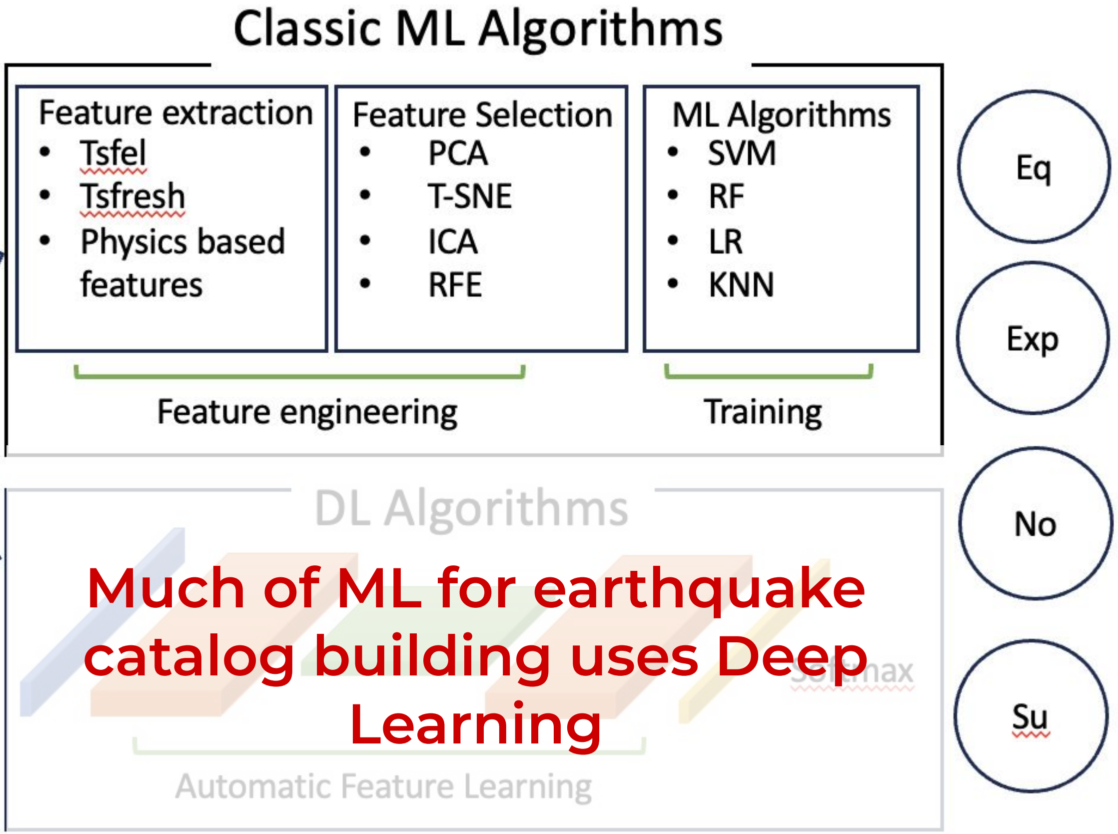


Much of ML in environmental seismology uses Random Forest

Kharita et al. (in prep)

Event Discrimination using Machine Learning

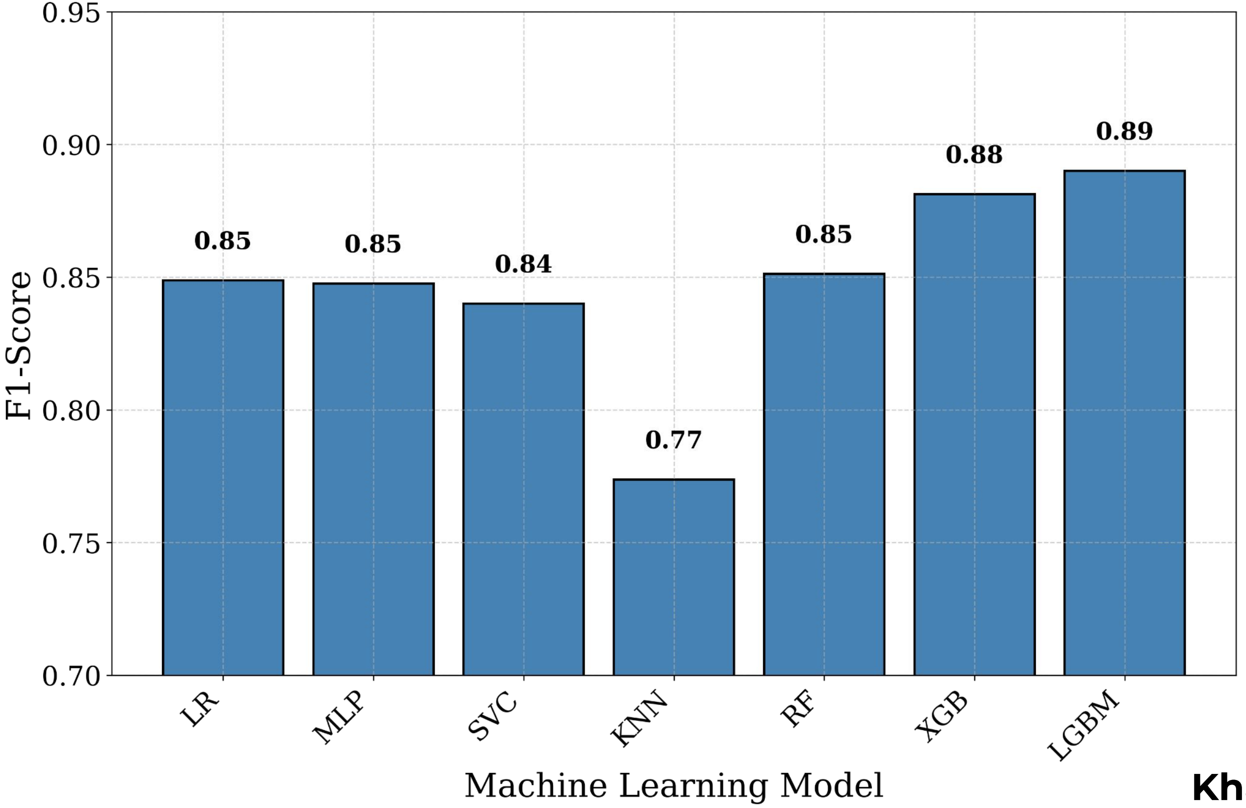




Much of ML for earthquake catalog building uses Deep Learning

Kharita et al. (in prep)

Classic Machine Learning: Decision-Trees-based models always win



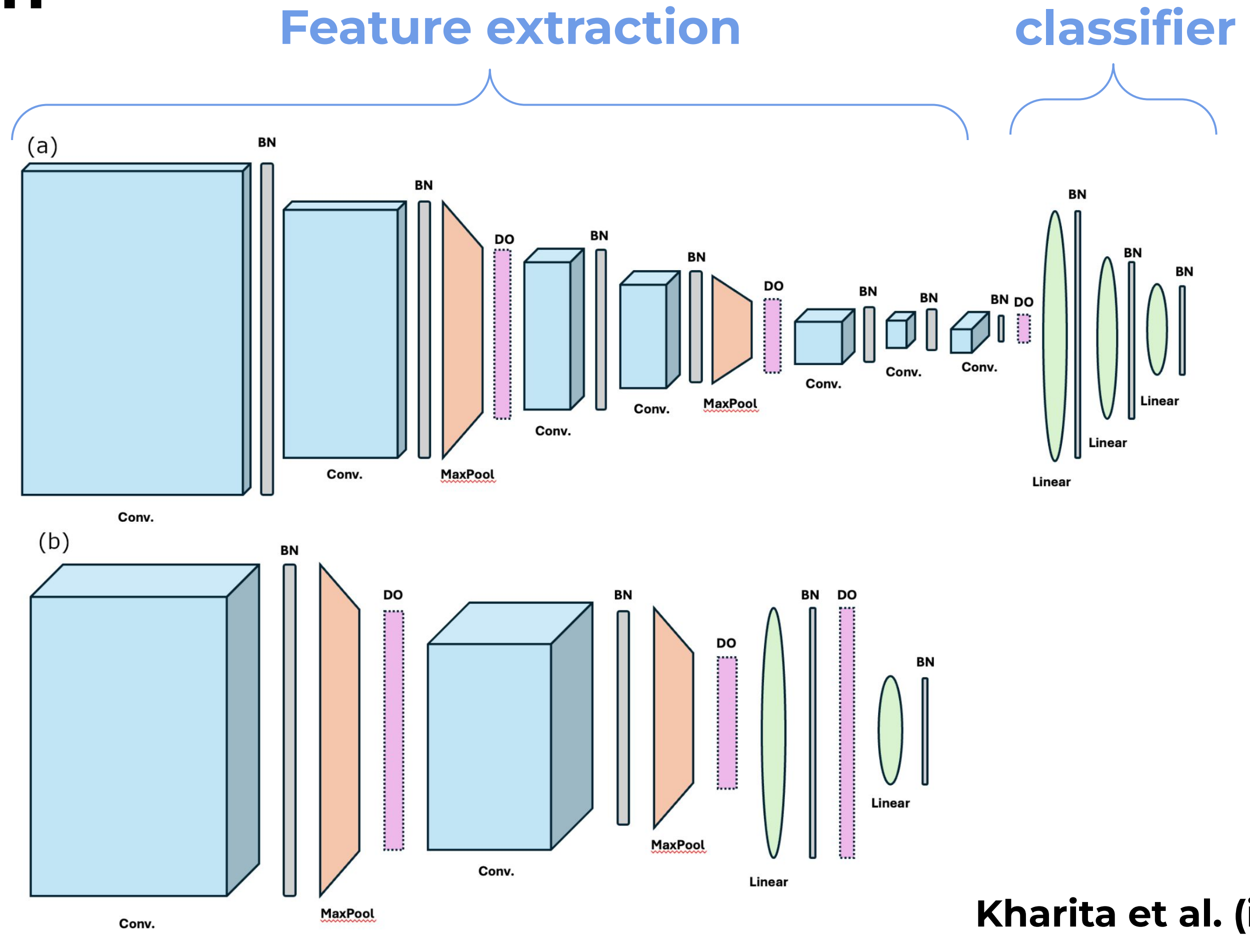
Kharita et al. (in prep)

Deep Learning: CNNs are great and simple for classification

Long skinny

QuakeXNet

Shallow & wide

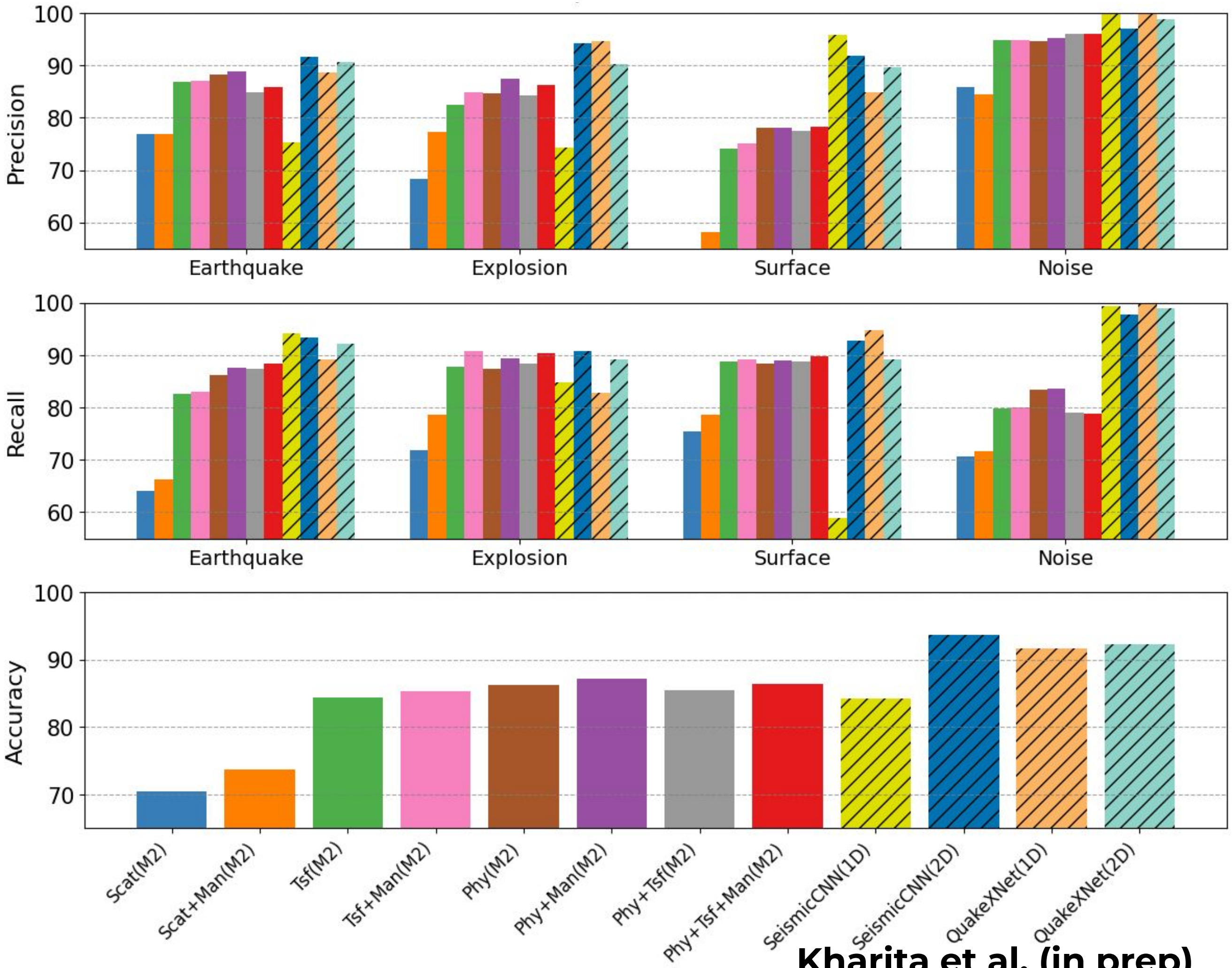


Kharita et al. (in prep)

Performance

Deep Learning outperforms classic Machine Learning

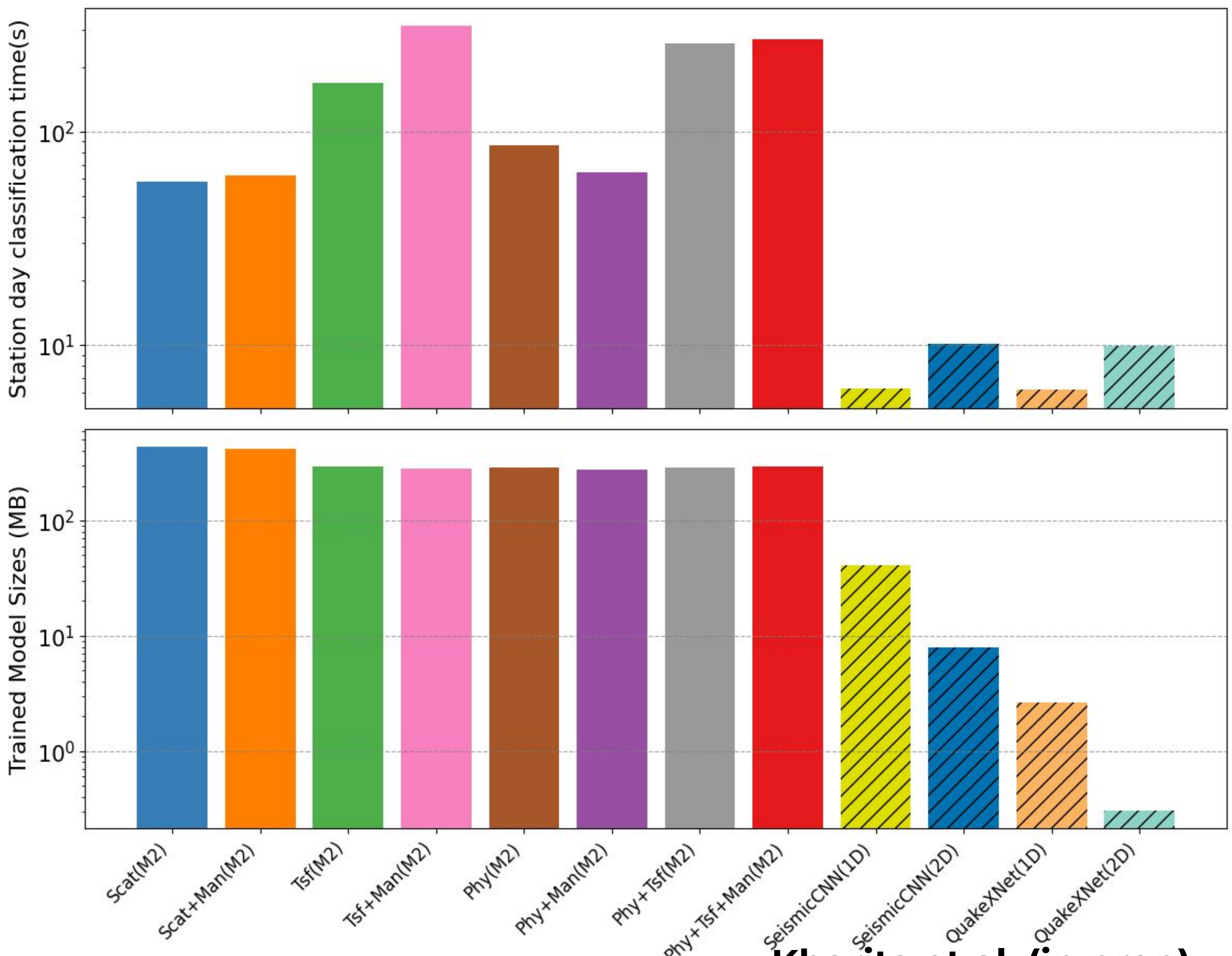
(on balanced data sets)



Kharita et al. (in prep)

Computational performance

Extracting features from time series takes much longer than DL inference.



Kharita et al. (in prep)

Merci

github.com/Denolle-lab
github.com/niyiyu
github.com/seisscoped/quakescope
github.com/congcy/ELEP
https://cascadiaquakes.github.io/2025_ML_TSC/intro.html

THE David & Lucile Packard Foundation
PNSN Pacific Northwest Seismic Network
NSF
Earth & Space Sciences