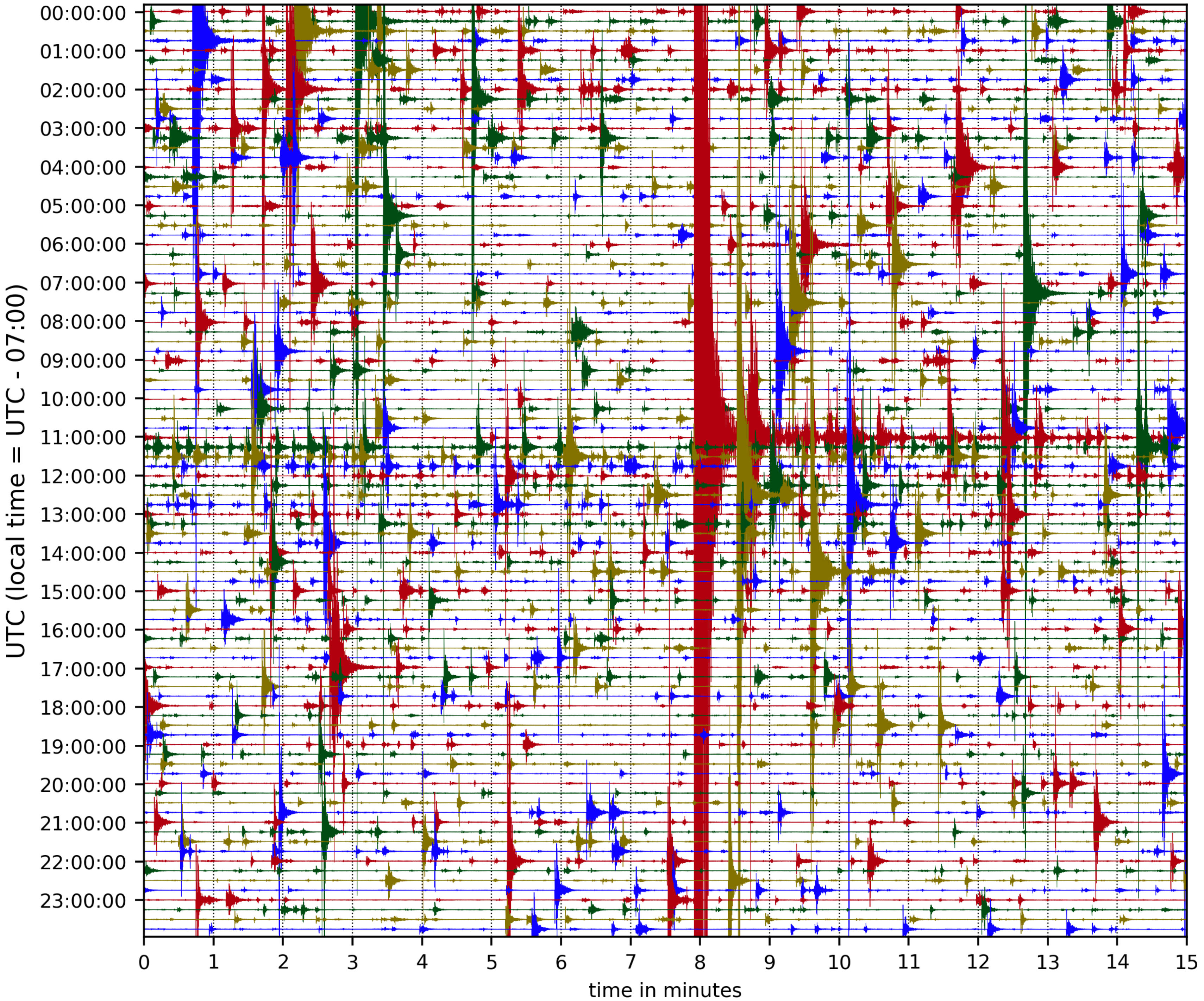
Searching for hidden earthquakes in Southern California

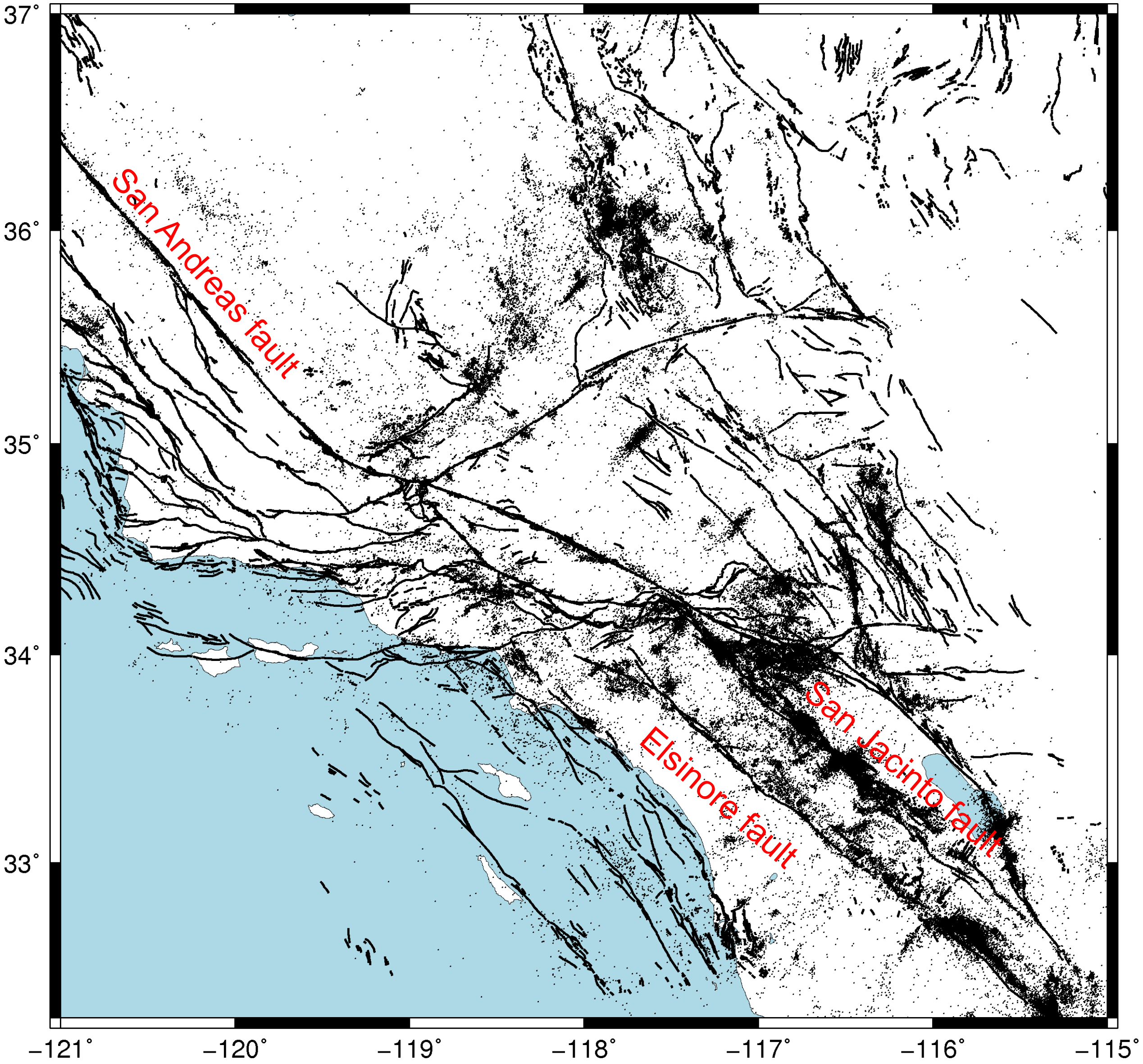
Zachary E. Ross
California Institute of Technology

It is well known that the smallest earthquakes are the most numerous

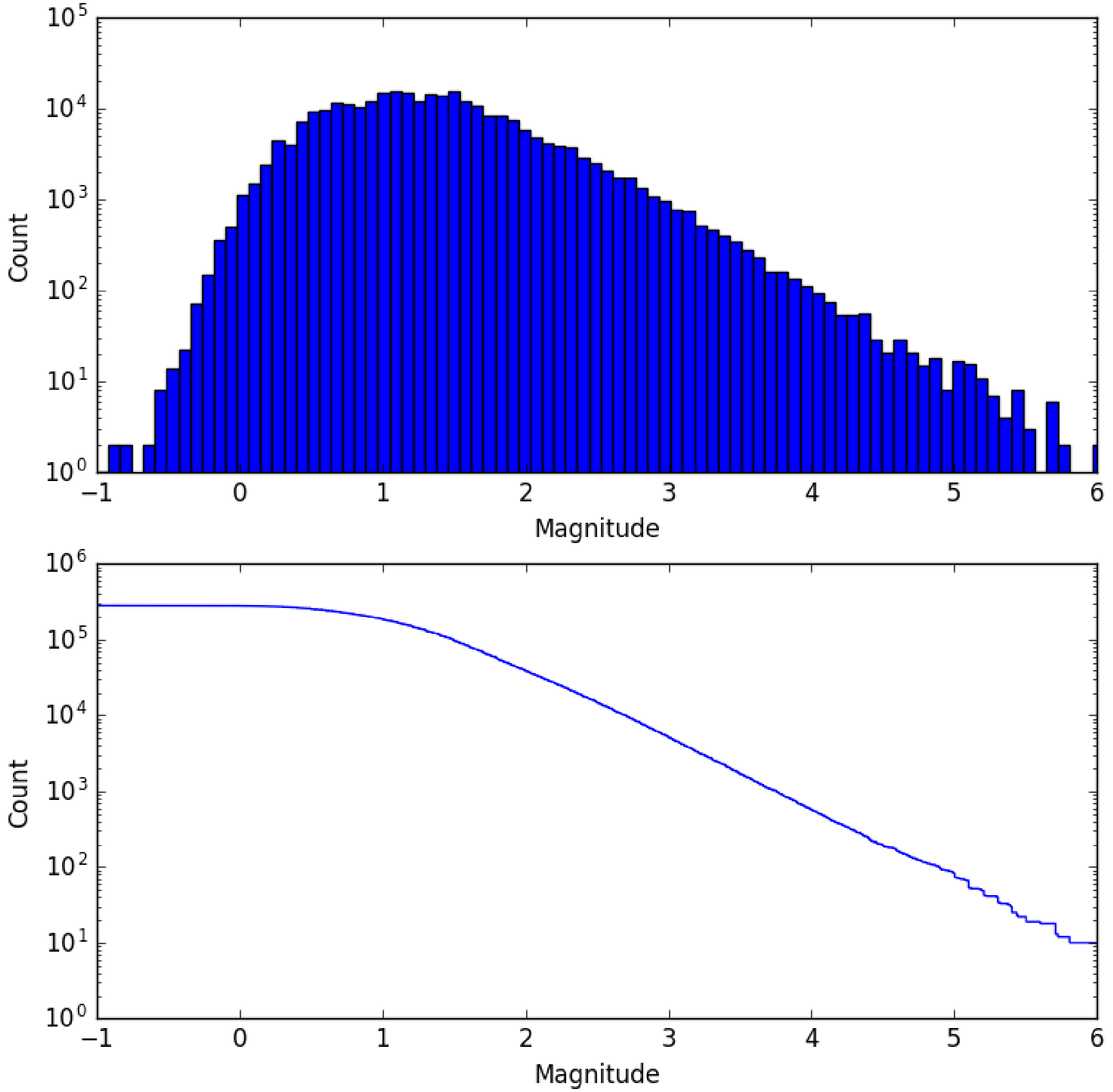
Standard techniques for detecting earthquakes typically miss >90% of events recorded in the data

The hidden events fill in the gaps in the earthquake record and tell a more complete story



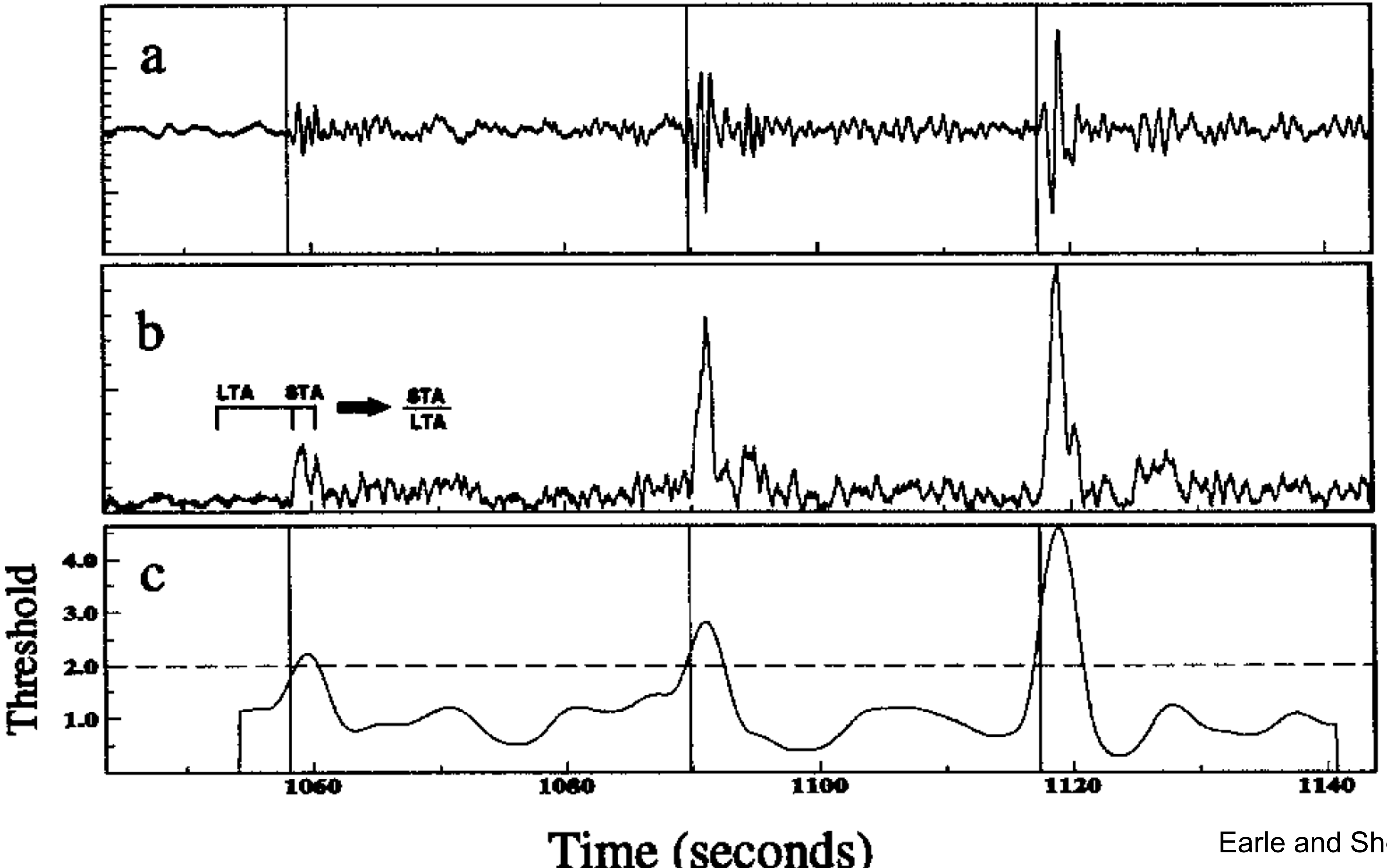


SCSN catalog (2008-present) ~180,000 events



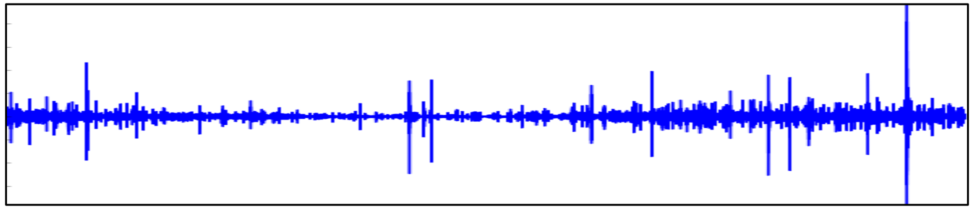
However, the SCSN catalog is incomplete for M < 1.3

Determination and Timing of Phase Arrivals



Earle and Shearer (1994)

A typical seismic network workflow



↓

Run phase pickers on seismic data

↓

Associate phase detections on multiple sensors that share a common source

↓

Obtain earthquake locations from picks

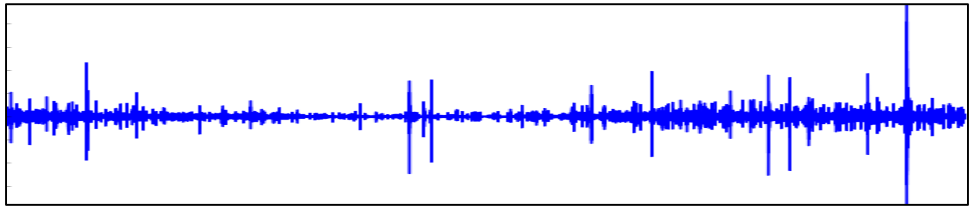
↓

Calculate magnitudes

↓

...

A typical seismic network workflow



↓

Run phase pickers on seismic data

↓

Associate phase detections on multiple sensors that share a common source

↓

Obtain earthquake locations from picks

↓

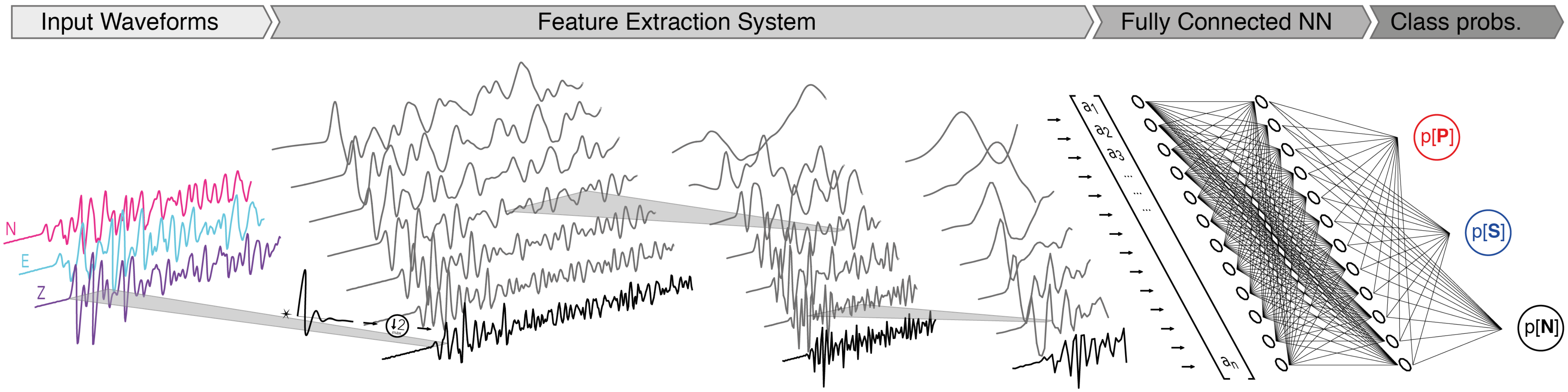
Calculate magnitudes

↓

...

Convolutional neural networks

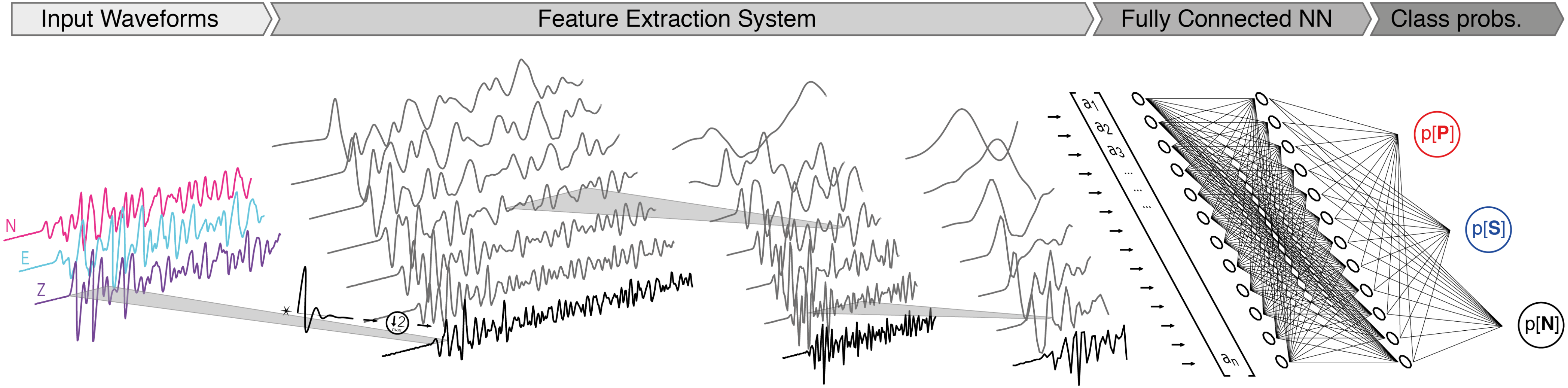
- Input images directly (or seismograms), without need for feature extraction
- Learnable feature extraction systems
- Hierarchical network of learned filters to be convolved with input (convolution layer)
- Periodic decimation of information between layers to span all length scales (pooling layer)



Ross et al. (2018), BSSA

Convolutional neural networks

- Excellent at invariant pattern recognition (e.g. translation, reflection, distortion)
- Capable of generalizing the knowledge contained in extremely large datasets
  - Not necessary for input to exactly match something previously observed
- Major limitation is large amounts of labeled data samples

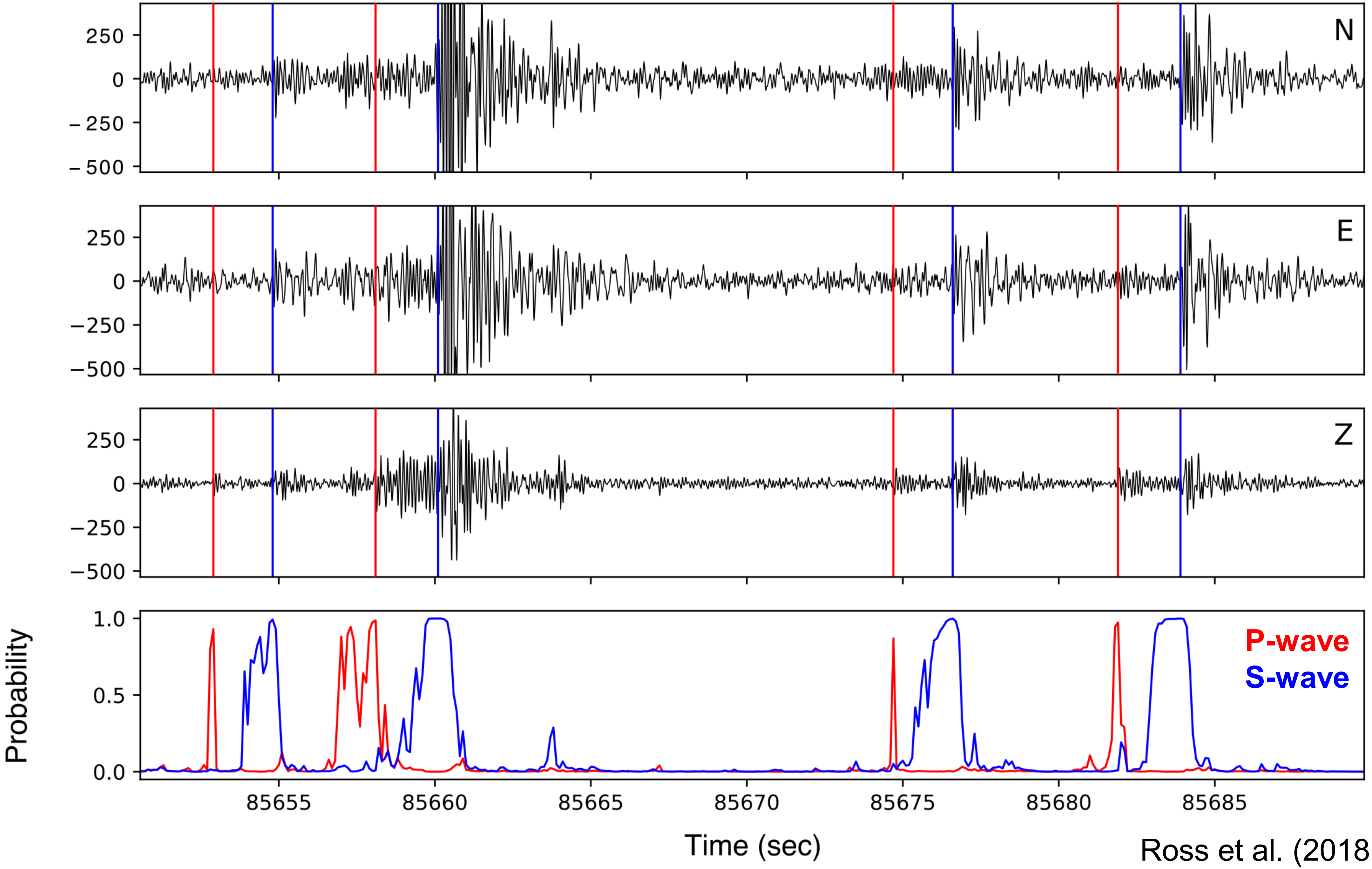


Ross et al. (2018), BSSA

Examples of labeled datasets in seismology

- Southern California Seismic Network has manually labeled phase data since 1932
    - Earthquake catalog consists of > 2,000,000 events
    - > 20 million hand-measured P & S arrival times
    - > 5 million hand-measured first-motion polarities

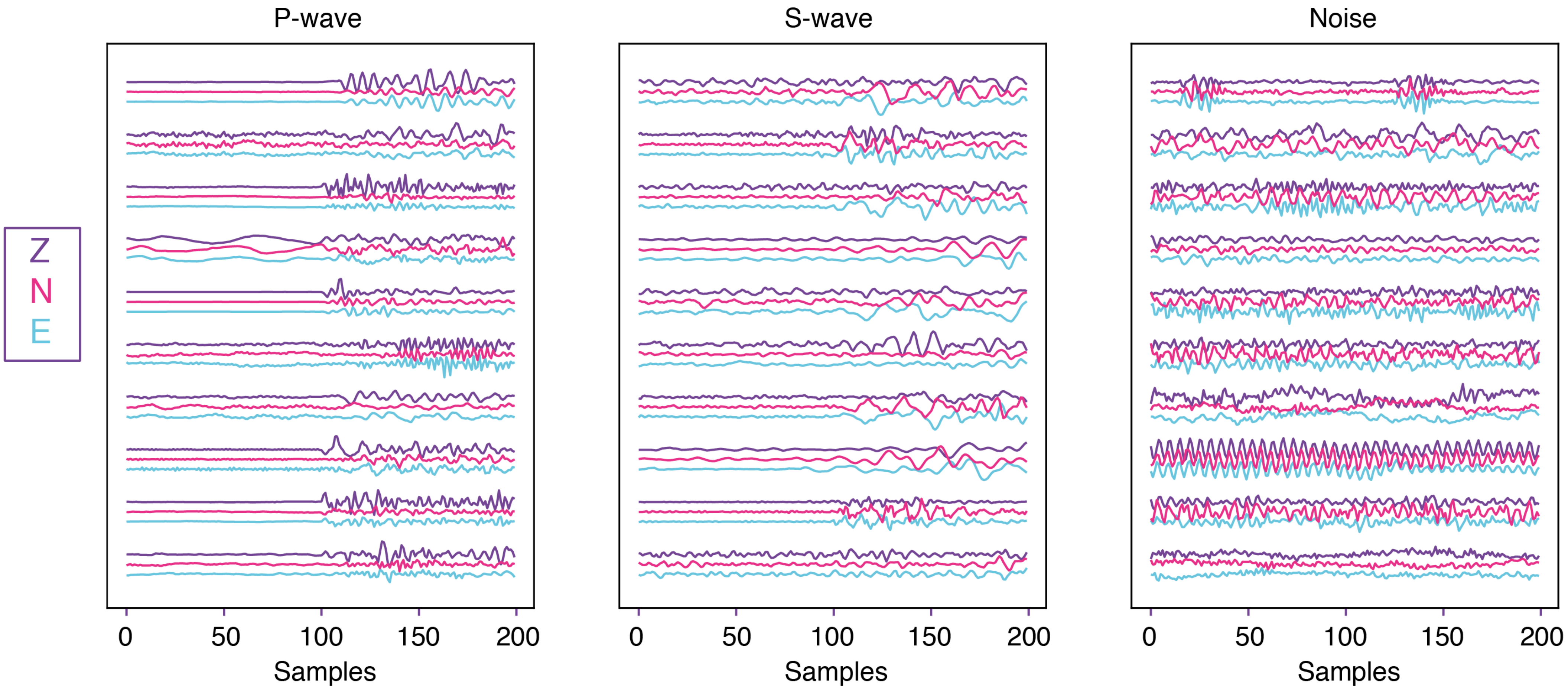
- Goal: use these datasets to train deep neural networks for earthquake detection



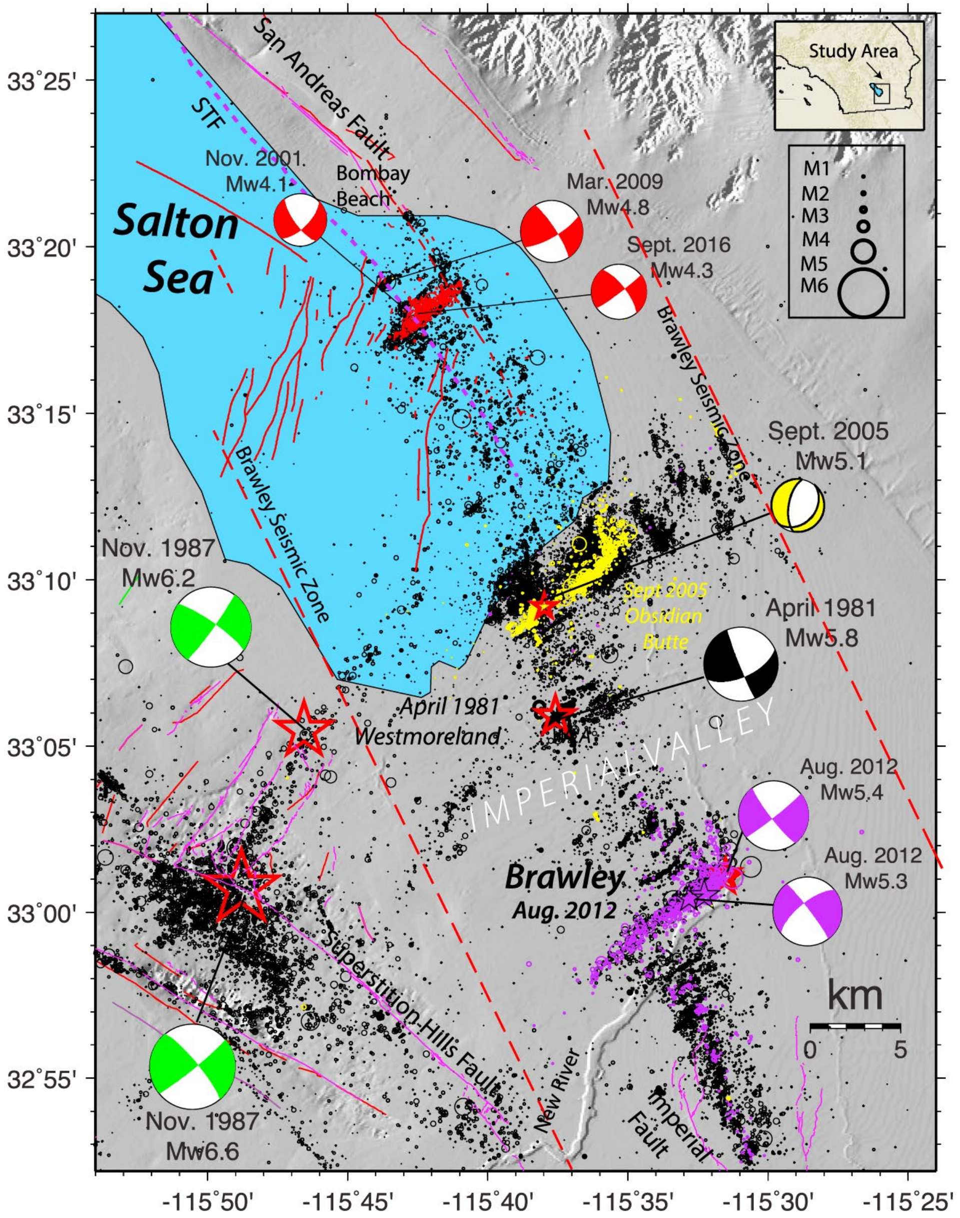
Ross et al. (2018), BSSA

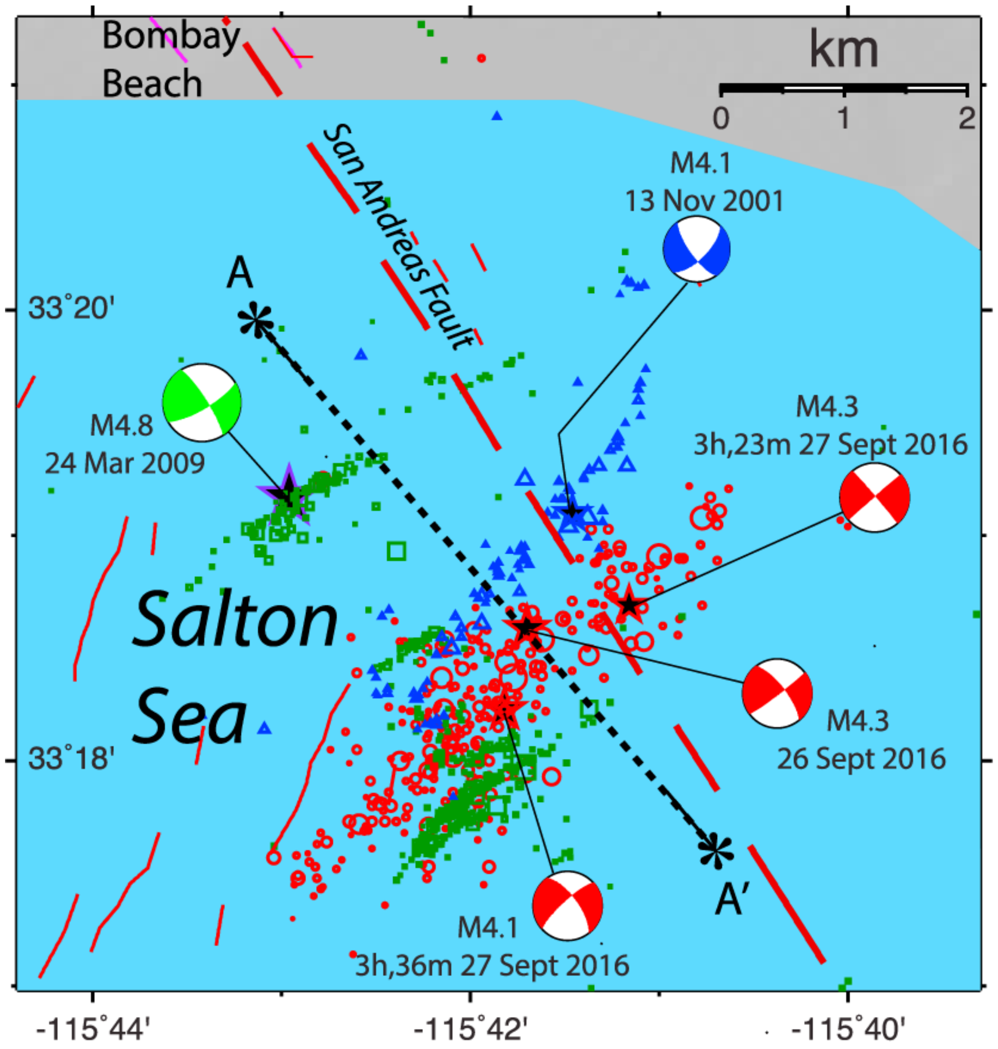
Training/validation dataset:

- 273,882 earthquakes recorded by SCSN (2000-2017)
- 692 broadband and short-period seismic stations
- 1.5 million P-wave 3-c seismograms
- 1.5 million S-wave 3-c seismograms
- 1.5 million pre-event noise seismograms

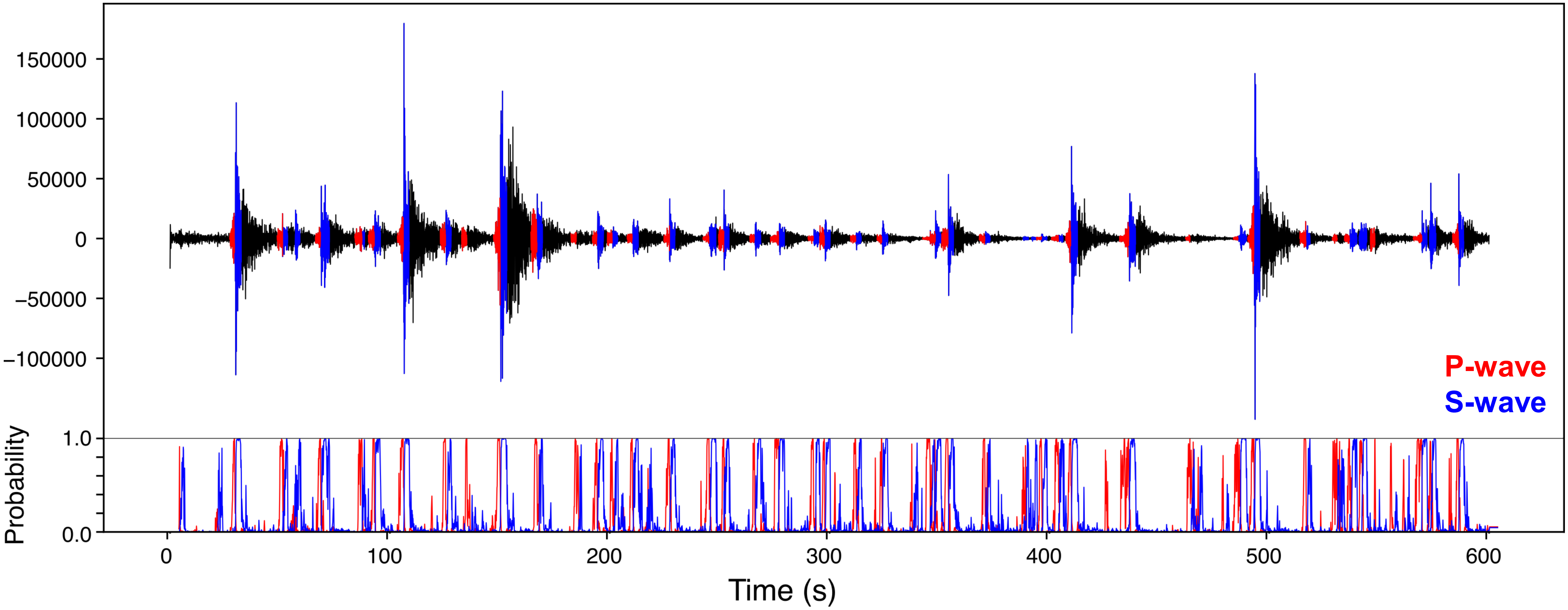


Application to 2016 Bombay Beach, CA swarm





Hauksson et al. (2017) *GRL*

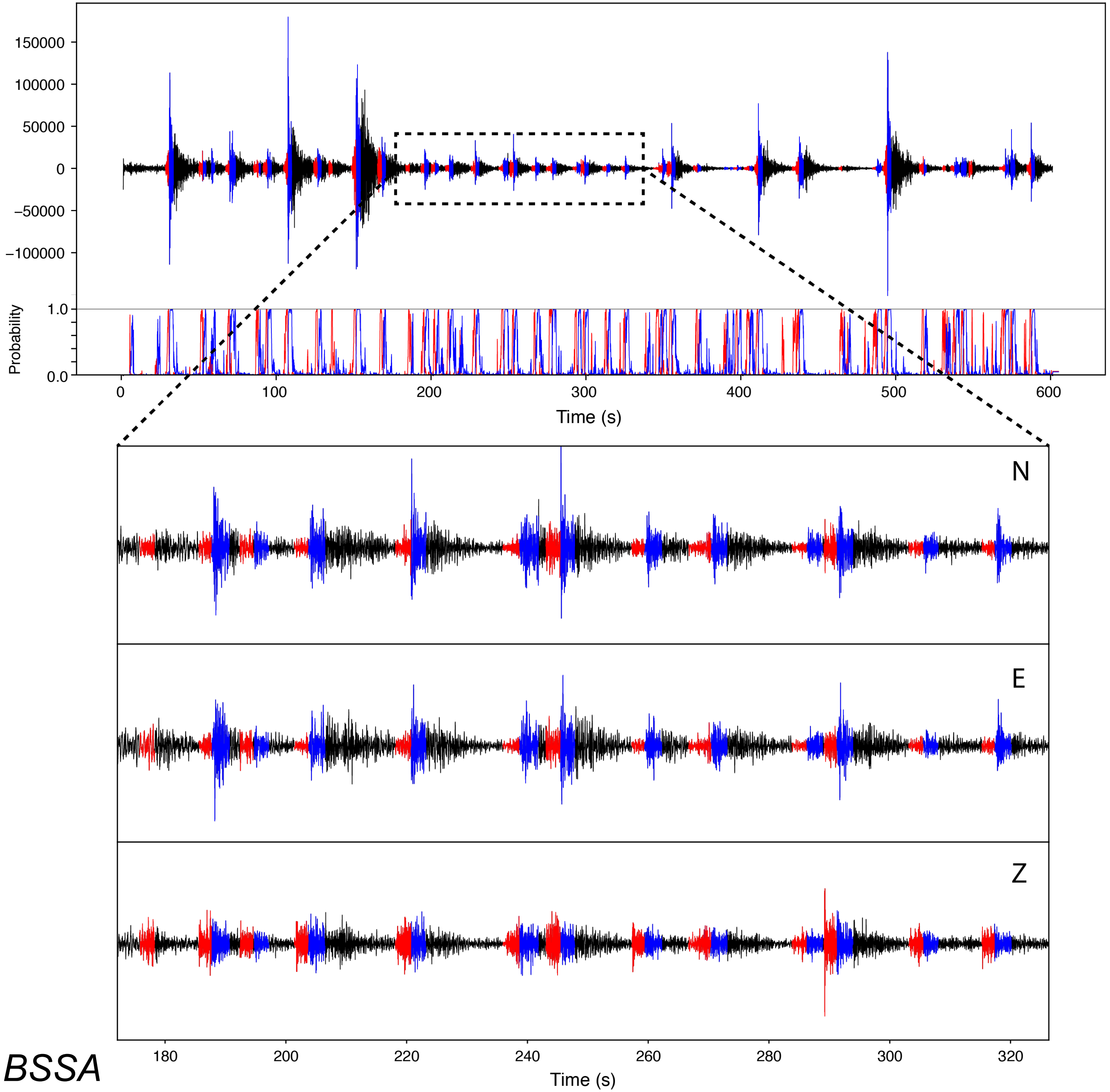


Ross et al. (2018), BSSA

Here we focus on the first 10 minutes after the onset of the swarm

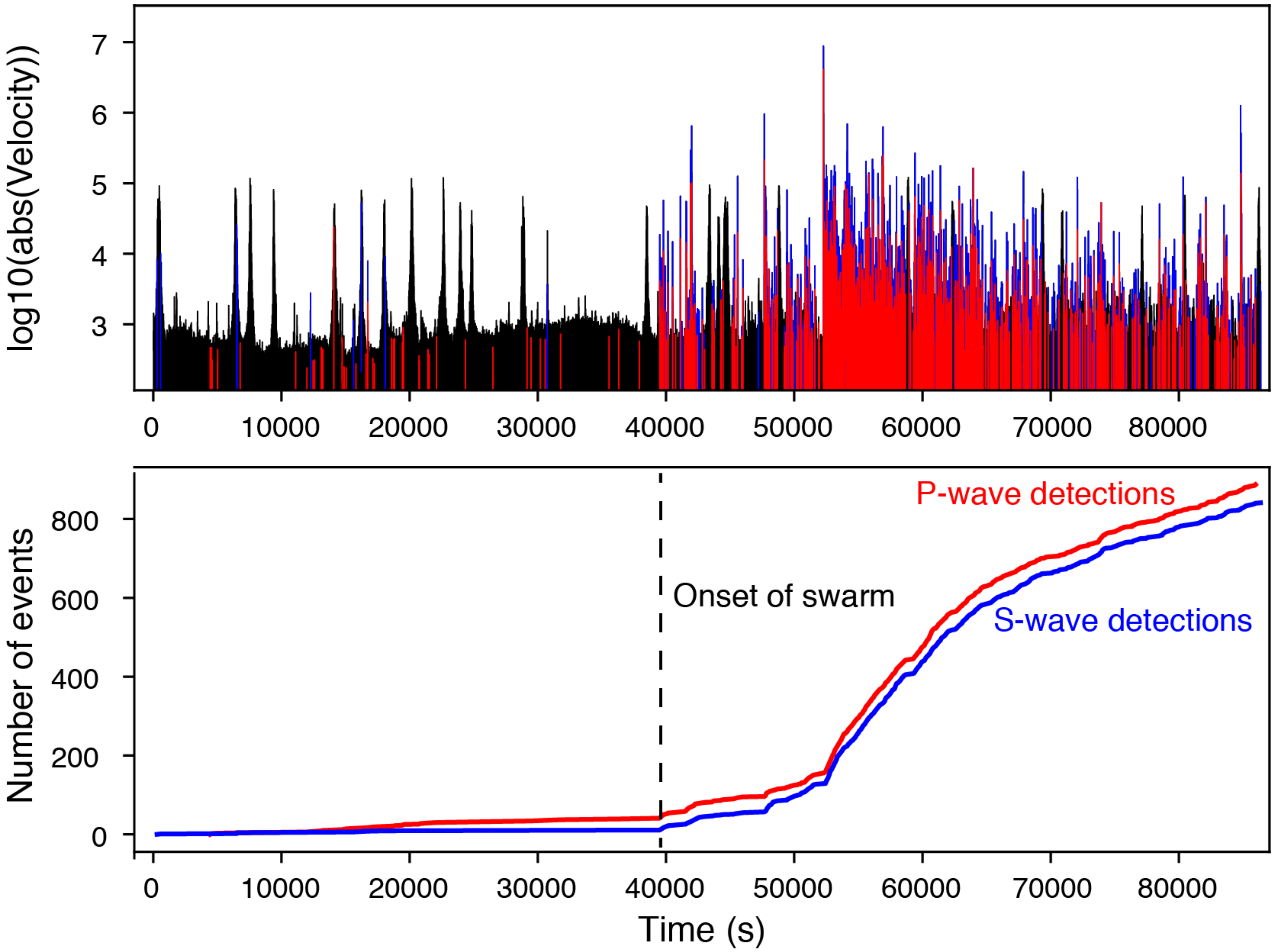
Within a span of 2 minutes, 13 P-waves and 12 S-waves are detected

Several of these events are separated less than 2 sec apart



Ross et al. (2018), *BSSA*

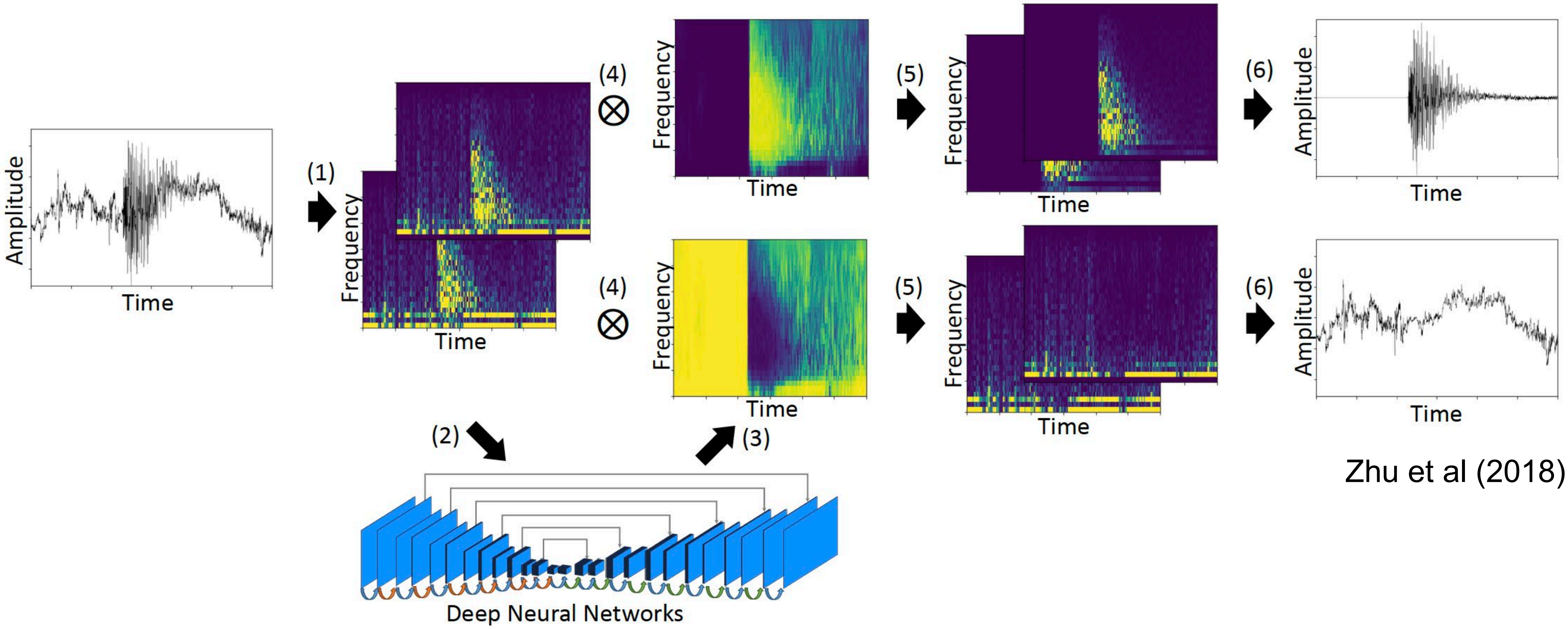
Results over a 24 hour period



Here we detect 8.3 times as many events as in the SCSN catalog

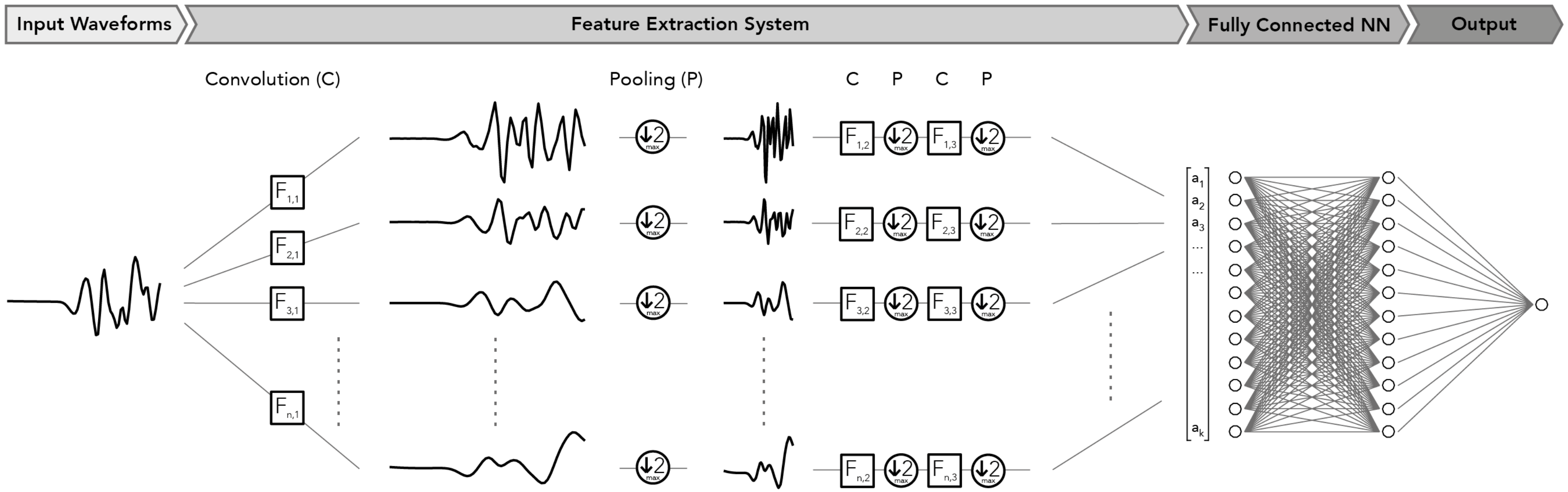
Ross et al. (2018), BSSA

Deep learning shows exciting promise for signal denoising

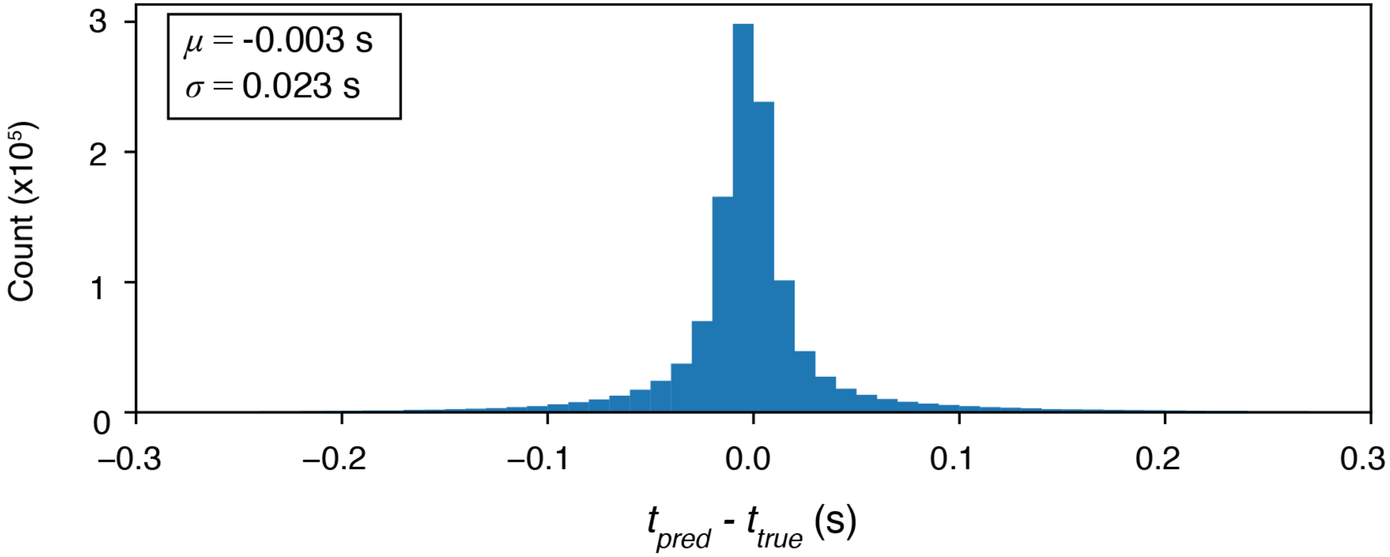


Zhu et al (2018)

P-wave arrival picking accuracy

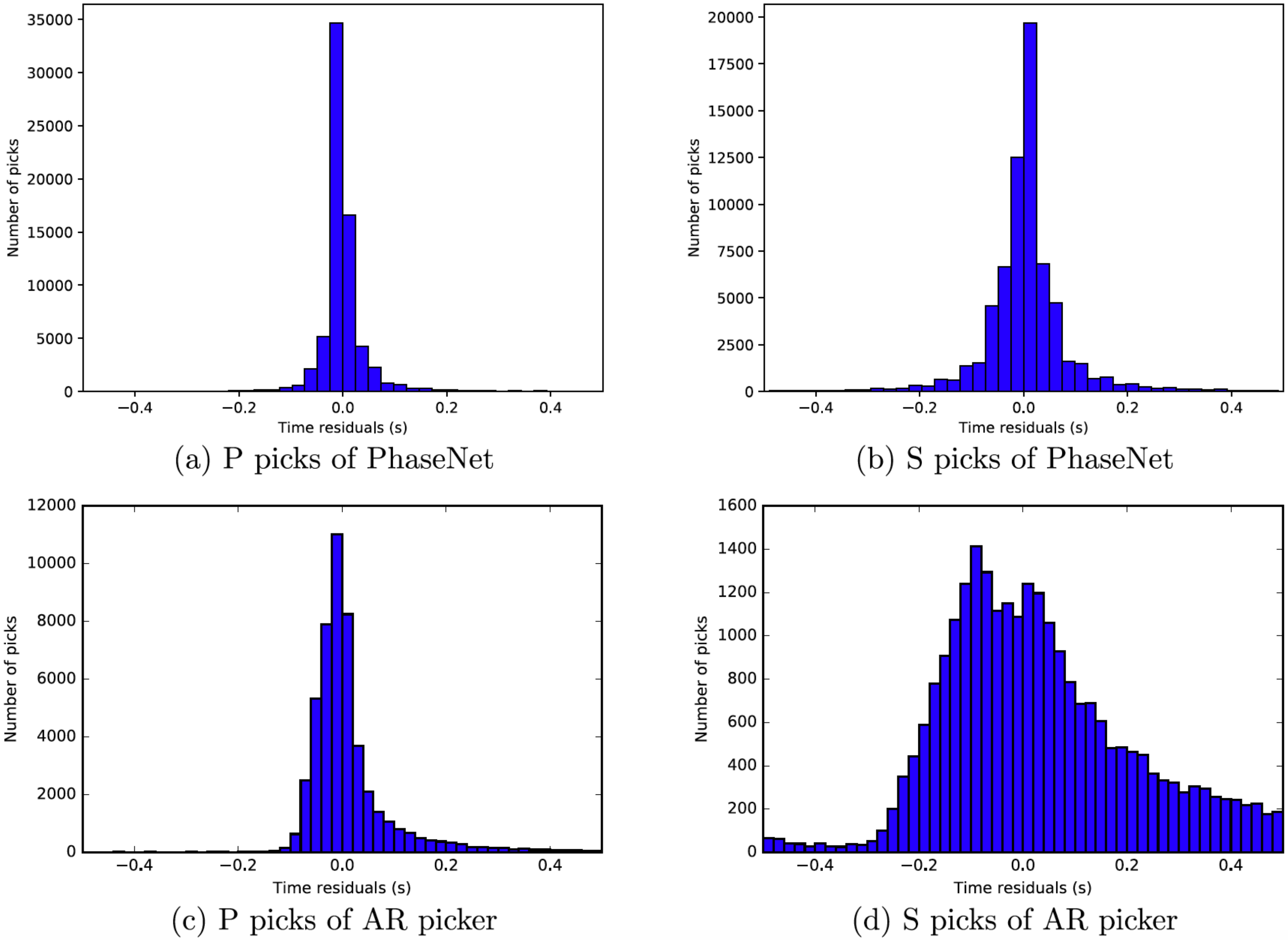


Ross et al. (2018), JGR



75% of 1.1 million picks are within 0.028 sec of analyst pick

Additional performance results with PhaseNet



(a) P picks of PhaseNet

(b) S picks of PhaseNet

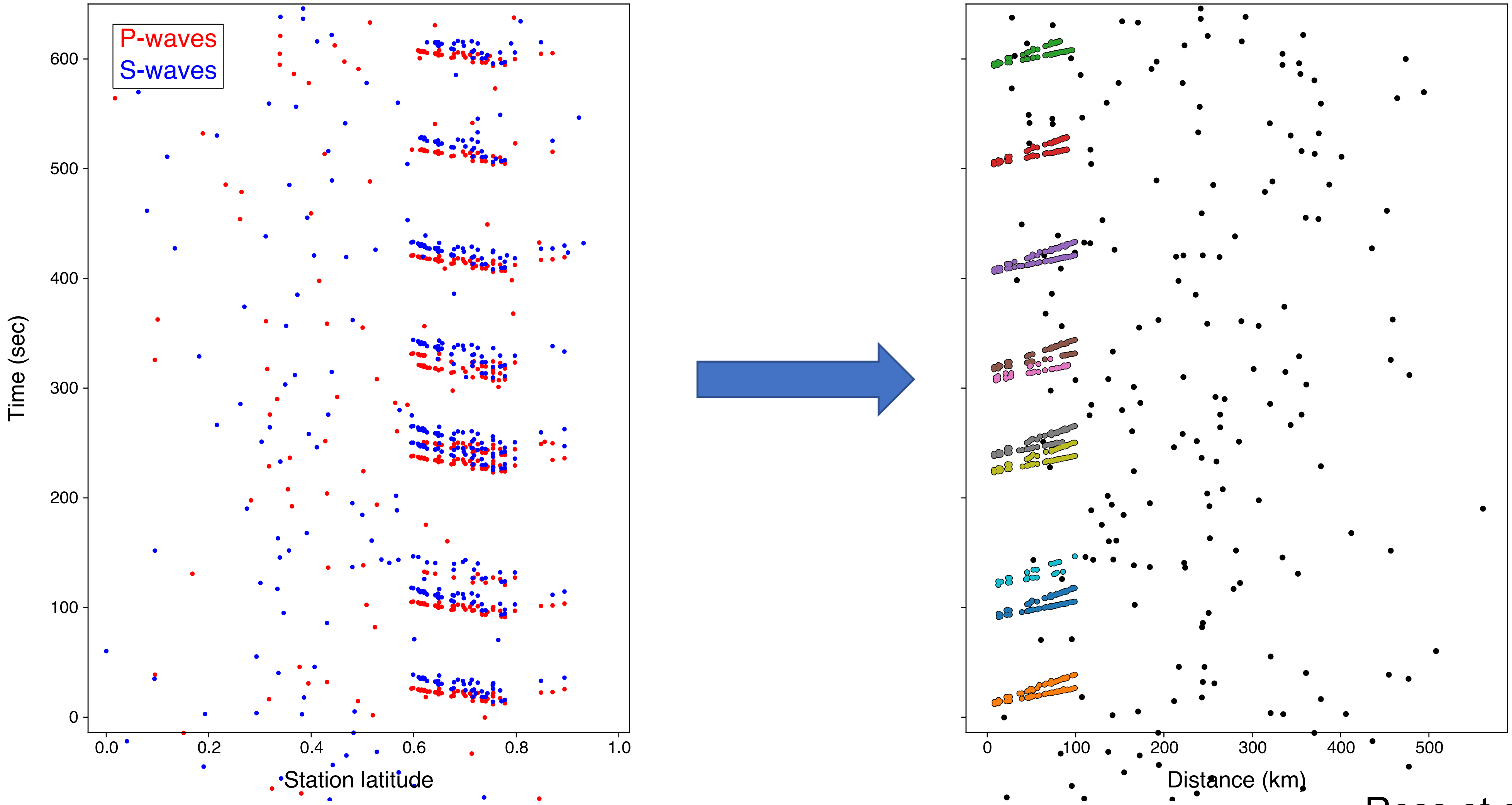
(c) P picks of AR picker

(d) S picks of AR picker

Zhu and Beroza (2018), *GJI*

The phase association problem:

Given an unknown number of earthquakes and a set of phase detections across a seismic network, assign each detection to the causative event



Ross et al. (2019), JGR

Background on Recurrent Neural Networks

- Standard neural networks lack a mechanism for learning structure in sequences
- RNN achieve this with an internal memory state

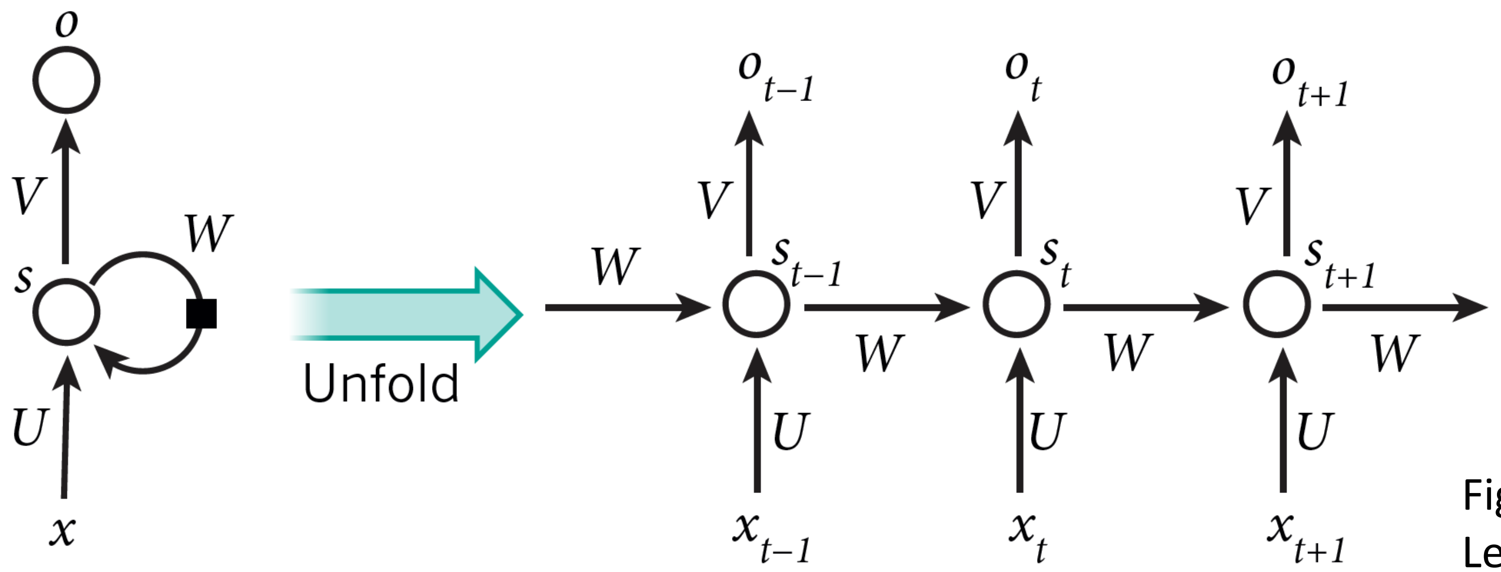


Figure from
LeCun et al. (2015)

Background on Recurrent Neural Networks

- Standard neural networks lack a mechanism for learning structure in sequences
- RNN achieve this with an internal memory state

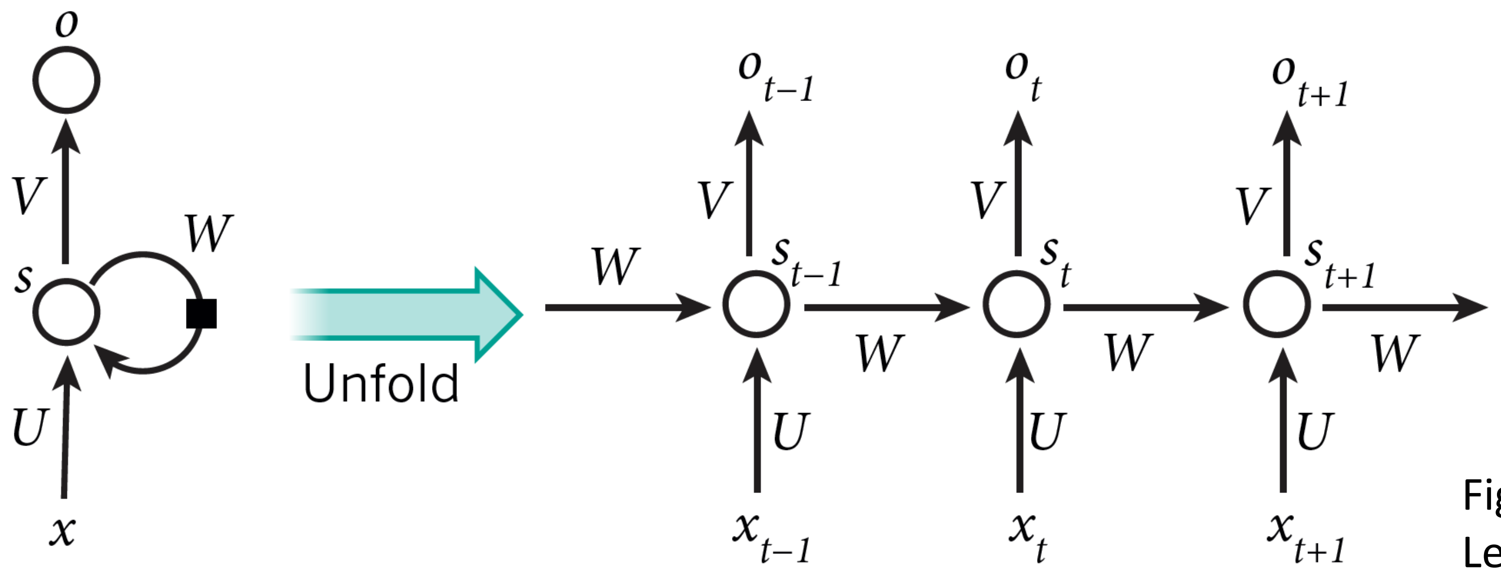
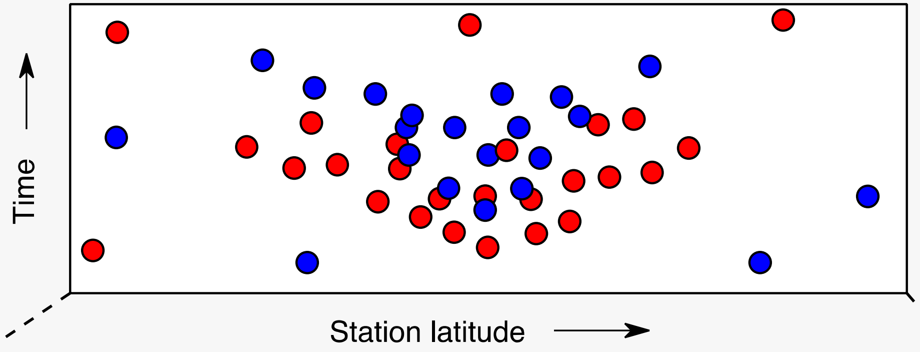


Figure from
LeCun et al. (2015)

- Predictions at each time step use context from earlier in sequence
- State of the art for speech recognition, natural language processing, many other domains of AI
- Common variants: LSTM (long short term memory), GRU (gated recurrent unit)

PhaseLink: A deep learning approach to seismic phase association

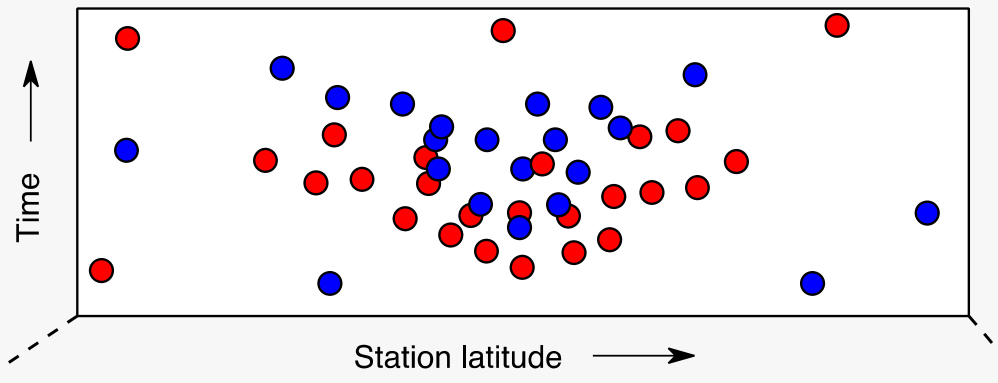
1. Collect picks over a seismic network



Station lat.	33.57	33.56	33.57	33.49	...	33.49	33.59	33.35
Station lon.	-116.67	-116.53	-116.22	-116.60	...	-116.60	-116.76	-116.56
Time (sec)	4.56	4.84	4.93	5.11	...	5.23	5.62	5.65
Phase type	P	P	S	P	...	S	P	S

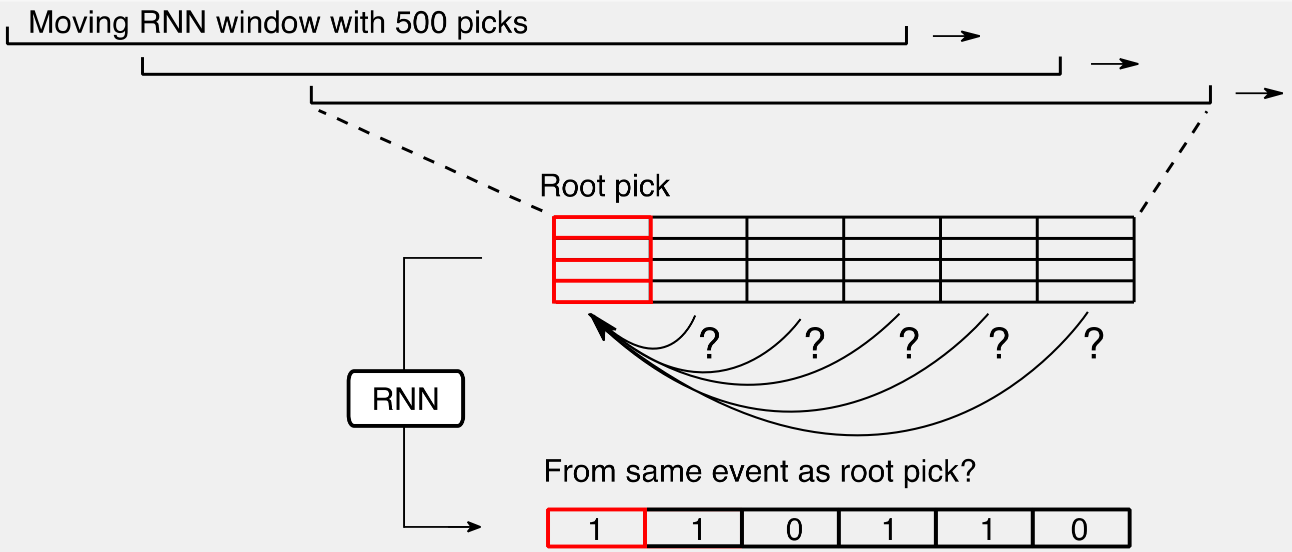
Ross et al. (2019), JGR

1. Collect picks over a seismic network



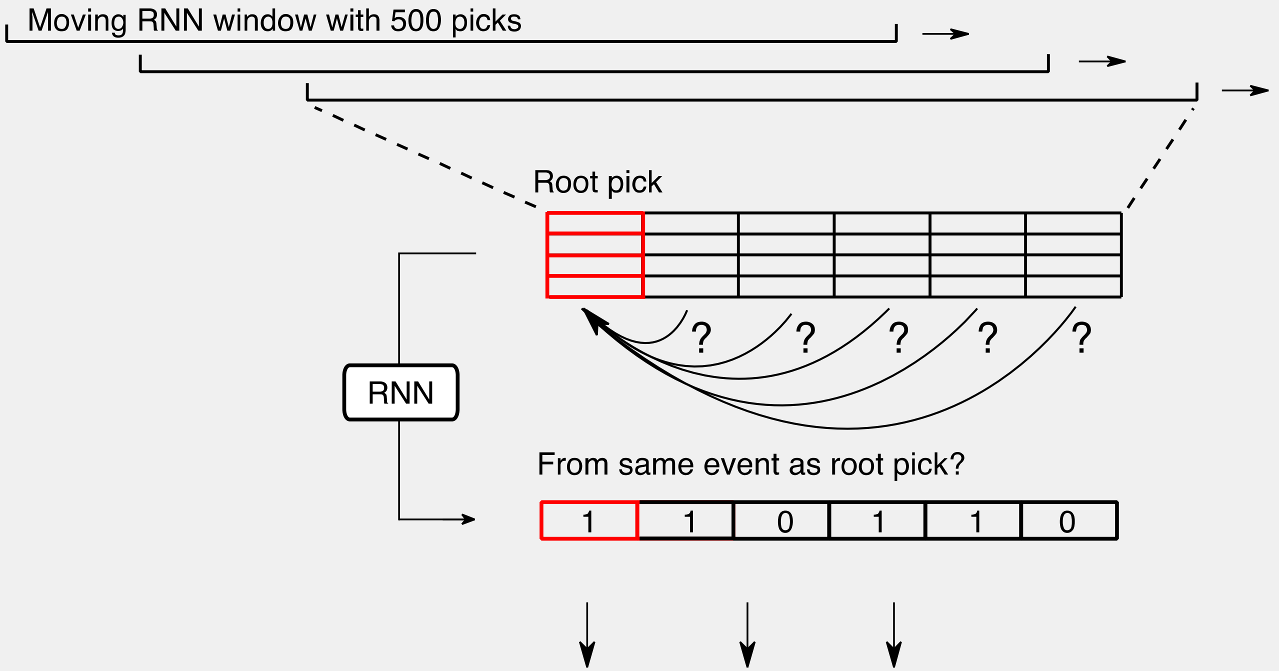
Station lat.	33.57	33.56	33.57	33.49	...	33.49	33.59	33.35
Station lon.	-116.67	-116.53	-116.22	-116.60	...	-116.60	-116.76	-116.56
Time (sec)	4.56	4.84	4.93	5.11	...	5.23	5.62	5.65
Phase type	P	P	S	P	...	S	P	S

2. In moving window, predict which picks are from same event as root



Ross et al. (2019), JGR

2. In moving window, predict which picks are from same event as root



3. Aggregate predictions for all windows

Event	x	x	A	A	x	A	B	B	B	C	C	C	x	C	...
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	...
		1	0	0	0	0	0	0	0	0	0	0	0	0	...
			1	1	0	1	0	0	0	0	0	0	0	0	...
				1	0	1	0	0	0	0	0	0	0	0	...
					1	0	0	0	0	0	0	0	0	0	...
						1	0	0	0	0	0	0	0	0	...
							1	1	1	0	0	0	0	0	...
								1	1	0	0	0	0	0	...
									1	0	0	0	0	0	...
										1	1	1	0	1	...
											1	1	0	1	...
												1	0	1	...
													1	0	...
														1	...
															...

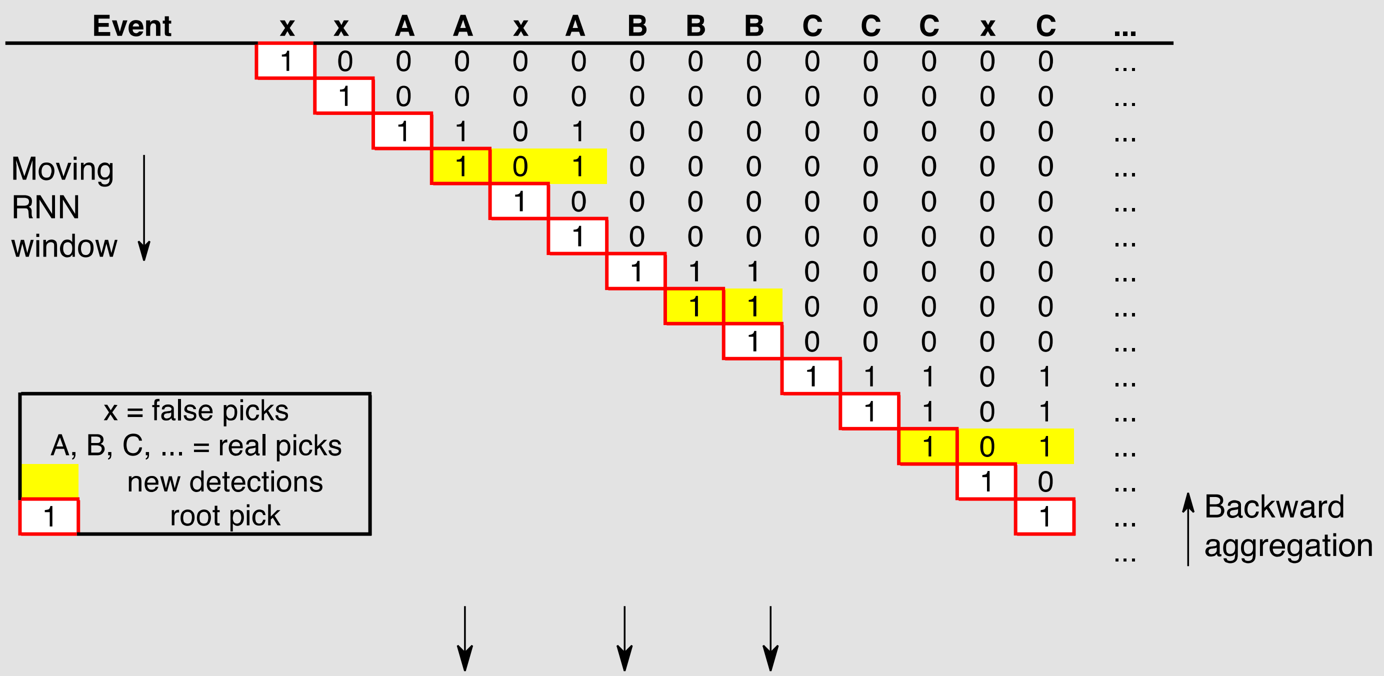
Moving RNN window

Backward aggregation

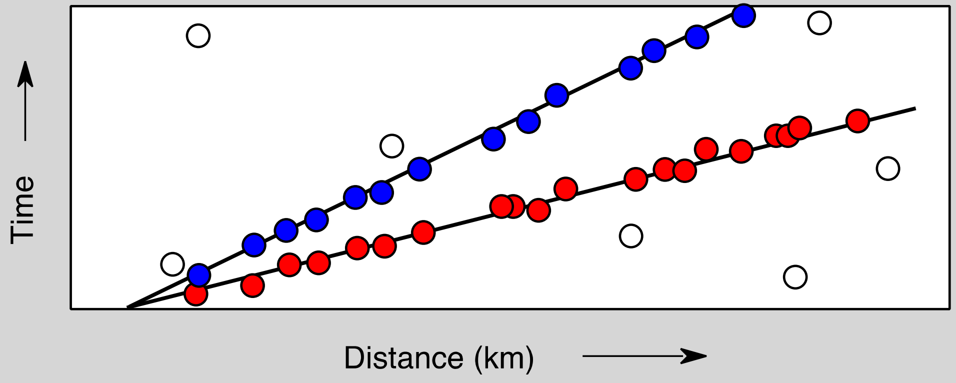
x = false picks
A, B, C, ... = real picks
new detections
1 root pick

Ross et al. (2019), JGR

3. Aggregate predictions for all windows

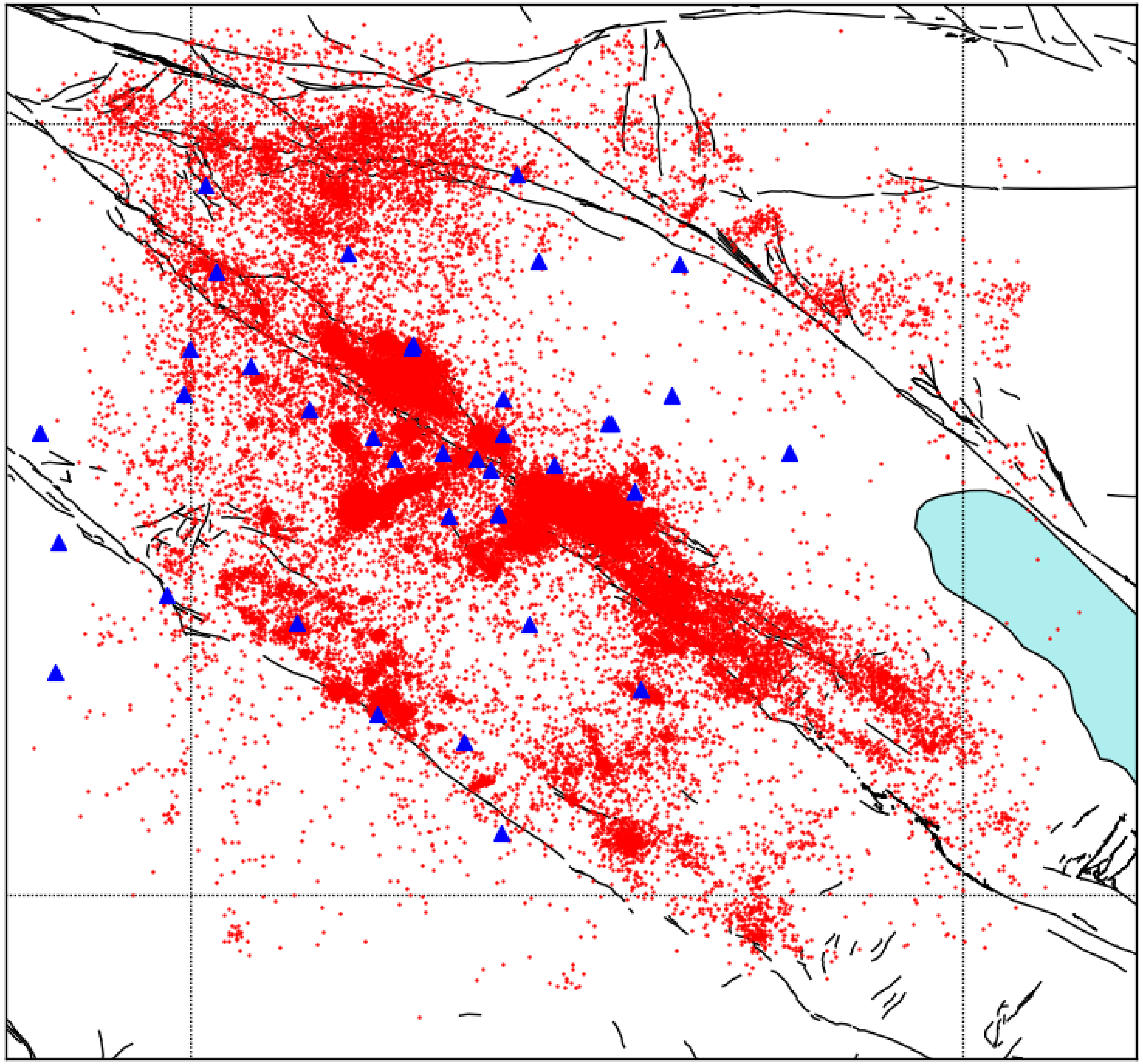


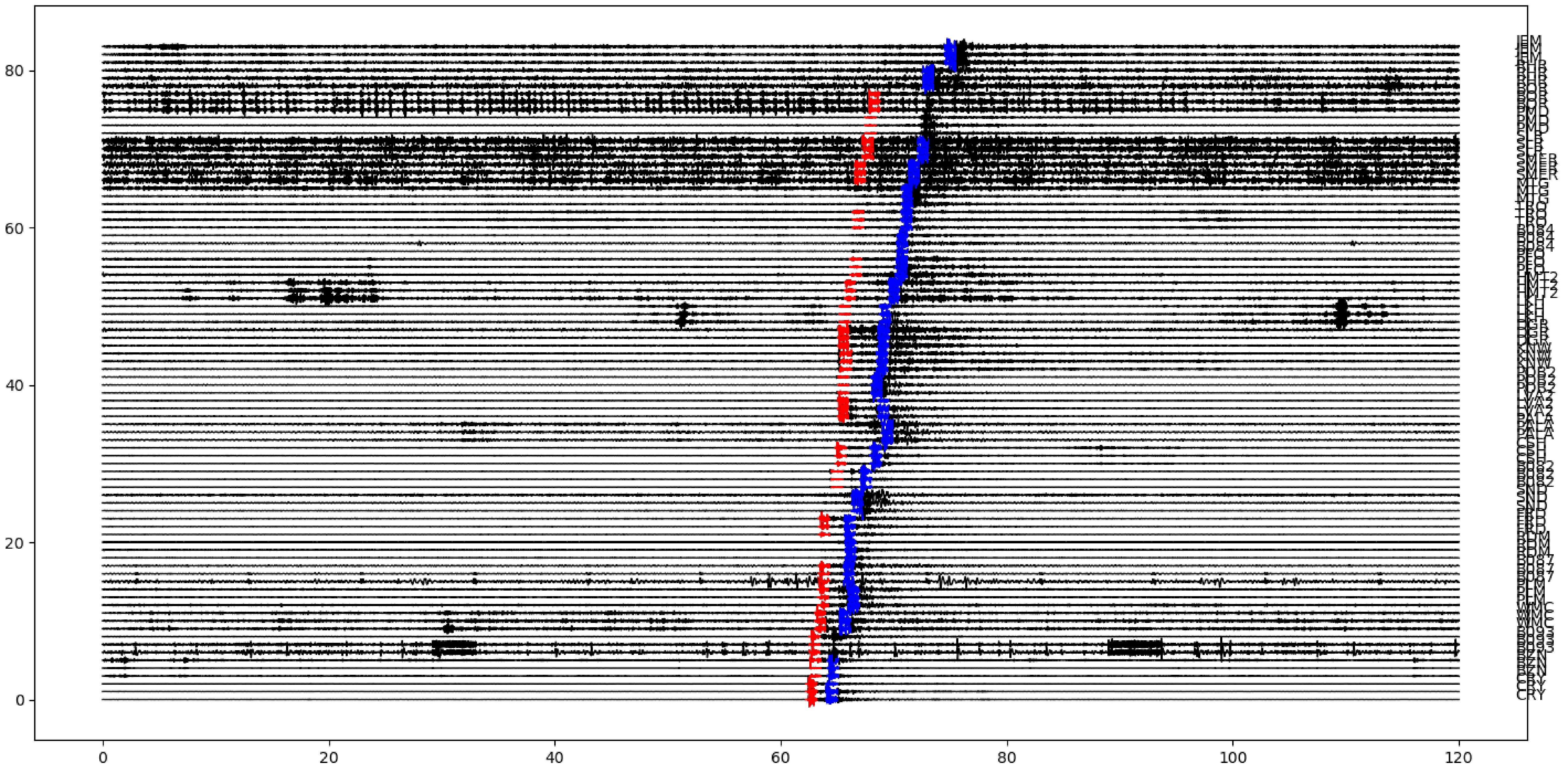
4. Pick sequence is fully associated



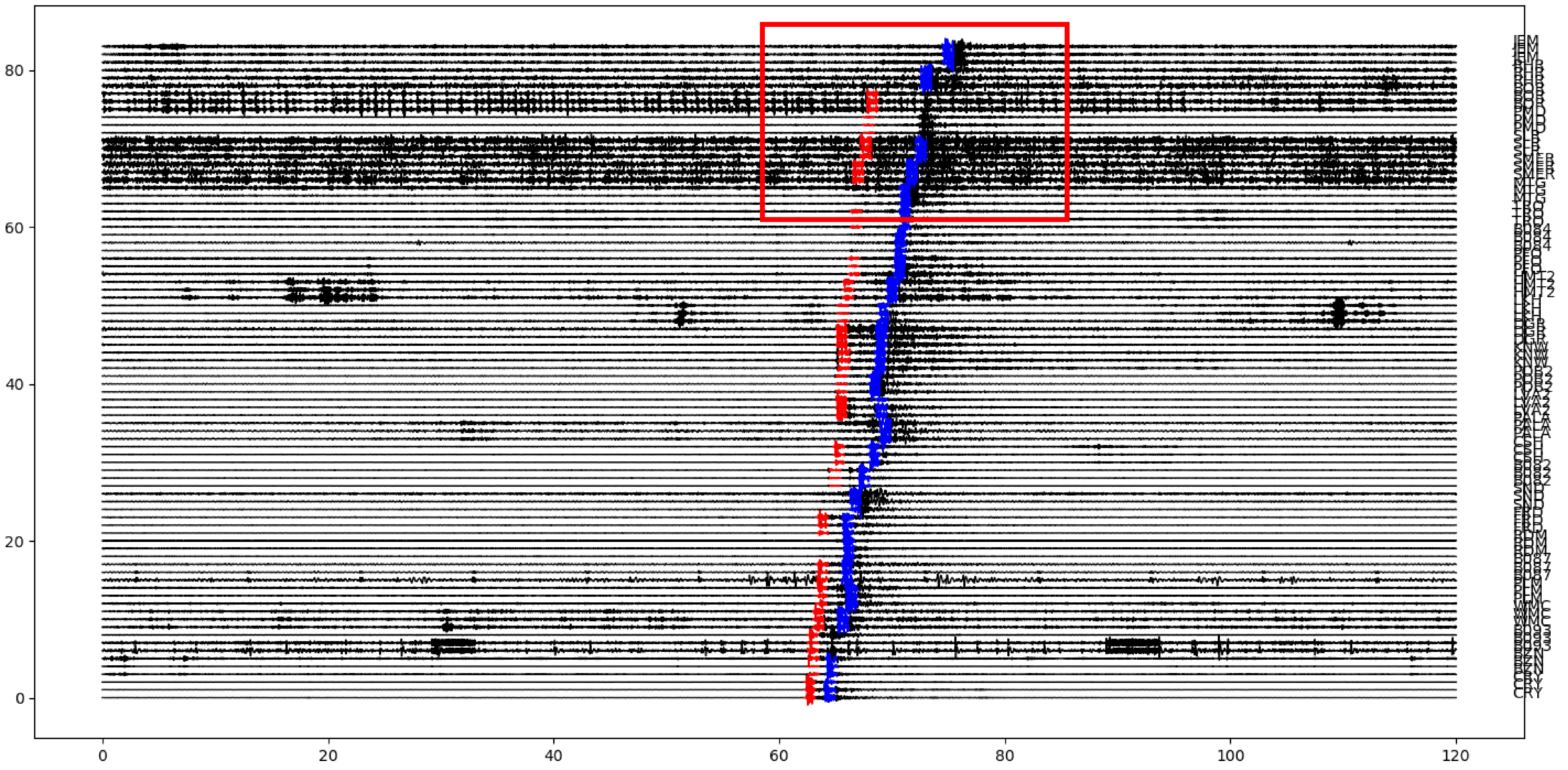
Ross et al. (2019), JGR

End-to-end processing of 3 years continuous data from scratch: 86,000+ events detected

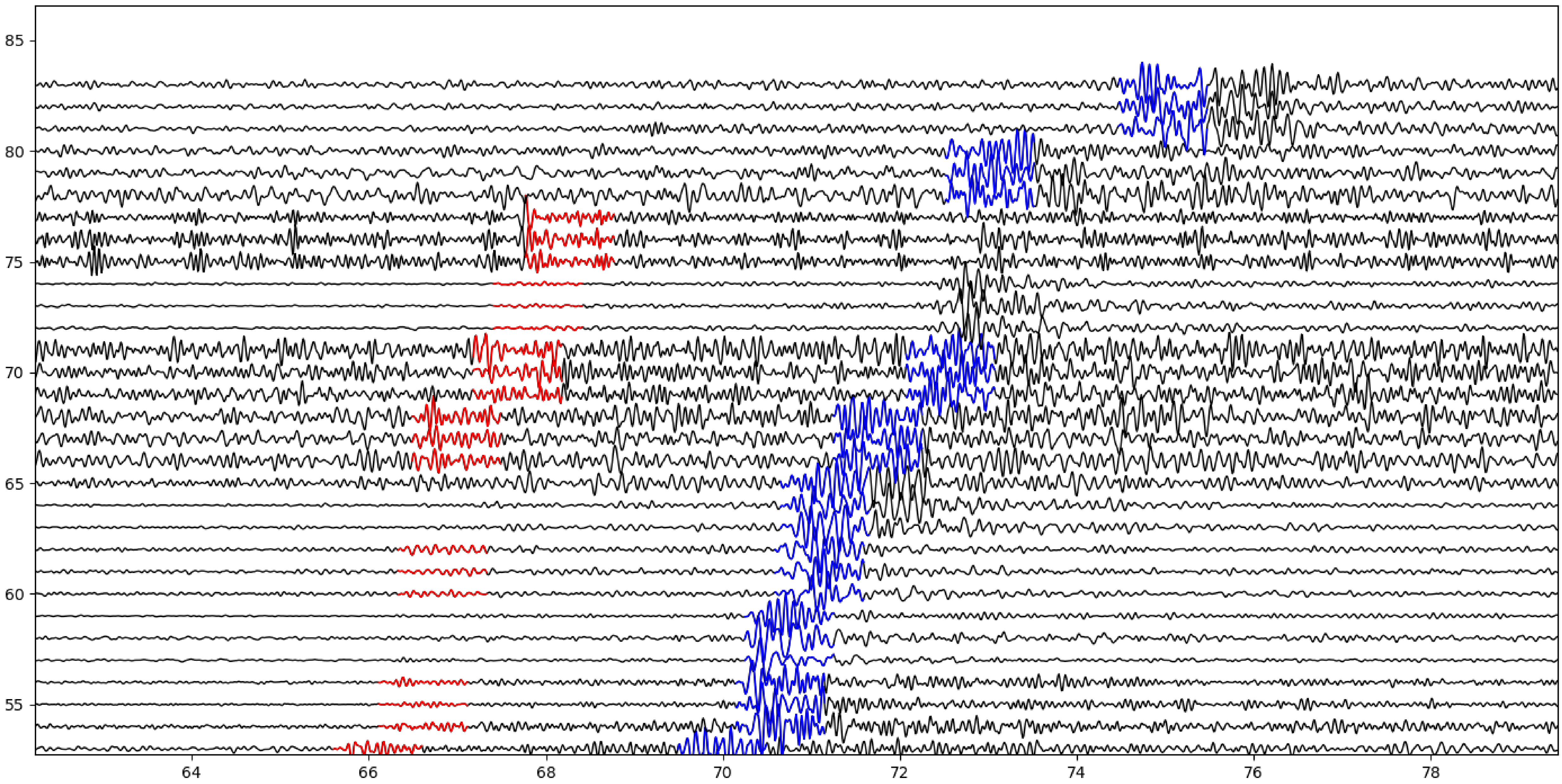




80
60
40
20
0
0
20
40
60
80
100
120

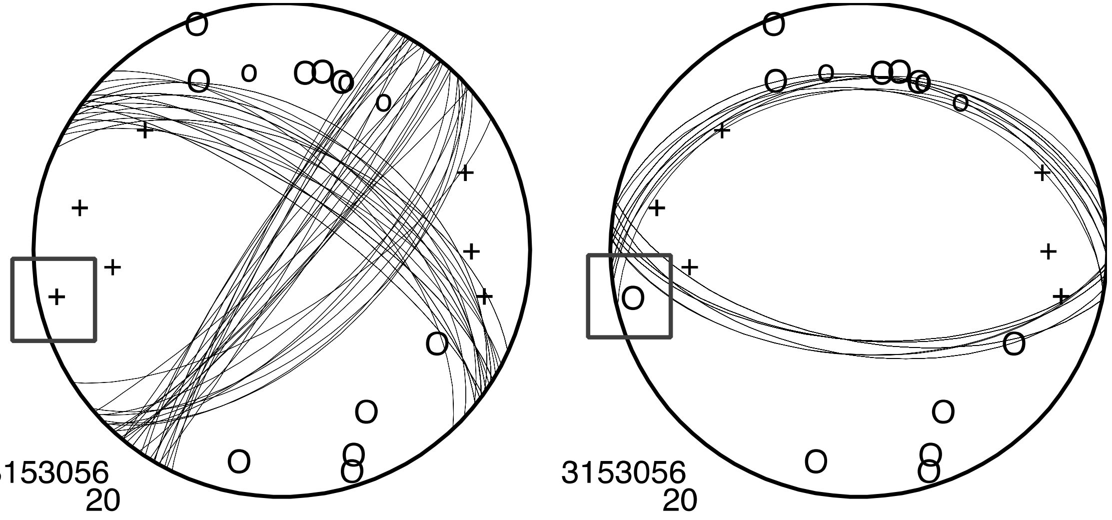


0
20
40
60
80
100
120



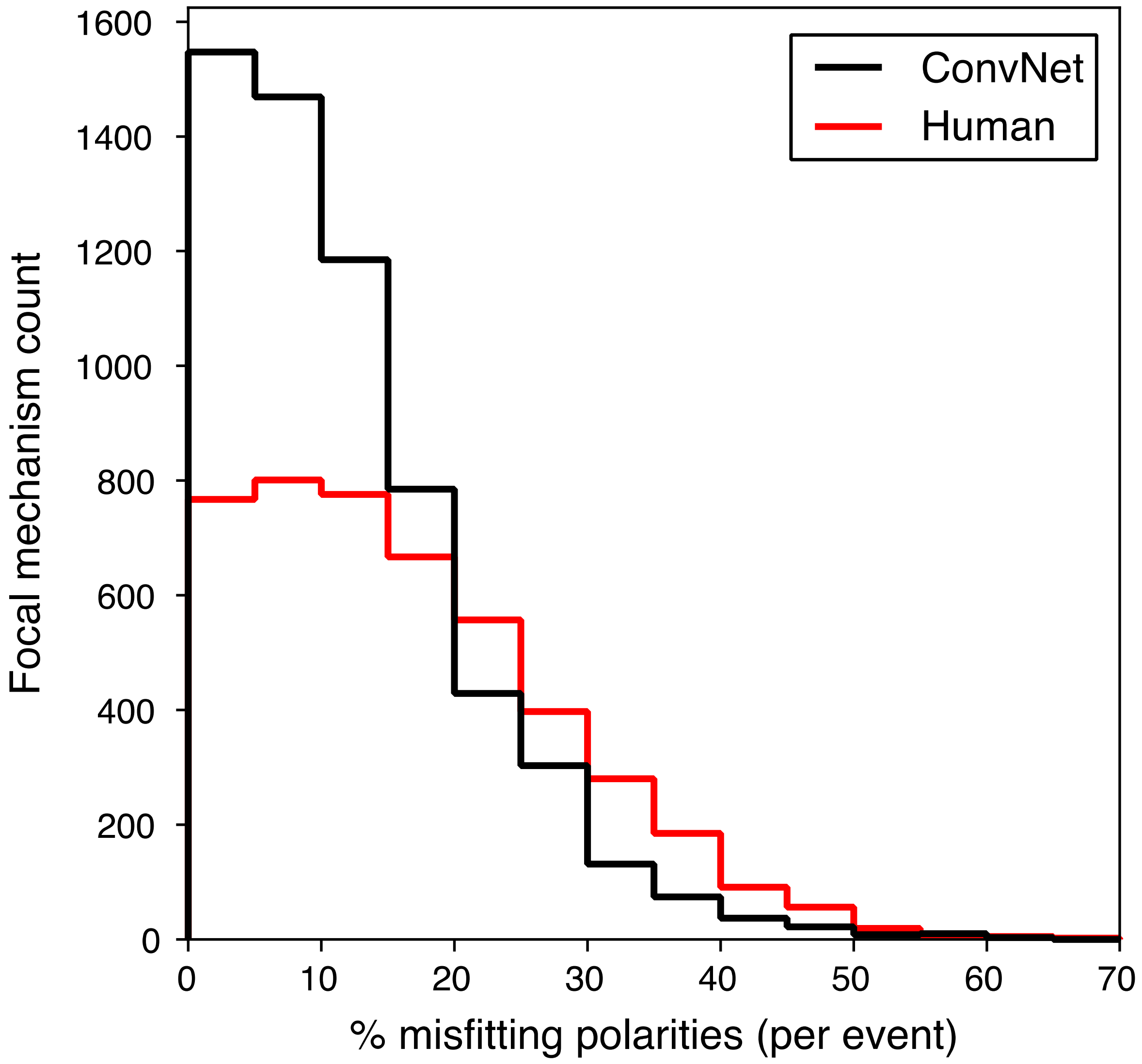
85
80
75
70
65
60
55
64
66
68
70
72
74
76
78

Focal mechanism determination with deep learning



Hardebeck and Shearer (2002)

Ross et al (2018), JGR



- We used a labeled dataset of 4.8 million seismograms to train a deep convolutional network to measure P-wave first-motion polarities
- ConvNet significantly outperforms human experts
- Results in many more focal mechanisms, with lower misfit per event

Looking forward

- Seismology has quickly become a leader in applying cutting edge ML algorithms
- The inability to build earthquake catalogs for arbitrary datasets has always been a barrier to downstream science
  - ML is providing end-to-end solutions that are making these disappear
  - Will transform experimental seismology
- Automated measurements are as good human ground truth
  - Expanded catalogs
  - Better locations
  - Improved tomography results
  - Better estimates of earthquake source properties