

Earthquake Catalog: Quality Control (QC)

- I created a custom (deep-learning) enhanced earthquake catalog automatically. How good is it?
 - No ground truth available
- Before using or interpreting your enhanced catalog, do this first:
 - Compare with a **reference catalog**
 - **Data visualization** (make plots!) to check for quality

What is a **reference catalog**?

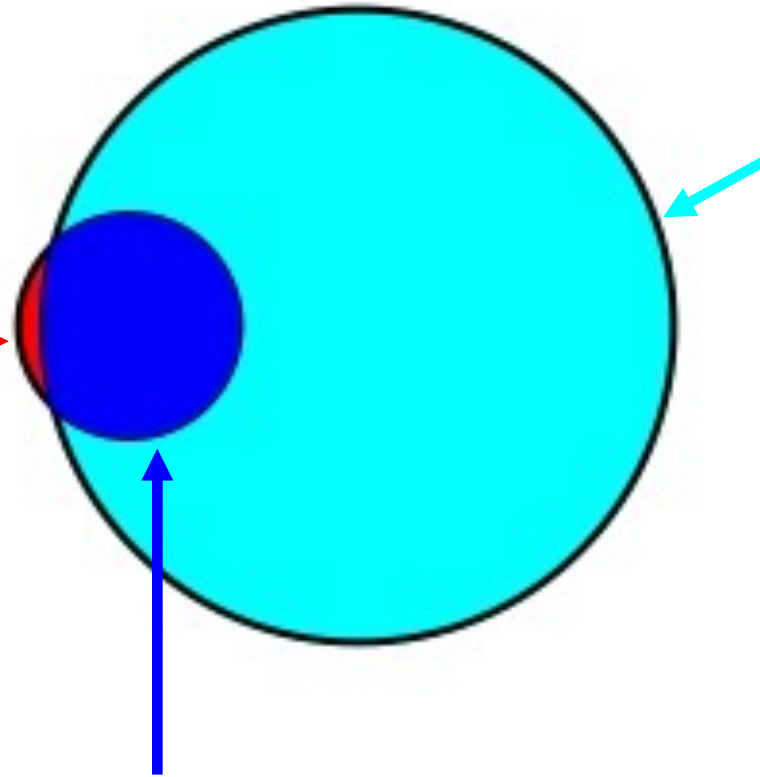
- *Authoritative, high-quality* earthquake catalog for a region
 - Good baseline for comparison: manually reviewed, meets performance standards
- **United States:** Comprehensive Catalog (**ComCat**), complete to M2.5-3.0, <https://earthquake.usgs.gov/earthquakes/search/> , <https://earthquake.usgs.gov/data/comcat/>
- **International:** ComCat is complete to M4.5 globally. For more complete catalogs, refer to the country/region's authoritative monitoring website.
 - Turkey: <https://deprem.afad.gov.tr/event-catalog>
 - New Zealand: <https://quakesearch.geonet.org.nz/>
 - Italy: <https://terremoti.ingv.it/bsi>

Event comparison: reference catalog vs. enhanced catalog

(Need locations) (small circle) (big circle)

MISSED events (only in reference catalog)

- *Manually detected and/or picked by analysts*
- *Check why they were missed (esp. larger events?)*



MATCH events (common to both catalogs)

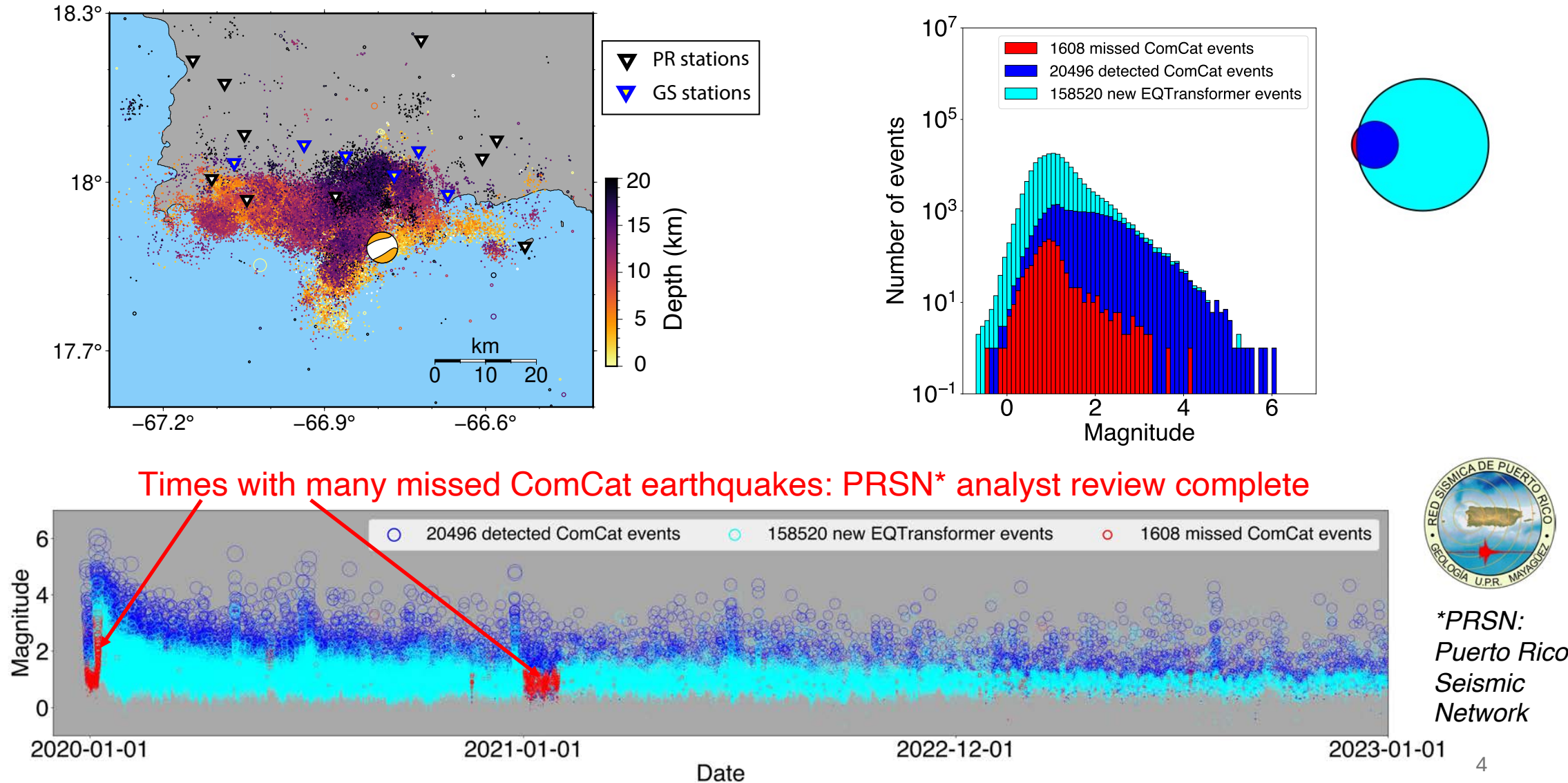
- *Origin times match within $X \sim 5$ seconds*
- *Hypocenters match within $Y \sim 25$ km*

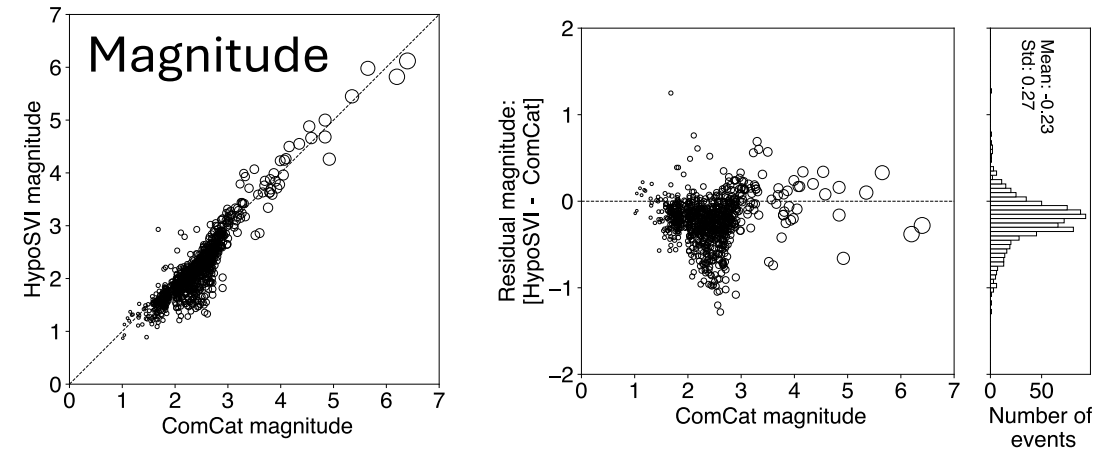
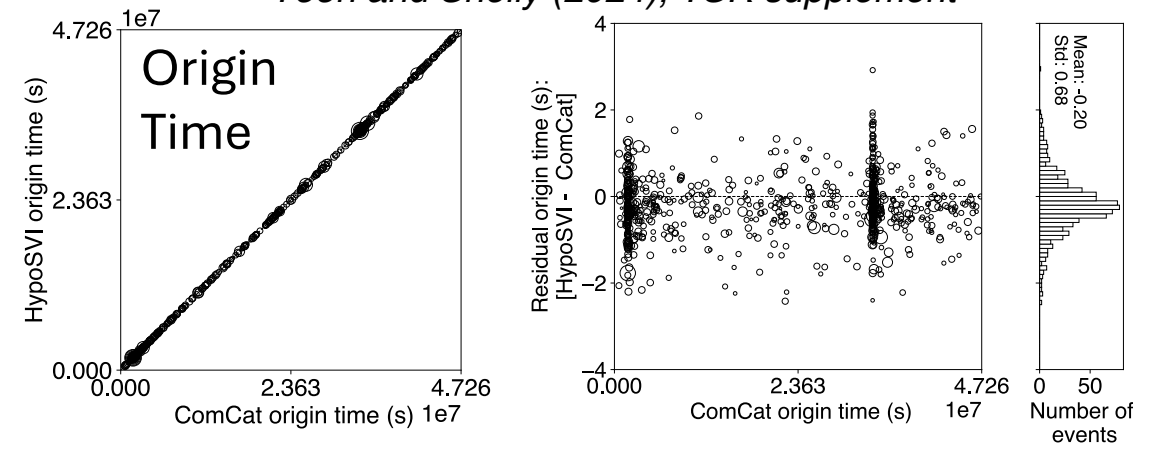
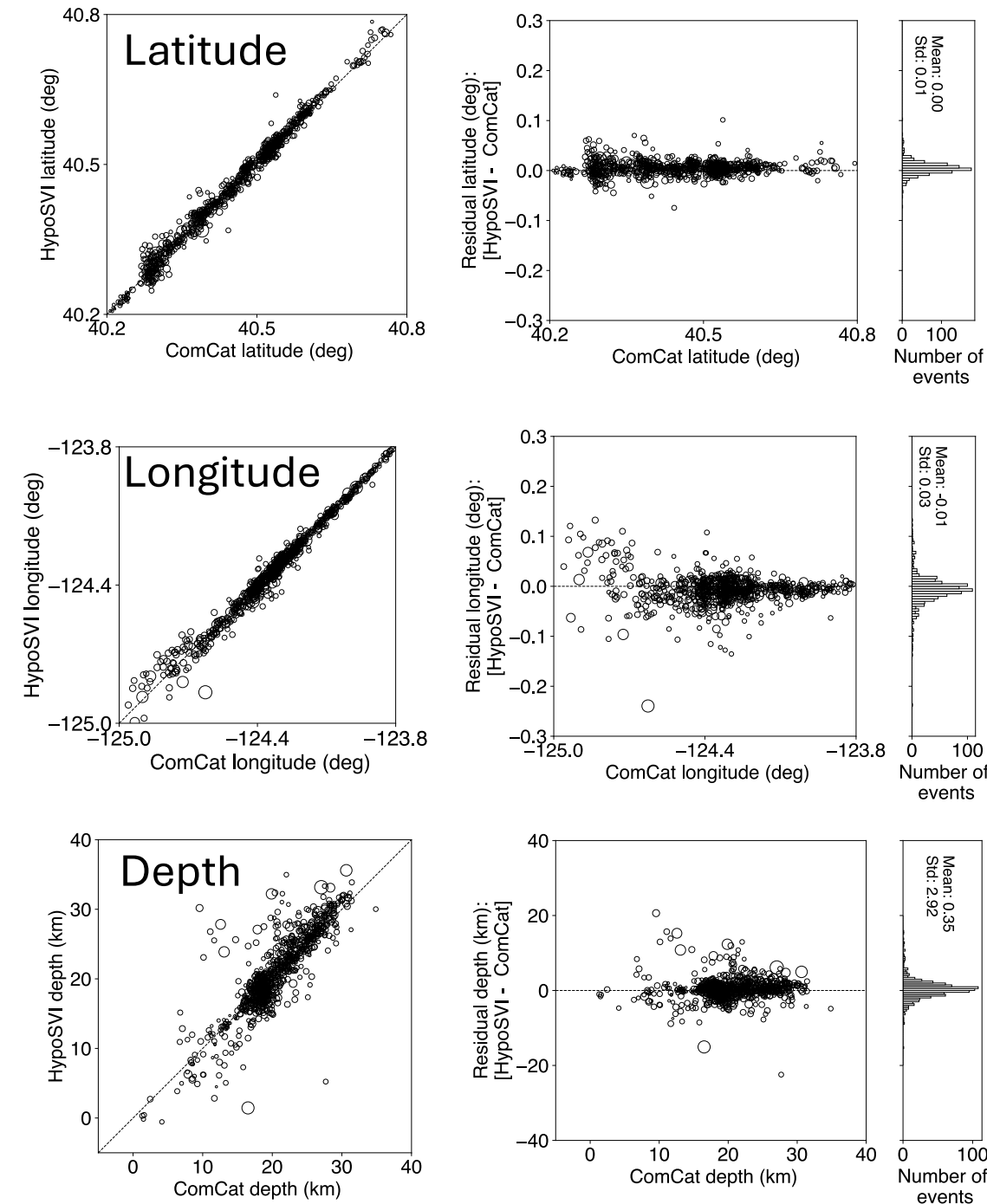
NEW events (only in enhanced catalog)

- *Newly detected small local earthquakes? 😊*
- *False detections from noise? 😞*
- *Other signals (quarry blasts, sonic booms, explosions)? 🤔*
- *Regional or teleseismic earthquakes? 🤔*

Event comparison example: SW Puerto Rico sequence

Yoon et al. (2023), BSSA





MATCH events: location, magnitude comparison

Horizontal axis: ComCat (reference catalog)
Vertical axis: Enhanced catalog

MATCH events: Pick time comparison for each station

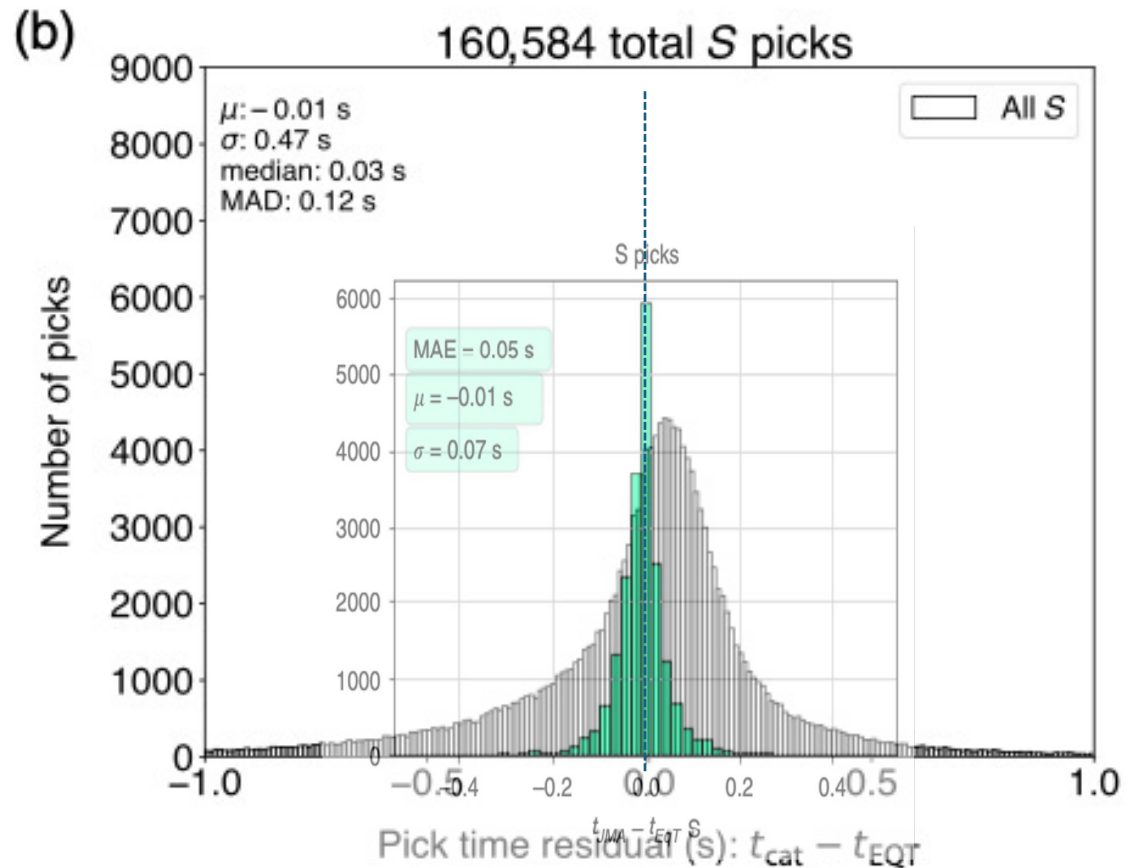
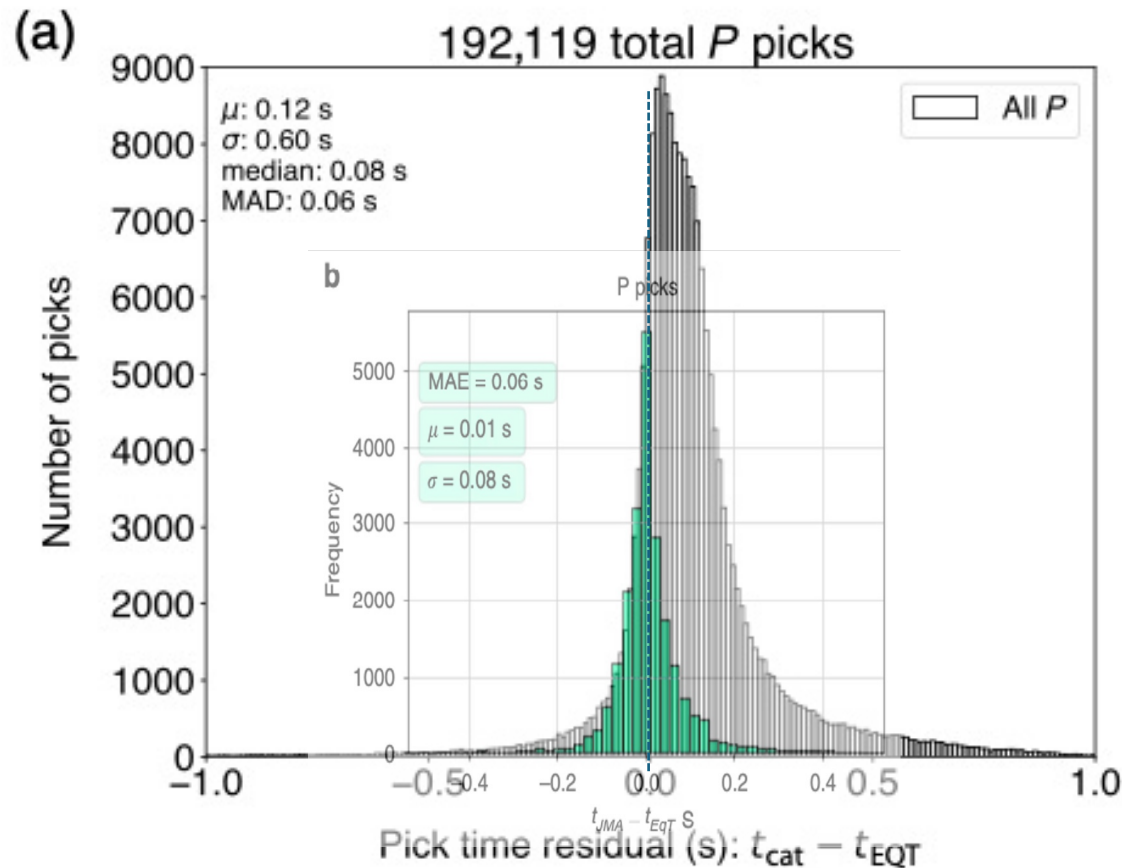
→ residual histograms

EQTransformer on Puerto Rico seismic data:

less accurate pick times than

Mousavi et al. (2020) on STEAD test dataset

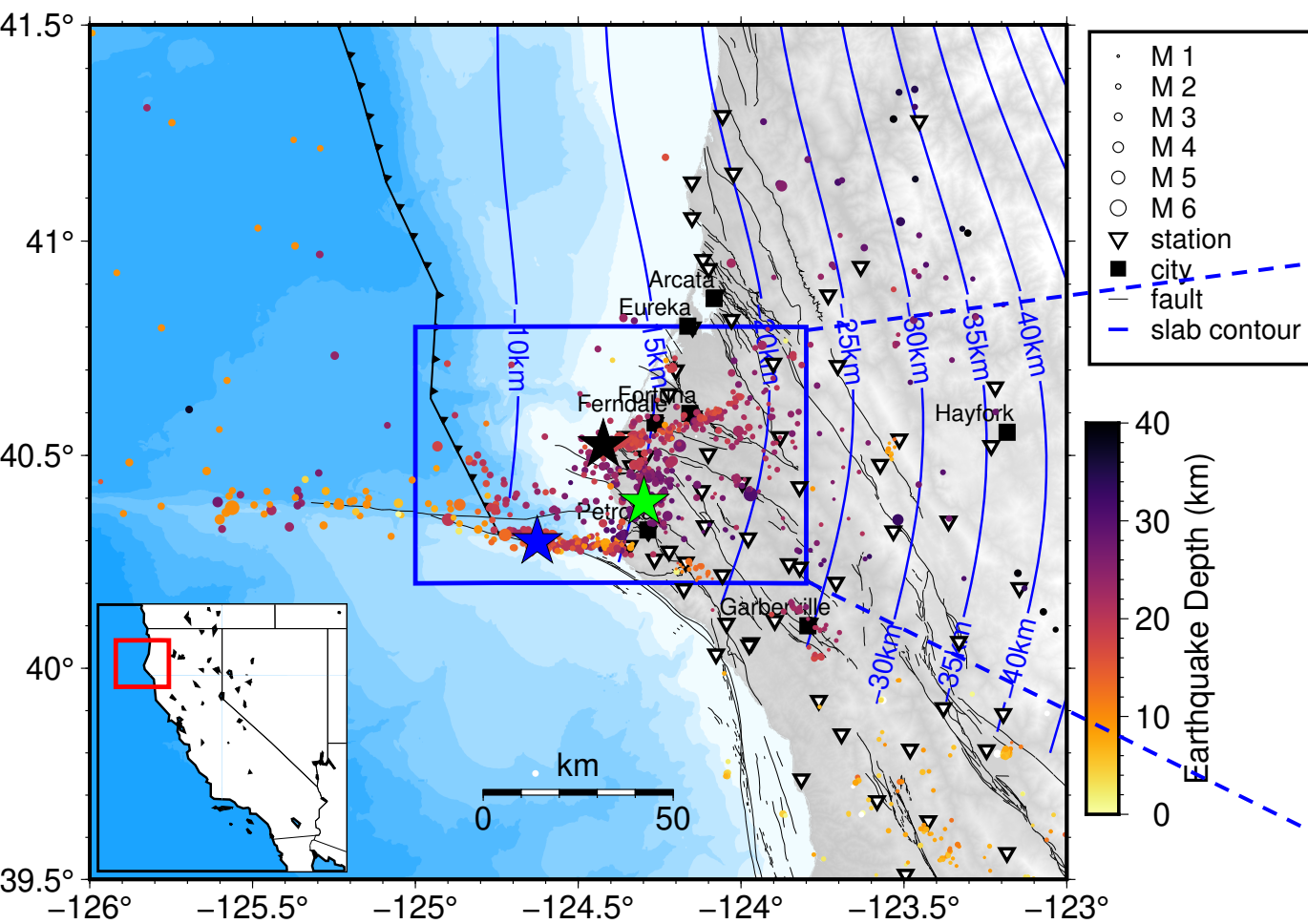
... but **good enough** for automatic catalog workflow



Reference Catalog Tip: Select boundaries carefully

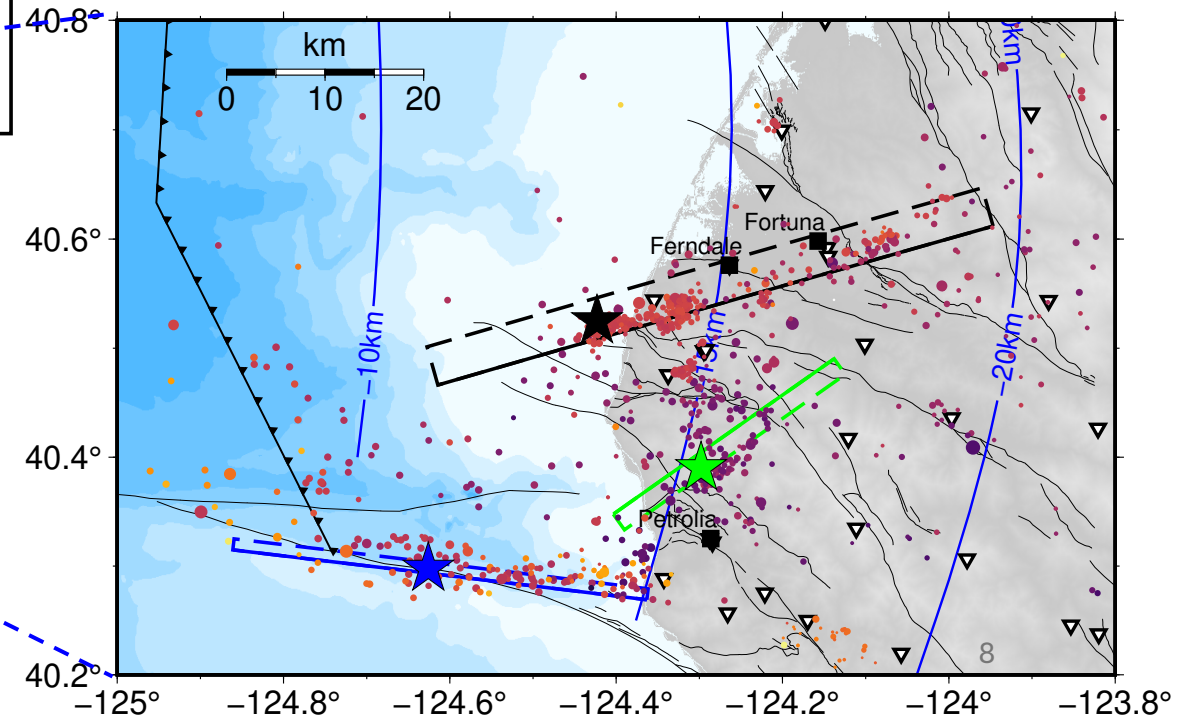
Example: Mendocino Triple Junction, CA

Region big enough: includes most seismicity and all stations used



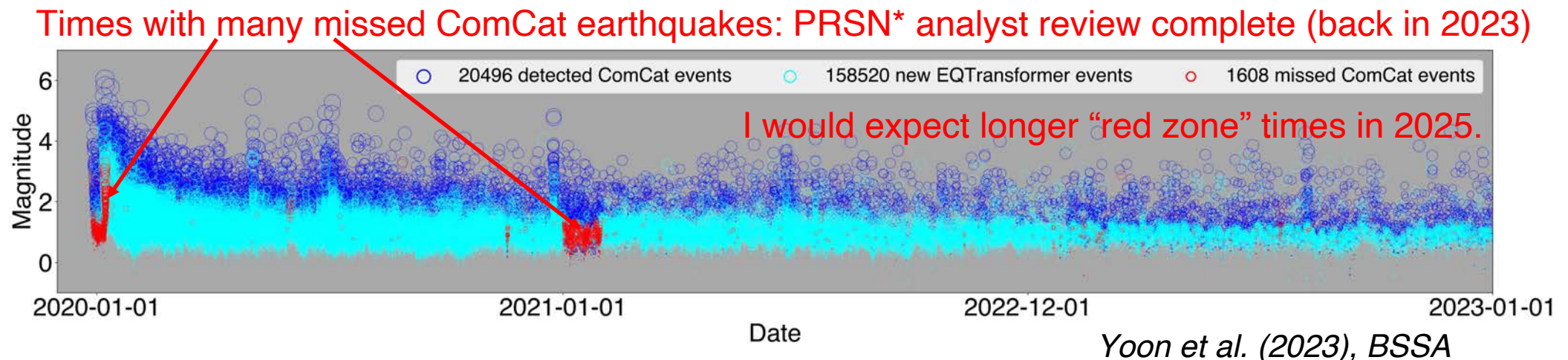
Yoon and Shelly (2024), TSR

Region too small: missing some regions with seismicity; not all stations



Reference Catalog Tip: It changes over time

- Reference catalog changes over time due to analyst review
 - New or deleted events; updated location/magnitude for existing events
 - Expect frequent changes to catalog: **hours to days** after large earthquake
 - Complete analyst review for active aftershock sequence: **months to years**
 - July 2019 M7.1 Ridgecrest CA sequence
 - 2019-07-07: 45% complete, 2019-07-08 to 2019-07-31: <10% complete each day
 - Catalog processing status (updated daily):
https://service.scedc.caltech.edu/ftp/catalogs/catalog_status/



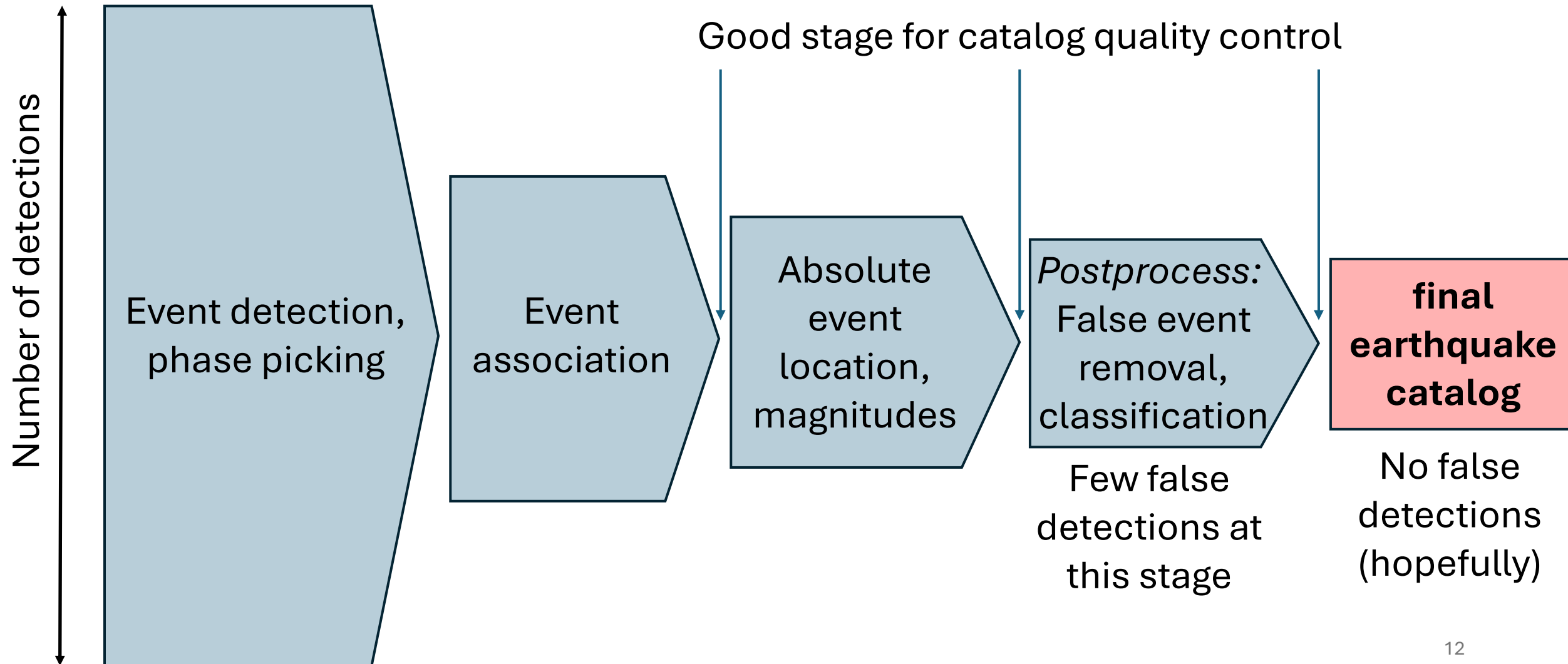
Reference Catalog Tip: It changes over time

- **Save the reference catalog file itself, with filename including the date you downloaded it.**
 - When reference catalog changes in the future, you still have the original version, so you can easily reproduce your results.

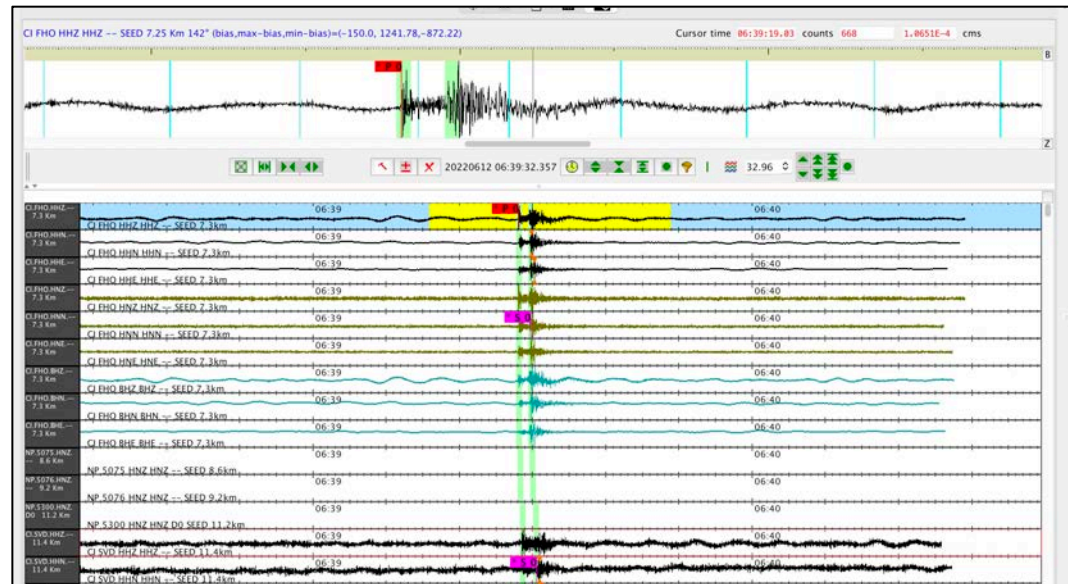
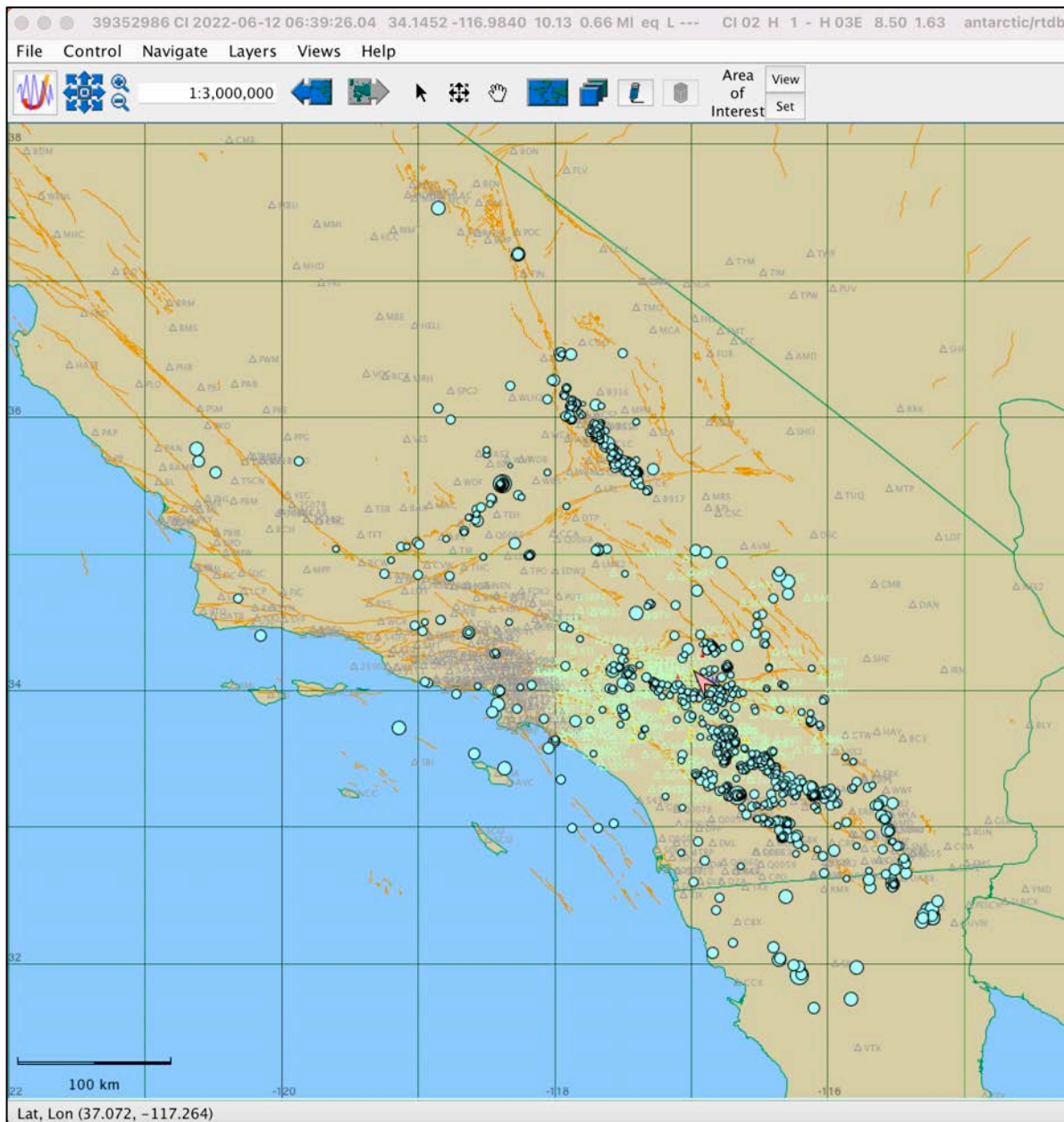
Earthquake Catalog: Quality Control (QC)

- I created a custom (deep-learning) enhanced earthquake catalog automatically. How good is it?
 - No ground truth available
- Before using or interpreting your enhanced catalog, do this first:
 - Compare with a reference catalog
 - **Data visualization** (make plots!) to check for quality

Catalog QC: Most false detections drop out at each step of automatic workflow, but not all.



Earthquake Catalog QC: Think like an analyst



AQMS Jiggle manual review and picking interface at SCSN

ID	ONMO	DATETIME	LAT	LONG	MZ	Z	MG	MTYP	AUTH	ETTYPE	QT	SRC	ESRC	QAT	DIST	RHS	EMR	EMR2	EMR3	EMR4	EMR5	LOBS	USED	Q	ST	PR	V	LOCATE	QAUTH	CORRECT
39352930		2002-06-12 05:24:00.880	33.74	-116.057	2.93	2.46	9.95	M1		earthquake	L	RTG	R	136.6	9.00	0.24	0.45	0.91	27	27	0.0	0.0	25.1	2002-06-12 05:20:41.41						
39352940		2002-06-12 05:58:19.320	33.991	-116.872	5.89	4.79	1.91	M1		earthquake	L	RTG	R	102.0	4.00	0.12	0.27	0.63	33	33	0.0	0.0	26.3	2002-06-12 06:03:05.01						
39352950		2002-06-12 06:01:18.720	33.288	-116.942	1.91	1.87	0.62	M1		earthquake	L	RTG	R	107.0	4.00	0.11	0.24	0.59	36	36	0.0	0.0	26.3	2002-06-12 06:03:11.01						
39352962		2002-06-12 06:25:06.730	33.661	-116.931	15.4	15.79	8.92	M1		earthquake	L	RTG	R	59.0	7.00	0.17	0.43	0.82	36	36	0.0	0.0	25.8	2002-06-12 06:29:45.41						
39352970		2002-06-12 06:34:02.930	33.239	-116.008	11.36	10.75	1.33	M1		earthquake	L	RTG	R	161.0	15.00	0.19	0.51	1.04	26	23	0.0	0.0	26.0	2002-06-12 06:30:36.36						
39352980		2002-06-12 06:35:00.160	33.239	-116.008	11.36	10.75	1.33	M1		earthquake	L	RTG	R	161.0	15.00	0.19	0.51	1.04	26	23	0.0	0.0	26.0	2002-06-12 06:30:36.36						
39352994		2002-06-12 07:05:15.190	33.286	-116.103	11.82	10.63	1.14	M1		earthquake	L	RTG	R	95.0	15.00	0.23	0.48	1.04	37	35	0.0	0.0	26.1	2002-06-12 07:09:58.58						
39353002		2002-06-12 08:24:17.870	32.121	-116.396	19.94	18.63	1.05	M1		earthquake	L	RTG	R	274.0	33.00	0.18	0.78	0.76	19	17	0.0	0.0	29.6	2002-06-12 08:20:09.41						
39353010		2002-06-12 09:08:01.440	33.122	-116.908	9.39	9.03	1.33	M1		earthquake	L	RTG	R	96.0	15.00	0.21	0.52	2.03	27	22	0.0	0.0	26.0	2002-06-12 09:13:17.11						
39353020		2002-06-12 09:24:13.320	33.239	-116.008	11.36	10.75	1.33	M1		earthquake	L	RTG	R	161.0	15.00	0.19	0.51	1.04	26	23	0.0	0.0	26.0	2002-06-12 09:13:17.11						
39353034		2002-06-12 09:40:38.300	33.232	-116.316	10.71	9.93	8.81	M1		earthquake	L	RTG	R	64.0	11.00	0.24	0.48	0.54	48	44	0.0	0.0	24.4	2002-06-12 09:45:31.41						
39353050		2002-06-12 09:55:38.990	33.232	-116.316	10.87	9.67	1.21	M1		earthquake	L	RTG	R	96.0	15.00	0.22	0.45	1.39	37	35	0.0	0.0	27.5	2002-06-12 10:00:08.71						
39353058		2002-06-12 10:01:05.960	33.456	-116.362	4.80	2.89	1.20	M1		earthquake	L	RTG	R	42.0	10.00	0.20	0.28	1.28	57	54	0.0	0.0	26.8	2002-06-12 10:05:52.51						
39353068		2002-06-12 10:40:48.530	33.495	-117.176	16.85	15.30	1.01	M1		earthquake	L	RTG	R	137.0	12.00	0.13	0.49	1.12	38	28	0.0	0.0	24.6	2002-06-12 10:41:00.01						
39353090		2002-06-12 10:35:12.760	35.589	-118.395	8.	7.95	1.23	M1		earthquake	L	RTG	R	102.0	10.00	0.15	0.36	1.17	35	33	0.0	0.0	27.8	2002-06-12 10:39:41.41						
39353100		2002-06-12 10:56:44.810	33.907	-117.876	16.21	15.52	1.22	M1		earthquake	L	RTG	R	57.0	12.00	0.20	0.26	0.88	59	55	0.0	0.0	26.5	2002-06-12 11:01:00.11						
39353122		2002-06-12 11:27:16.330	35.945	-116.930	15.99	14.68	0.94	M1		earthquake	L	RTG	R	66.0	11.00	0.17	0.27	0.58	49	47	0.0	0.0	25.1	2002-06-12 11:33:12.11						
39353128		2002-06-12 11:37:37.820	33.223	-116.316	10.71	9.93	8.81	M1		earthquake	L	RTG	R	64.0	11.00	0.24	0.48	0.54	48	47	0.0	0.0	25.0	2002-06-12 11:33:12.11						
39353138		2002-06-12 12:00:48.150	33.215	-115.969	11.70	11.67	1.87	M1		earthquake	L	RTG	R	33.0	18.00	0.29	0.30	0.93	84	64	0.0	0.0	29.8	2002-06-12 12:13:15.11						
39353162		2002-06-12 12:11:28.640	32.895	-115.538	1.51	5.11	1.05	M1		earthquake	L	RTG	R	171.0	3.00	0.10	1.59	0.68	18	10	0.0	0.0	25.7	2002-06-12 12:15:27.11						
39353165		2002-06-12 12:15:48.490	33.232	-116.316	10.87	9.67	1.21	M1		earthquake	L	RTG	R	96.0	15.00	0.22	0.45	1.39	37	35	0.0	0.0	27.5	2002-06-12 12:15:27.11						
39353186		2002-06-12 13:02:27.320	35.173	-118.611	6.12	7.35	1.57	M1		earthquake	L	RTG	R	68.0	7.00	0.18	0.25	0.85	51	49	0.0	0.0	28.3	2002-06-12 13:07:05.11						
39353210		2002-06-12 13:11:57.250	33.433	-117.875	10.48	9.94	1.22	M1		earthquake	L	RTG	R	58.0	18.00	0.16	0.35	1.15	28	28	0.0	0.0	26.5	2002-06-12 13:16:23.11						
39353218		2002-06-12 13:25:23.440	36.449	-117.965	4.25	2.66	1.52	M1		earthquake	L	RTG	R	162.0	16.00	0.23	1.42	1.01	18	16	0.0	0.0	28.8	2002-06-12 13:30:27.11						
39353220		2002-06-12 13:40:27.680	33.298	-116.377	10.48	13.04	0.91	M1		earthquake	L	RTG	R	100.0	10.00	0.22	0.45	1.06	33	32	0.0	0.0	25.9	2002-06-12 13:53:52.51						
39353234		2002-06-12 14:51:48.760	31.925	-116.195	23.43	22.94	1.99	M1		earthquake	L	RTG	R	235.0	45.00	0.17	1.05	2.31	30	15	0.0	0.0	30.4	2002-06-12 14:57:22.71						
39353250		2002-06-12 15:45:17.990	33.267	-116.793	10.39	12.67	1.46	M1		earthquake	L	RTG	R	56.0	5.00	0.19	0.32	0.71	40	39	0.0	0.0	27.7	2002-06-12 15:49:58.11						
39353266		2002-06-12 16:12:07.290	33.962	-117.214	13.22	12.69	8.77	M1		earthquake	L	RTG	R	206.0	11.00	0.12	0.86	1.24	14	14	0.0	0.0	24.3	2002-06-12 16:16:07.11						
39353280		2002-06-12 16:38:52.360	33.174	-116.952	17.72	17.22	1.36	M1		earthquake	L	RTG	R	107.0	10.00	0.11	0.24	0.59	36	36	0.0	0.0	26.3	2002-06-12 16:43:11.01						
39353290		2002-06-12 16:24:47.800	34.638	-117.255	15.55	14.88	1.23	M1		earthquake	L	RTG	R	127.0	12.00	0.20	0.62	0.87	26	24	0.0	0.0	26.6	2002-06-12 16:29:13.11						
39353298		2002-06-12 16:20:33.640	35.028	-116.907	5.09	4.13	1.24	M1		earthquake	L	RTG	R	58.0	17.00	0.17	0.16	0.67	103	44	0.0	0.0	31.2	2002-06-12 16:33:53.11						
39353314		2002-06-12 16:33:47.370	34.664	-117.520	14.68	13.95	1.34	M1		earthquake	L	RTG	R	105.0	10.00	0.17	0.45	1.10	28	21	0.0	0.0	27.7	2002-06-12 16:38:33.11						
39353326		2002-06-12 16:38:26.690	34.269	-117.381	16.85	15.30	1.01	M1		earthquake	L	RTG	R	107.0	10.00	0.17	0.45	1.10	28	21	0.0	0.0	27.7	2002-06-12 16:43:11.01						
39353338		2002-06-12 16:38:36.760	35.028	-117.261	16.68	15.25	1.25	M1		earthquake	L	RTG	R	116.0	11.00	0.15	0.69	0.91	22	21	0.0	0.0	26.1	2002-06-12 16:45:16.11						
39353346		2002-06-12 16:40:51.780	34.277	-118.437	9.19	8.61	1.04	M1		earthquake	L	RTG	R	134.0	4.00	0.16	0.63	0.93	19	17	0.0	0.0	25.6	2002-06-12 16:45:25.11						
39353366		2002-06-12 16:56:48.950	33.147	-116.952	17.72	17.22	1.36	M1		earthquake	L	RTG	R	107.0	10.00	0.11	0.24	0.59	36	36	0.0	0.0	26.3	2002-06-12 16:51:11.01						
39353380		2002-06-12 17:58:53.590	35.523	-118.308	7.48	6.44	8.54	M1		earthquake	L	RTG	R	118.0	17.00	0.13	0.47	1.40	21	18	0.0	0.0	25.2	2002-06-12 18:03:19.11						
39353402		2002-06-12 18:20:57.160	35.597	-128.044	28.38	27.81	1.78	M1		earthquake	L	RTG	R	264.0	27.00	0.26	0.74	1.96	12	11	0.0	0.0	29.4	2002-06-12 18:33:34.11						
39353420		2002-06-12 19:36:38.680	36.151	-117.929	3.72	2.32	1.53	M1		earthquake	L	RTG	R	108.0	9.00	0.32	0.64	1.12	22	21	0.0	0.0	28.1	2002-06-12 19:41:14.11						
39353430		2002-06-12 19:38:18.160	36.151	-117.929	3.72	2.32	1.53	M1		earthquake	L	RTG	R	108.0	9.00	0.32	0.64	1.12	22	21	0.0	0.0	28.1	2002-06-12 19:41:14.11						

Earthquake Catalog QC Tips: Check Parameters

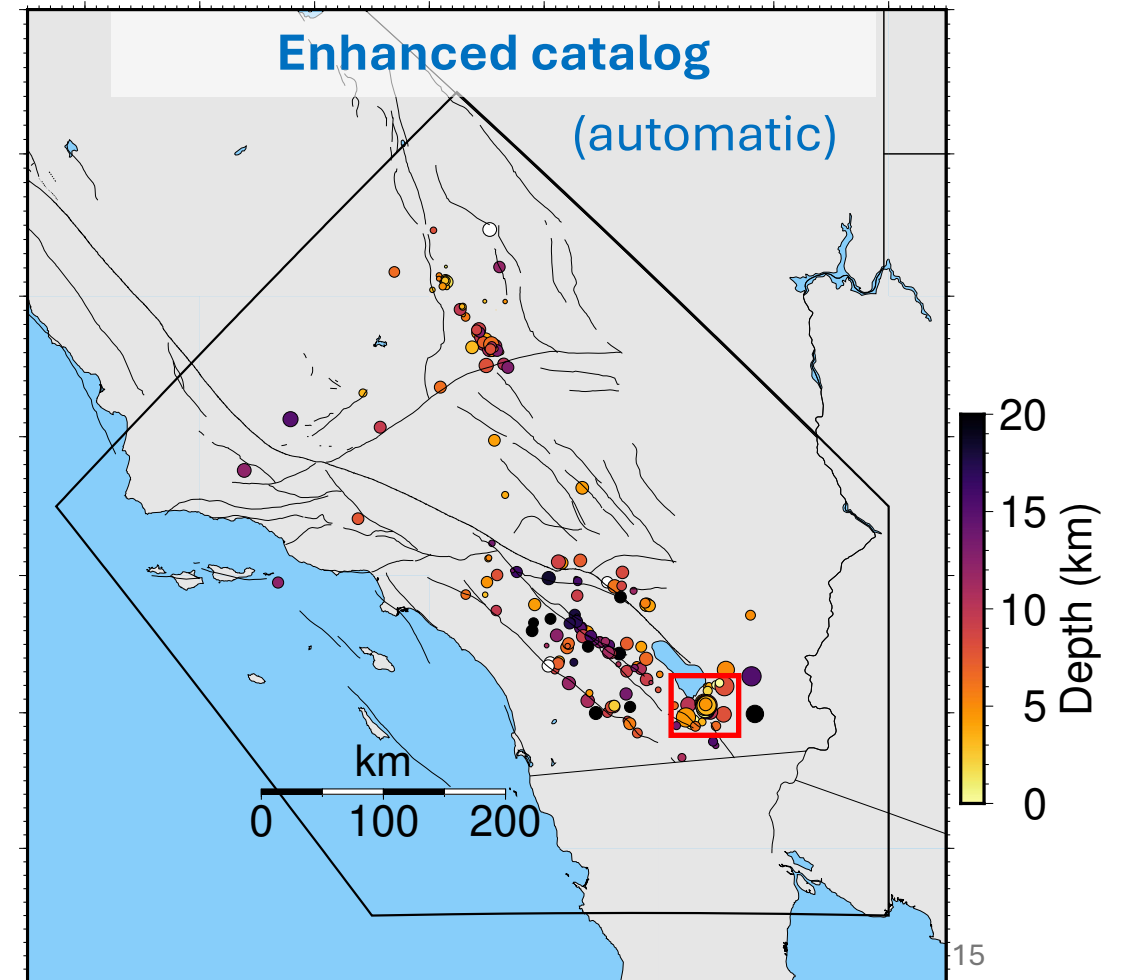
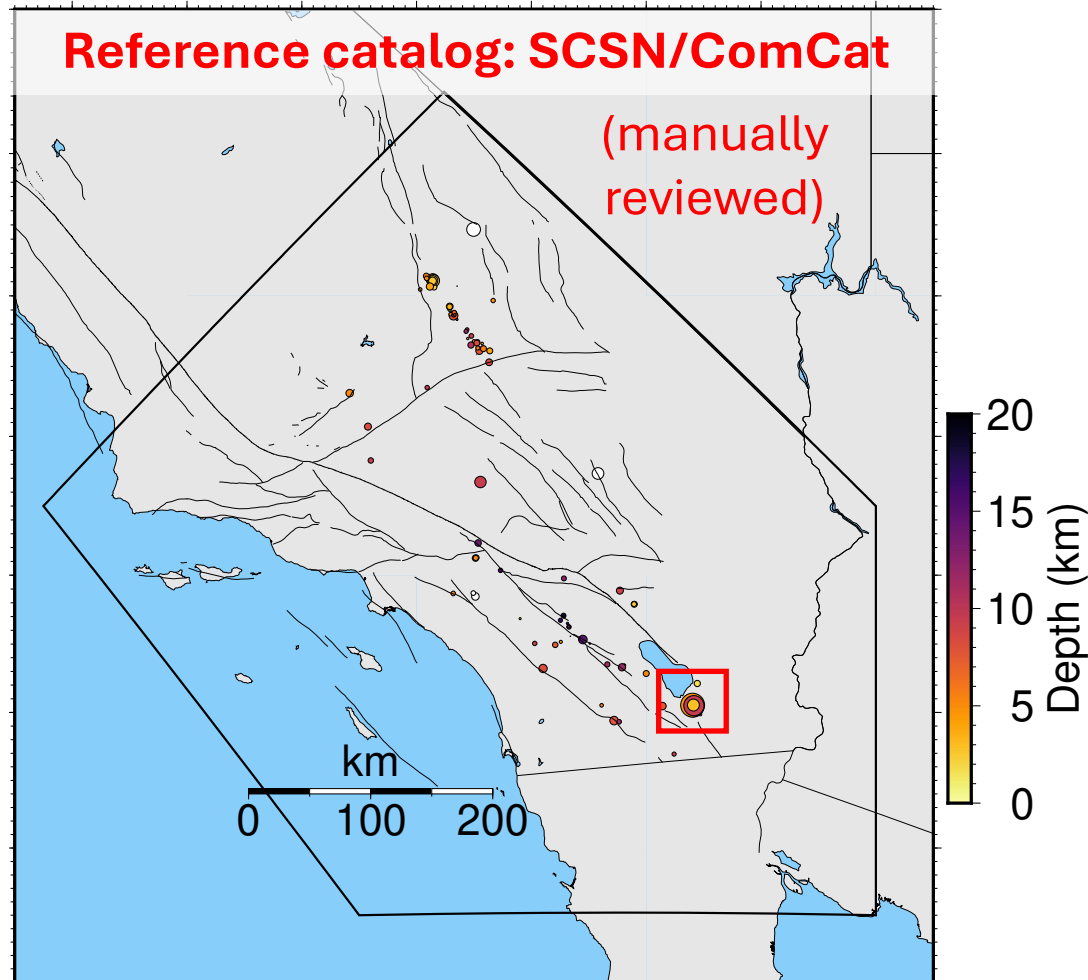
- Compare enhanced catalog with reference catalog on plots
 - Plot reference & enhanced catalogs side-by-side
 - Separate plots for MATCH, NEW, MISSED category events
- Visualize enhanced catalog → are results reasonable for earthquakes?
 - Locations (epicenters) in map view
 - Are events near known seismogenic areas, past seismicity, faults? Known quarries?
 - Depth cross-sections
 - Are events at expected seismogenic depths (0-30 km, unless in subduction zone)?
 - Magnitude-frequency distribution
 - Do event magnitudes follow Gutenberg-Richter distribution (lots of small events)?
 - Magnitude vs. time, Seismicity rate (number of events vs. time)
 - If aftershock sequence, does number of events decrease as $1/\text{time}$ after mainshock (Omori decay)?

Catalog QC, **locations in map view**: SCSN



2020-09-30 00:00:00 UTC to 2020-10-01 00:00:00 UTC (24 hours)

Enhanced catalog: events are on known faults with active seismicity

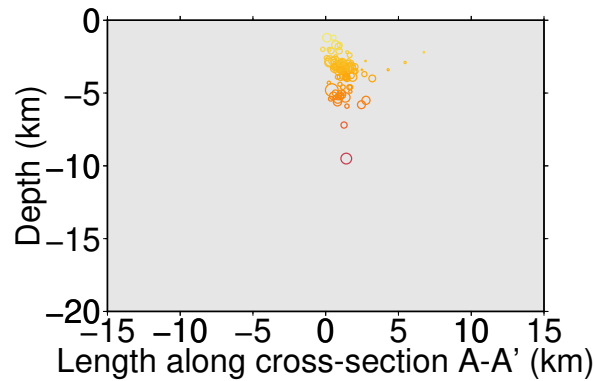
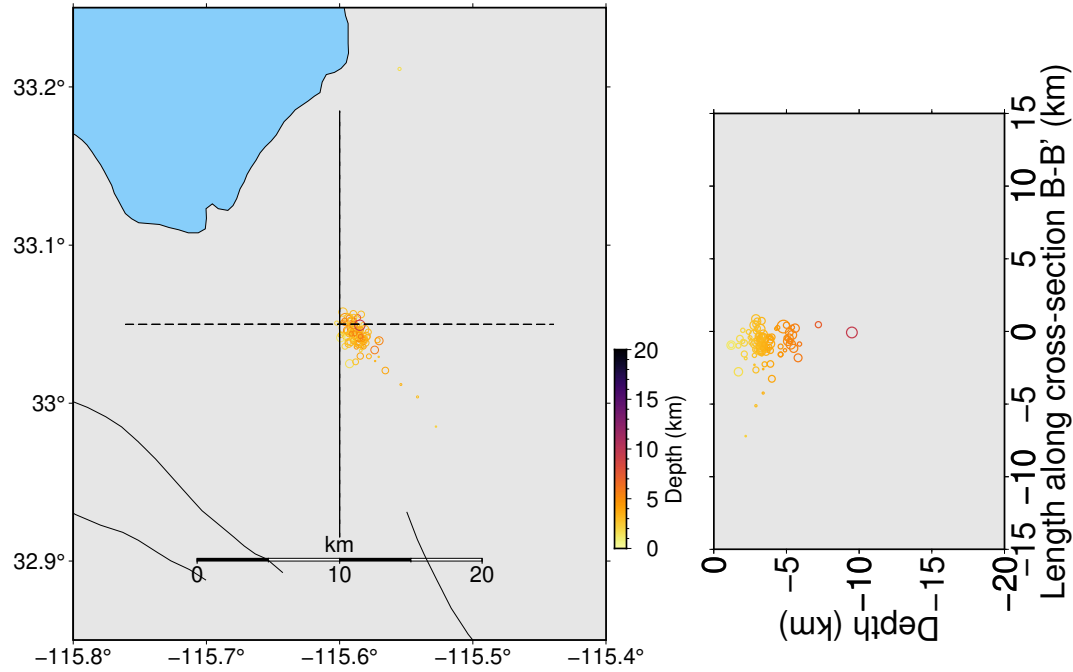


Catalog QC, zoomed **locations in map view & depth cross-section: SCSN**

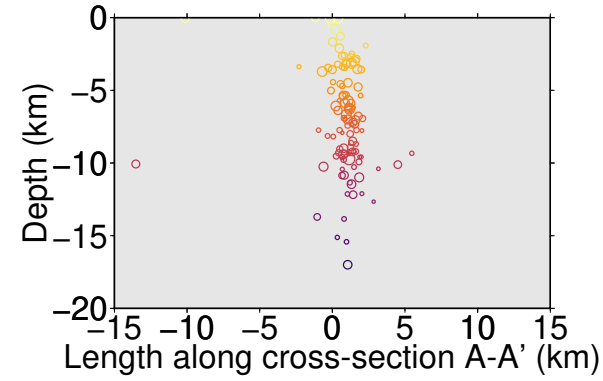
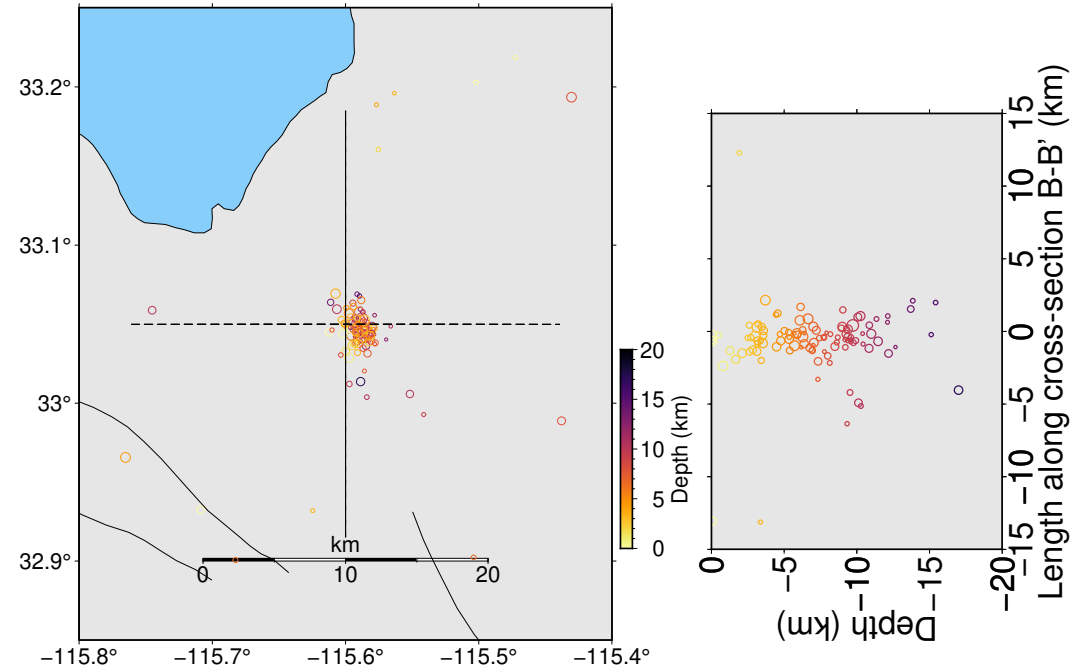


2020-09-30 00:00:00 UTC to 2020-10-01 00:00:00 UTC (24 hours)

Enhanced catalog: Similar epicenters. Larger range of depths, but still reasonable



**Reference catalog:
SCSN/ComCat
(manually reviewed)**



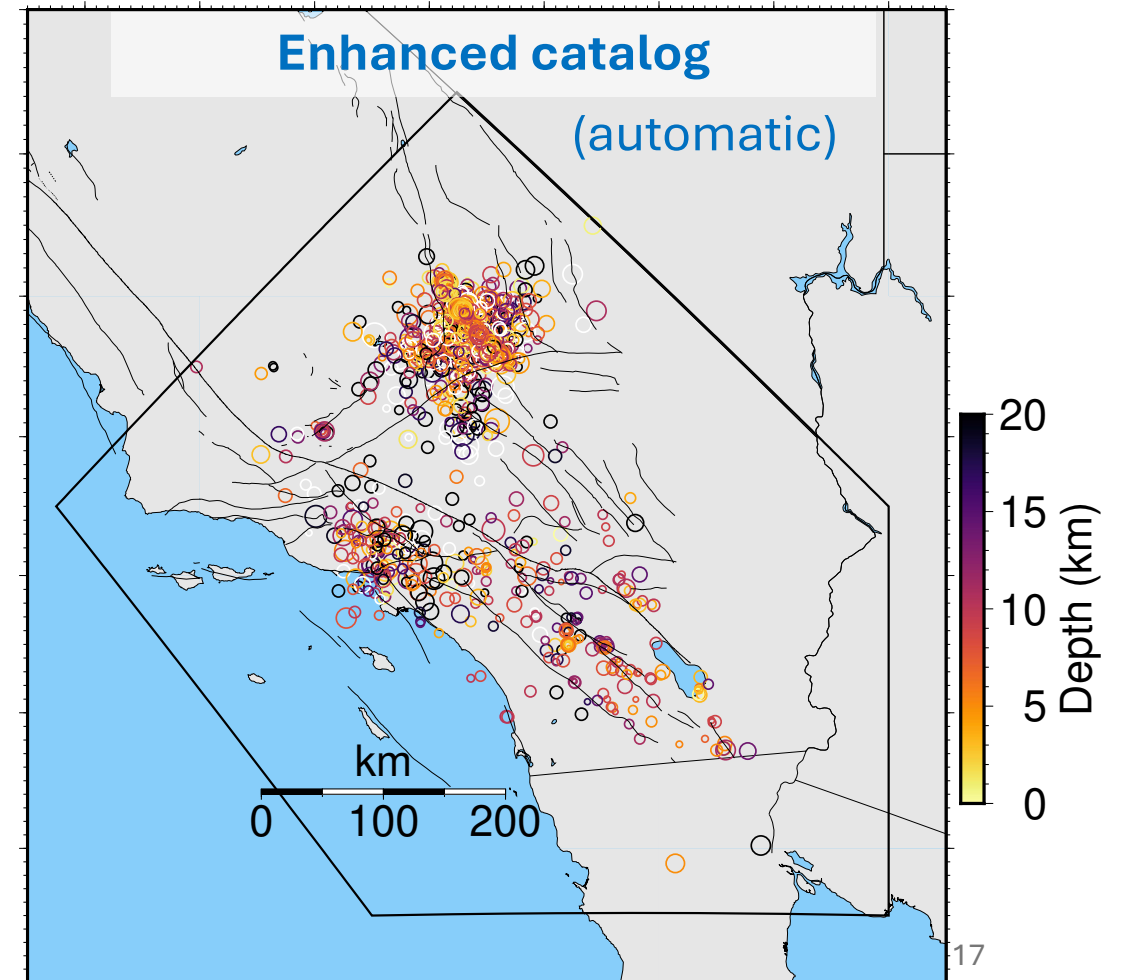
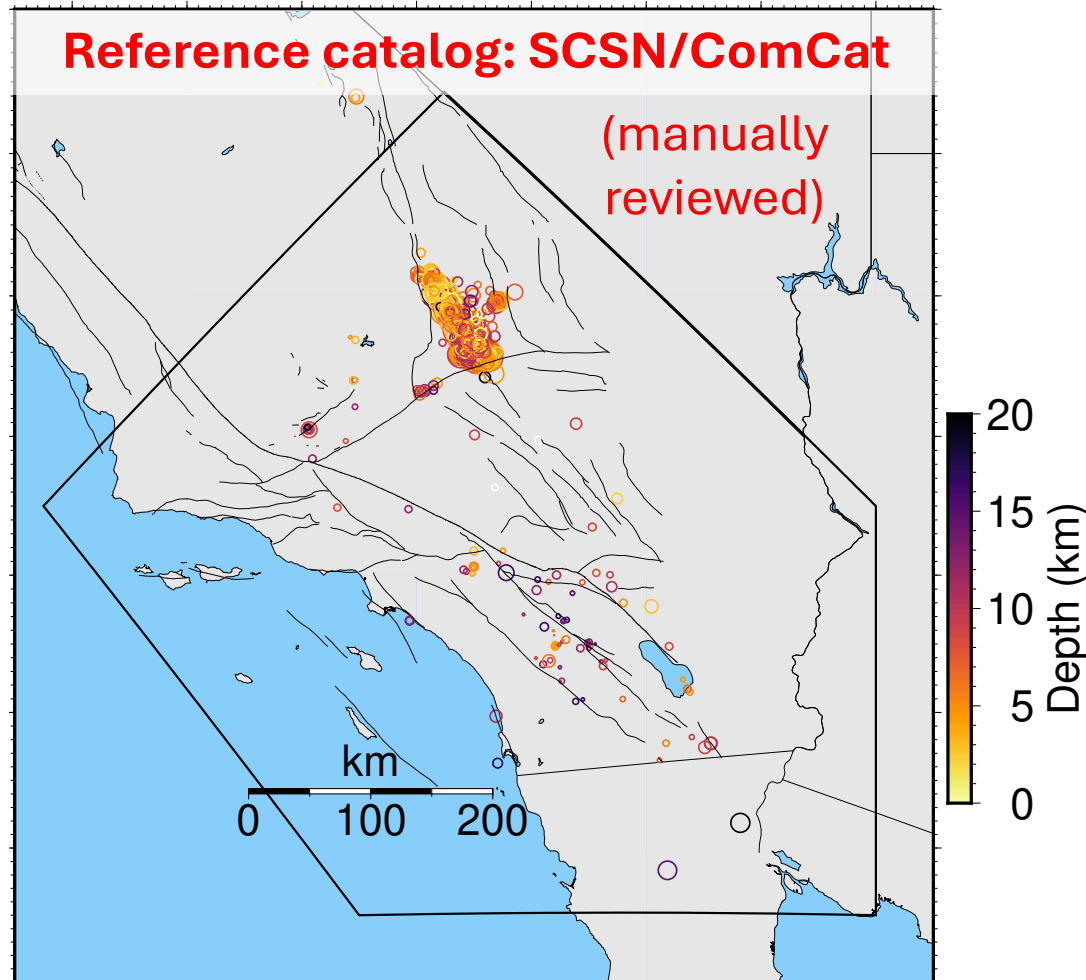
**Enhanced catalog
(automatic)**

Catalog QC, **locations in map view**: SCSN



2019-07-04 00:00:00 UTC to 2019-07-07 01:14:00 UTC (73 hours)

Ridgecrest first 3 days (M_w 6.4 & 7.1, aftershocks), enhanced catalog:
association errors → false detections with bad locations/magnitudes



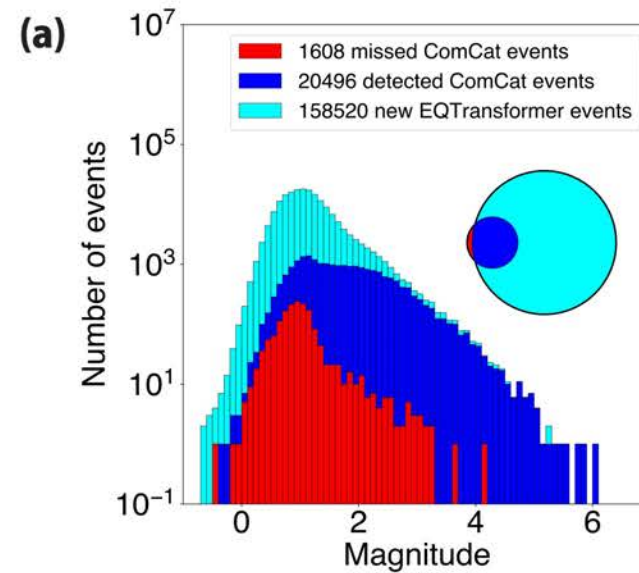
Earthquake Catalog QC Tips: Check Parameters

- Compare enhanced catalog with reference catalog on plots
 - Plot reference & enhanced catalogs side-by-side
 - Separate plots for MATCH, NEW, MISSED category events
- Visualize enhanced catalog → are results reasonable for earthquakes?
 - Locations (epicenters) in map view
 - Are events near known seismogenic areas, past seismicity, faults? Known quarries?
 - Depth cross-sections
 - Are events at expected seismogenic depths (0-30 km, unless in subduction zone)?
 - Magnitude-frequency distribution
 - Do event magnitudes follow Gutenberg-Richter distribution (lots of small events)?
 - Magnitude vs. time, Seismicity rate (number of events vs. time)
 - If aftershock sequence, does number of events decrease as $1/\text{time}$ after mainshock (Omori decay)?

Catalog QC: Puerto Rico

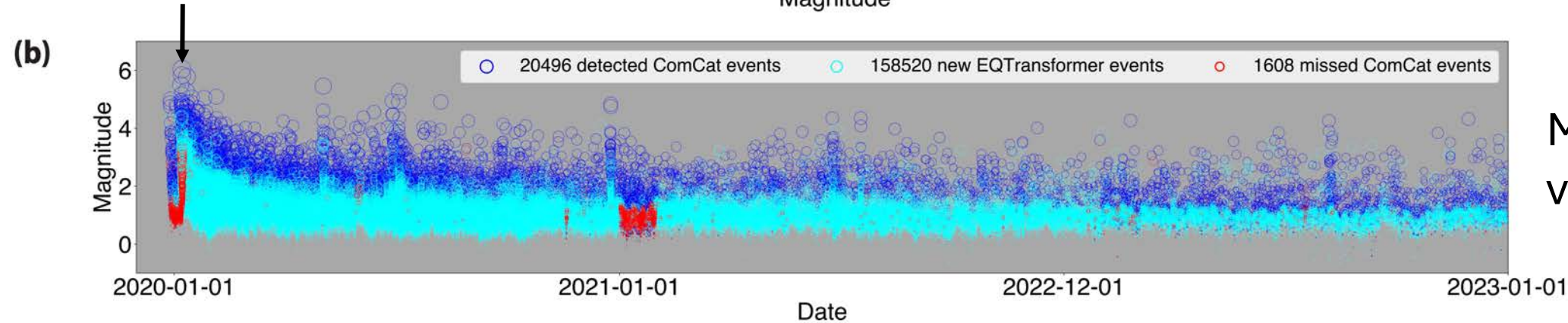


MATCH,
MISSED,
NEW

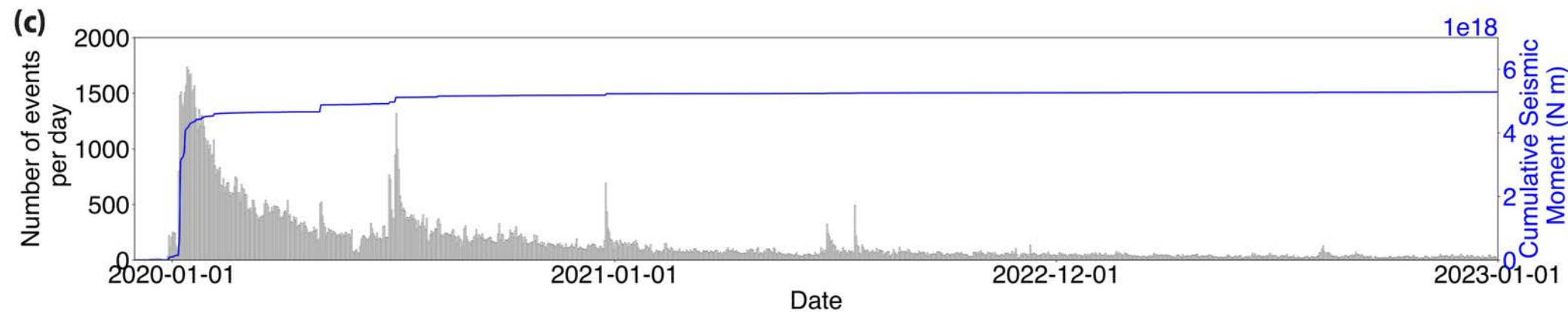


Magnitude-frequency
distribution (MFD)

M_w 6.4 mainshock



Magnitude
vs. time

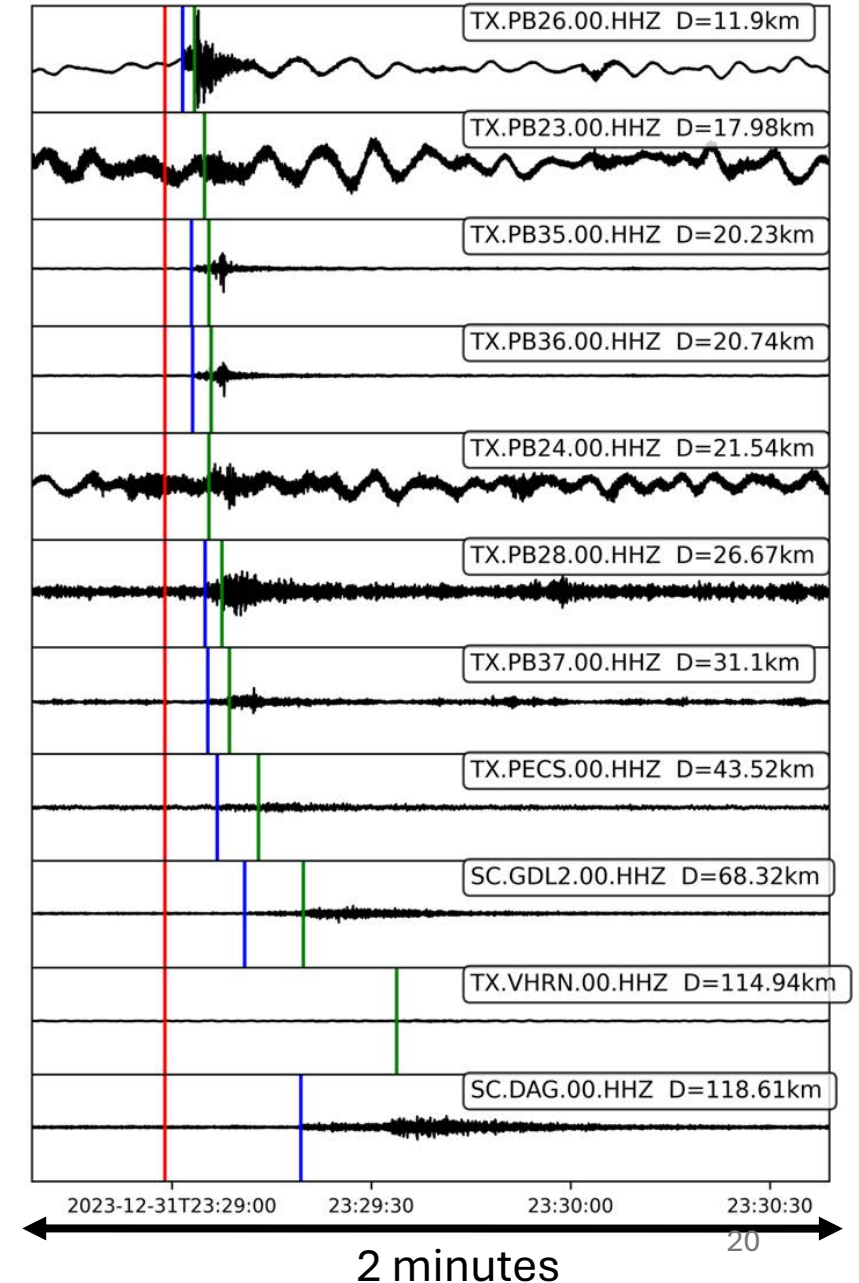


Number of
events/day,
**Cumulative
moment**
vs. time

Earthquake Catalog QC Tips: Check Waveforms

- When in doubt, **plot event waveforms**
 - Overlay picks on event waveforms, ordered by event-station distance, especially for “**NEW**” events not in reference catalog
 - **Earthquake or noise?**
 - **Distance** from event to **nearest station?**
 - **Frequency content & time duration** – local, regional, or teleseismic earthquake? Quarry blast?
 - **Moveout** (from **P**, **S** pick times) & **attenuation** (from amplitudes) with distance?
 - Do actual picks match predicted arrival times (from ray-tracing through velocity model)?

PyOcto Event #340 : 5/5/10 No Match to TexNet, Match to GaMMA, NSD<75.0
Red: Event, Blue: P Pick, Green: S Pick



Check “NEW” events in enhanced catalog, not in reference catalog

Sorted by **magnitude**
(Southwest Puerto Rico)

Origin time (UTC)	OT (seconds)	Latitude	Longitude	Depth	Magnitude	EventID
2020-01-07T08:35:15.030000	30915.030000	17.871833	-66.721667	6.420000	4.320000	1000335
2020-01-07T11:21:01.620000	40861.620000	18.034167	-66.800000	13.780000	4.360000	1000569
2020-01-07T08:29:18.320000	30558.320000	18.139667	-66.810000	7.660000	4.440000	1000328
2020-01-07T08:29:36.930000	30576.930000	17.963500	-66.752333	29.160000	4.480000	1000330
2020-01-07T08:51:09.400000	31869.400000	17.676333	-66.773667	27.610000	4.760000	1000356

Catalog QC: False detection in coda of larger earthquake (also, too deep?)

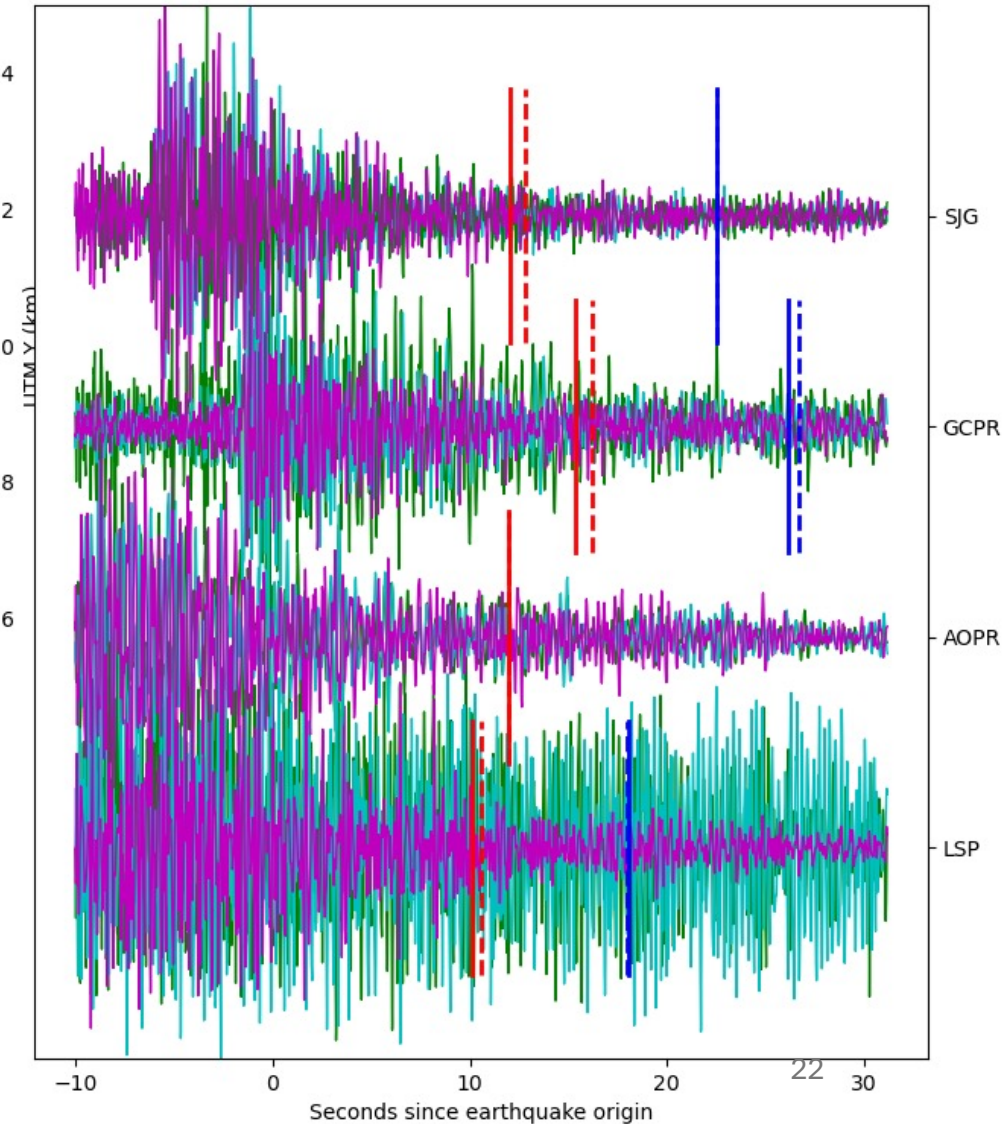
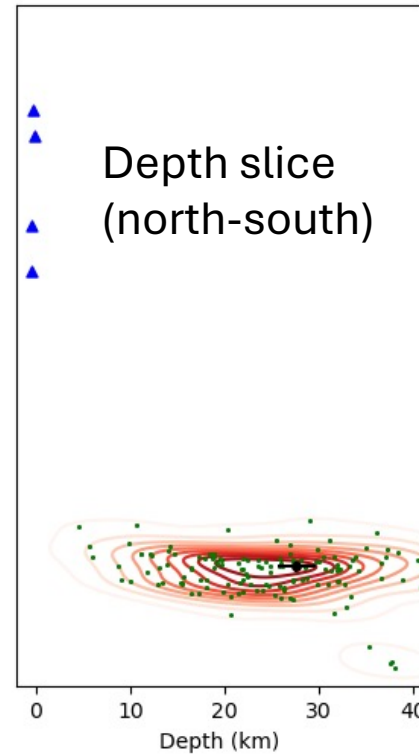
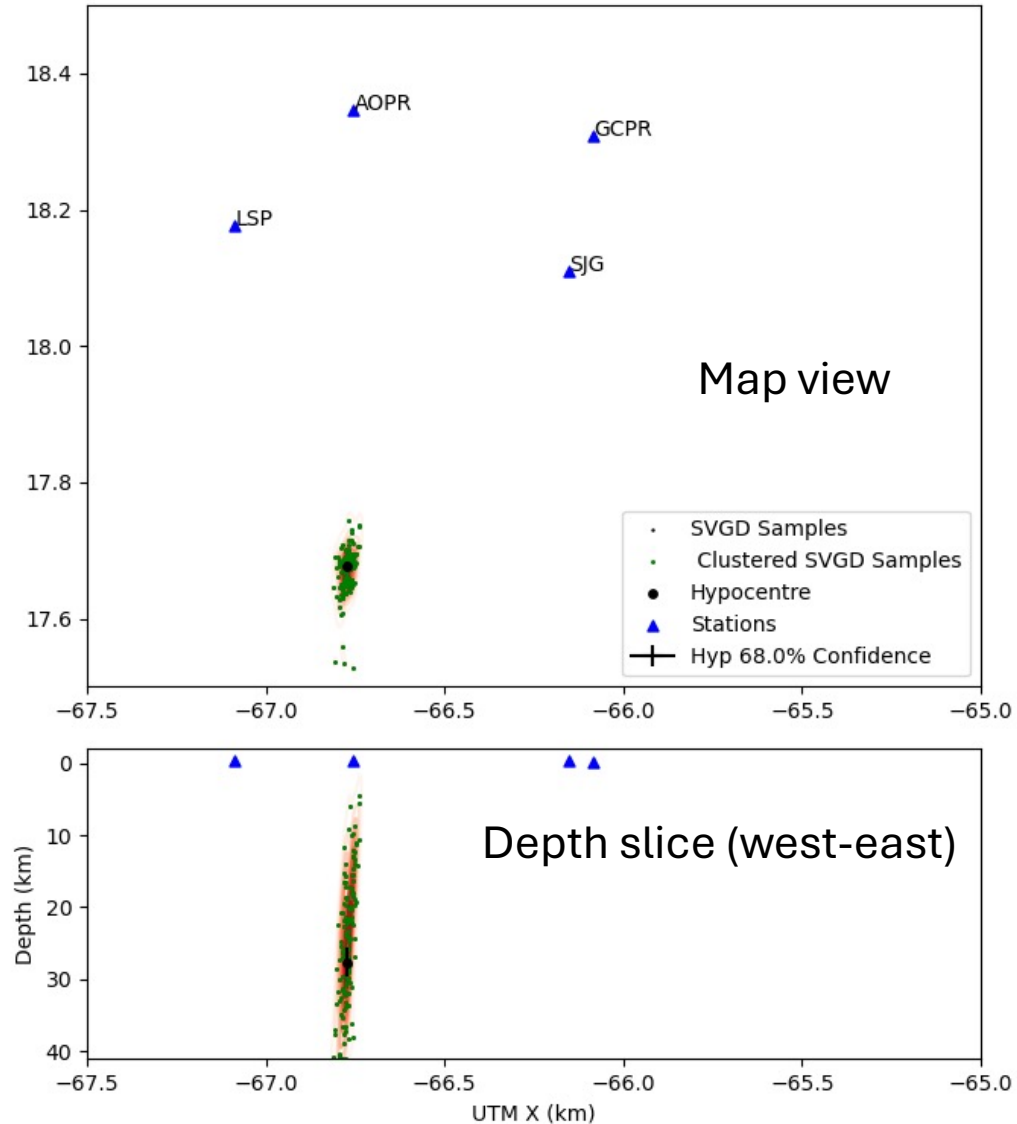


2020-01-07T08:51:09.400000 31869.400000 17.676333 -66.773667 27.610000 4.760000 1000356

Earthquake 2020-01-07 08:51:09.396302640 +/- 0.50s
Hyp=[-66.77,17.68,27.61] - Hyp Uncertainty (km) +/- [0.35,0.84,1.97]

Event waveforms with P, S picks

Solid: auto-picks, Dashed: predicted arrivals



Catalog QC: Real (newly detected) earthquake (reasonable depth)

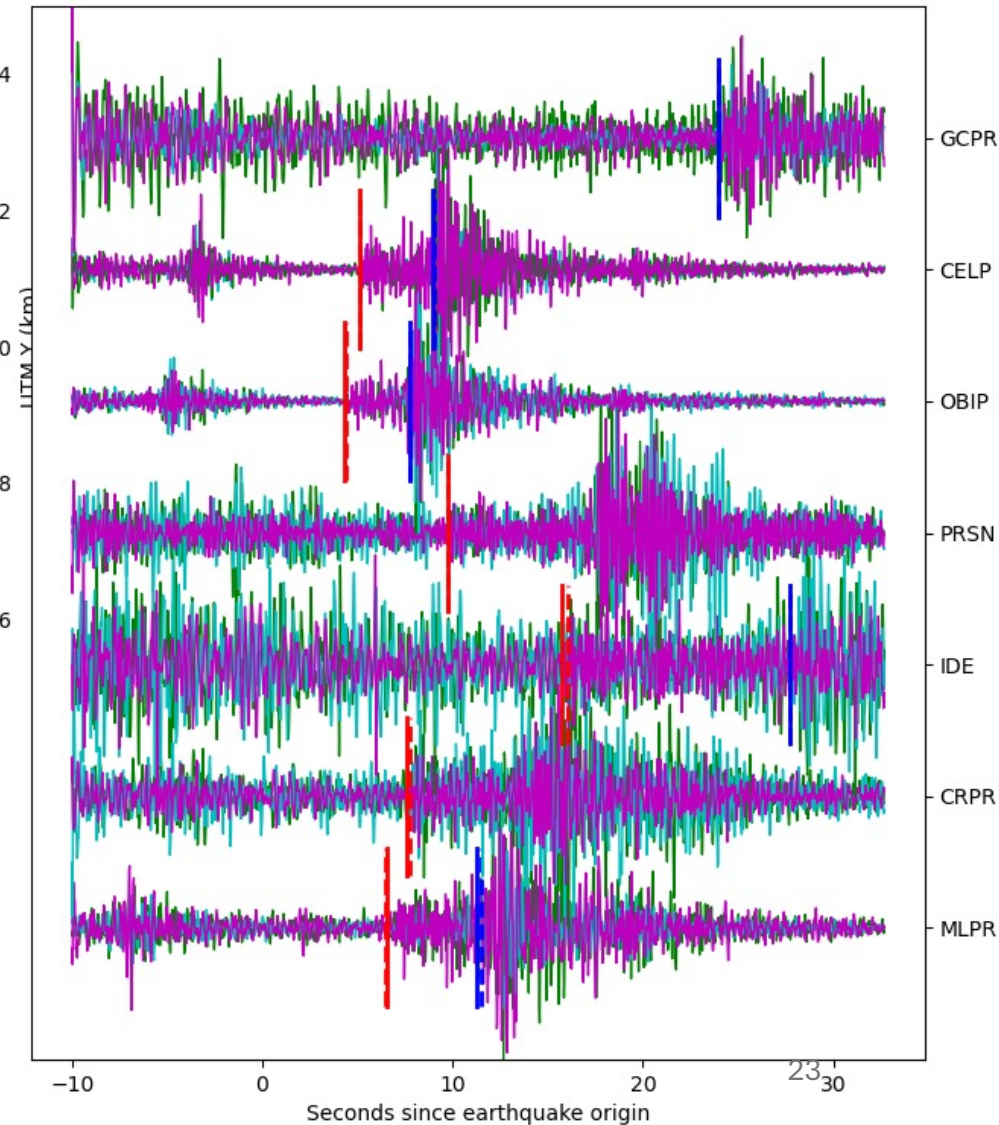
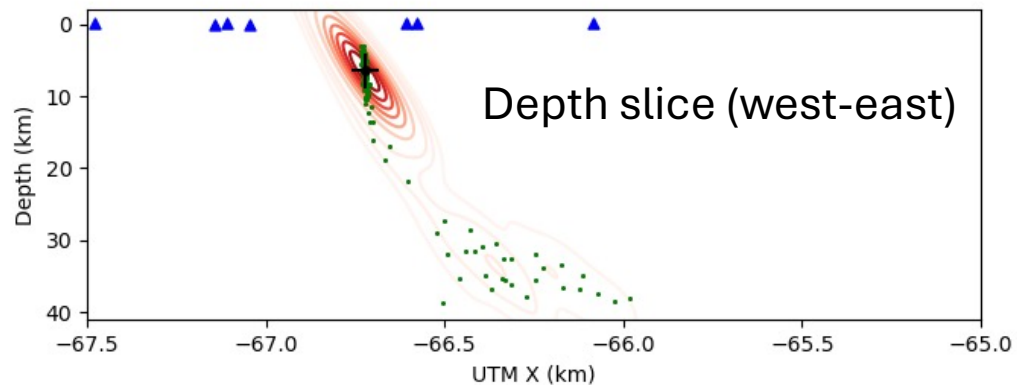
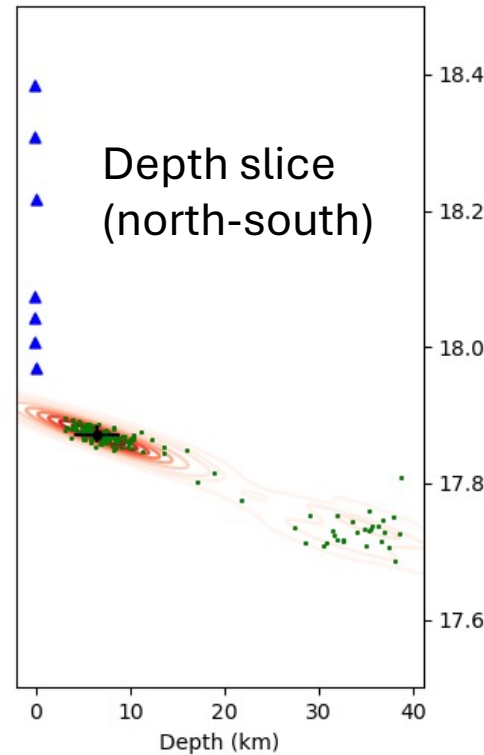
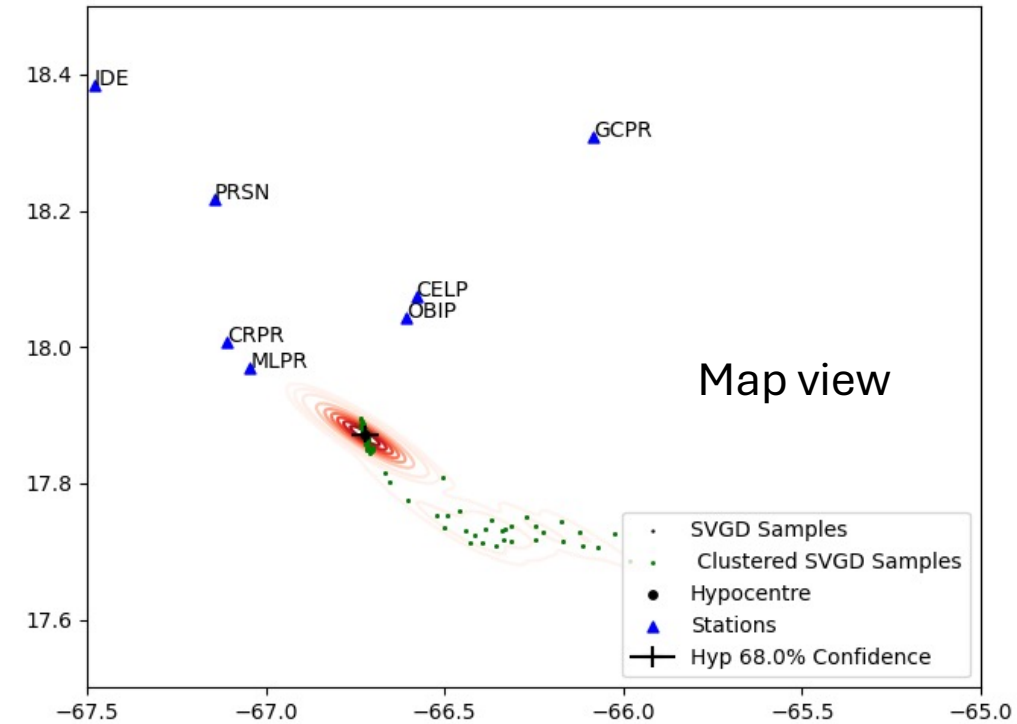
2020-01-07T08:35:15.030000 30915.030000 17.871833 -66.721667 6.420000 4.320000 1000335



Earthquake 2020-01-07 08:35:15.029341138 +/- 0.05s
Hyp=[-66.72,17.87,6.42] - Hyp Uncertainty (km) +/- [4.13,1.33,2.41]

Event waveforms with P, S picks

Solid: auto-picks, Dashed: predicted arrivals



Check “NEW” events in enhanced catalog, not in reference catalog

Sorted by **depth**
(Southwest Puerto Rico)

Origin time (UTC)	OT (seconds)	Latitude	Longitude	Depth	Magnitude	EventID
2020-01-07T16:19:28.490000	58768.490000	18.146667	-66.835500	37.840000	2.980000	1001012
2020-01-07T20:43:26.700000	74606.700000	17.933500	-66.923333	38.810000	1.570000	1001342
2020-01-07T15:58:52.370000	57532.370000	18.353667	-67.028000	39.160000	1.310000	1000980
2020-01-07T20:27:22.970000	73642.970000	17.931000	-66.846500	39.800000	1.560000	1001320
2020-01-07T22:21:53.870000	80513.870000	18.466000	-66.986833	40.580000	2.170000	1001466



Unrealistic depths
for the region?
(SW Puerto Rico)

Catalog QC: False detection in coda of larger earthquake (also, too deep?)

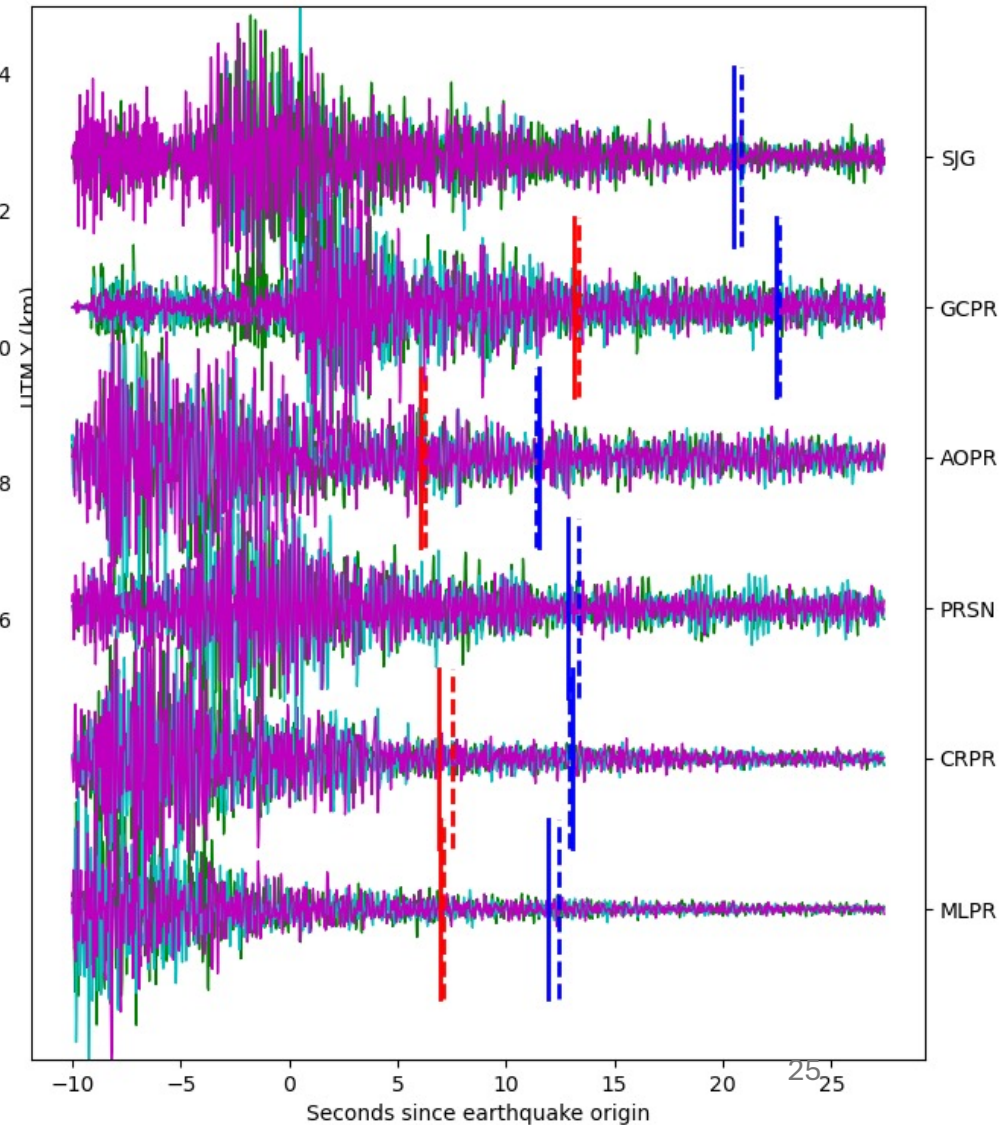
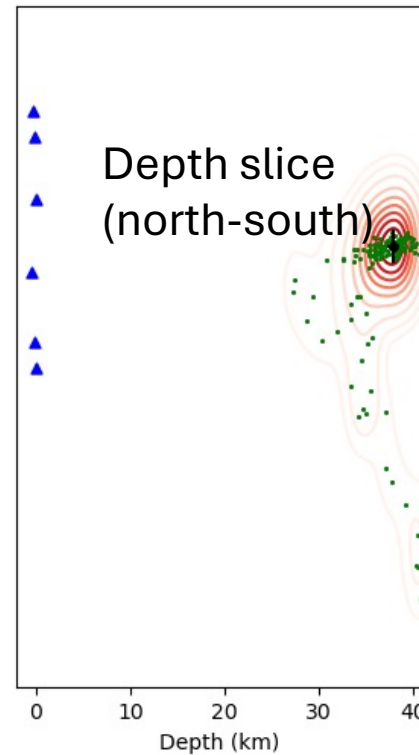
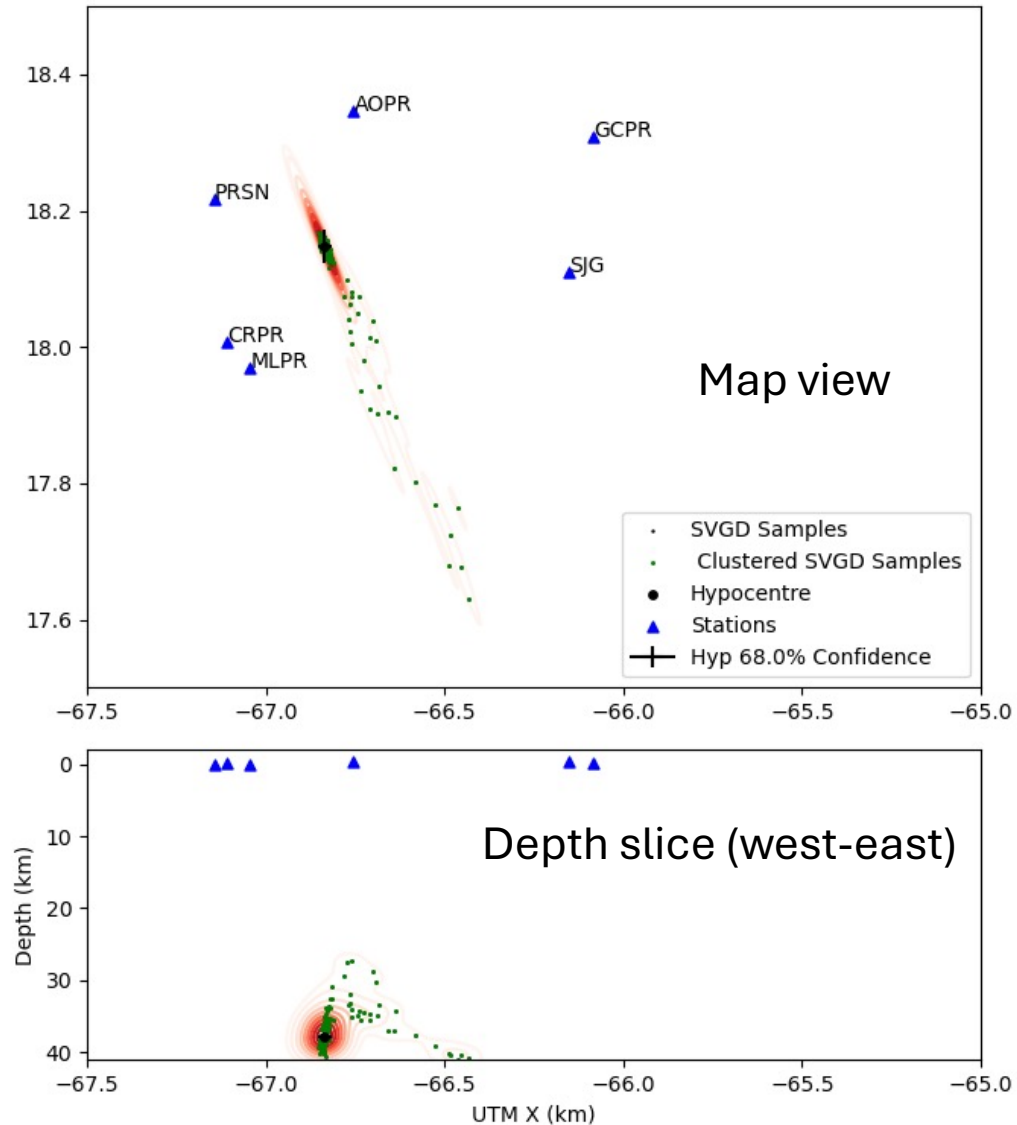


2020-01-07T16:19:28.490000 58768.490000 18.146667 -66.835500 37.840000 2.980000 1001012

Earthquake 2020-01-07 16:19:28.491959877 +/- 0.20s
Hyp=[-66.84,18.15,37.84] - Hyp Uncertainty (km) +/- [2.08,2.71,0.58]

Event waveforms with P, S picks

Solid: auto-picks, Dashed: predicted arrivals

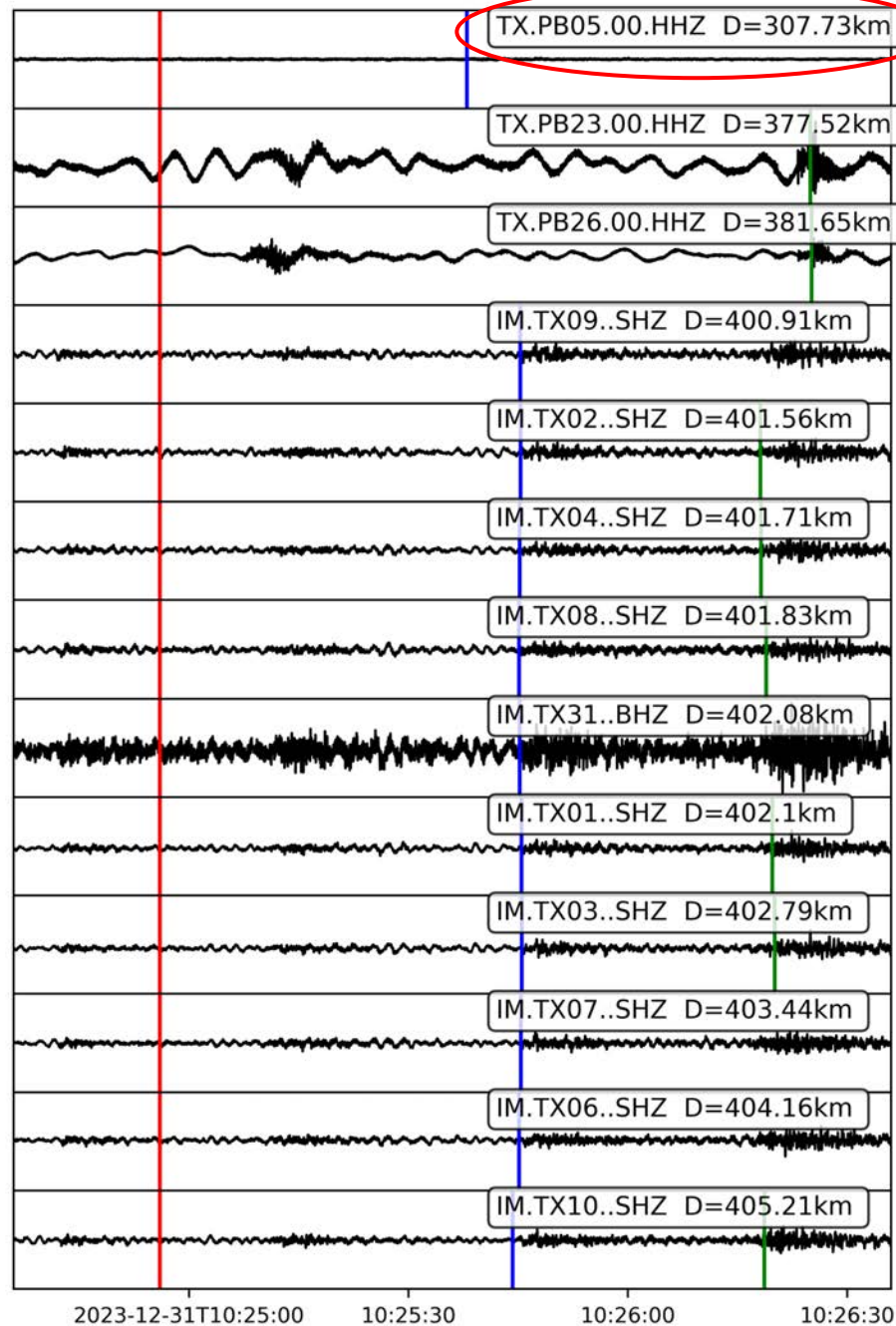
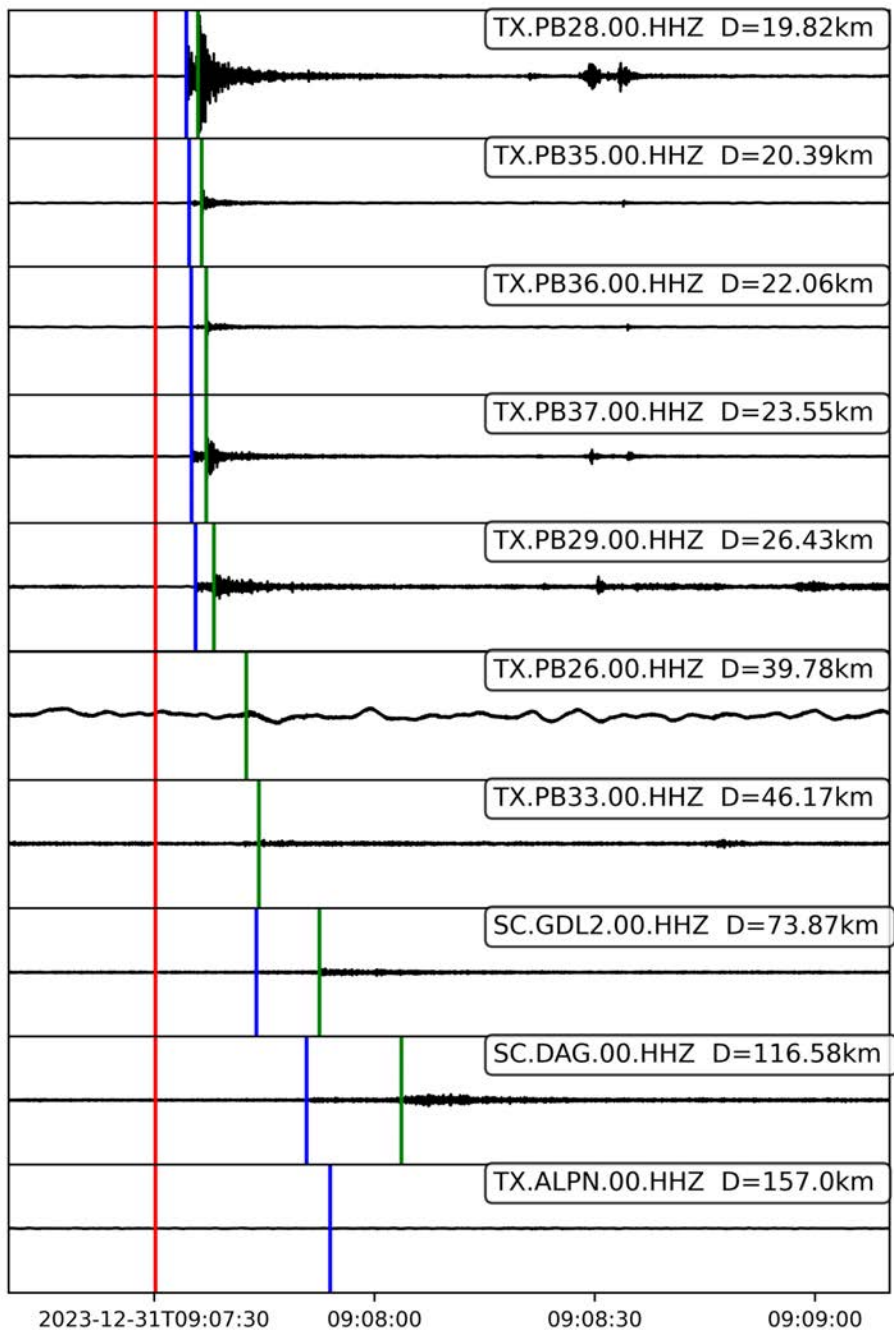


Post-processing: automatic removal of false detections in coda of larger earthquakes

- Visual detection of event waveforms with unrealistic depths (30-40 km): false detections in coda of larger earthquakes
 - EQTransformer (Mousavi et al., 2020) with low thresholds for event detection and P, S picks
- Devised an empirical algorithm to remove these specific false detections (Yoon et al., 2023, BSSA)
 - Apply only to NEW events not in reference catalog, not to MATCH events. Must meet ALL 4 criteria below for removal.
 - 1) current event was within 45 seconds of the previous event in time;
 - 2) current event had 14 or fewer phases;
 - 3) previous event had at least 5 more phases than the current event;
 - 4) current & previous events within 1 local magnitude unit of each other

PyOcto Event #82 : No Match to ComCat or GaMMA
 Red: Event, Blue: P Pick, Green: S Pick

PyOcto Event #90 : No Match to ComCat or GaMMA
 Red: Event, Blue: P Pick, Green: S Pick

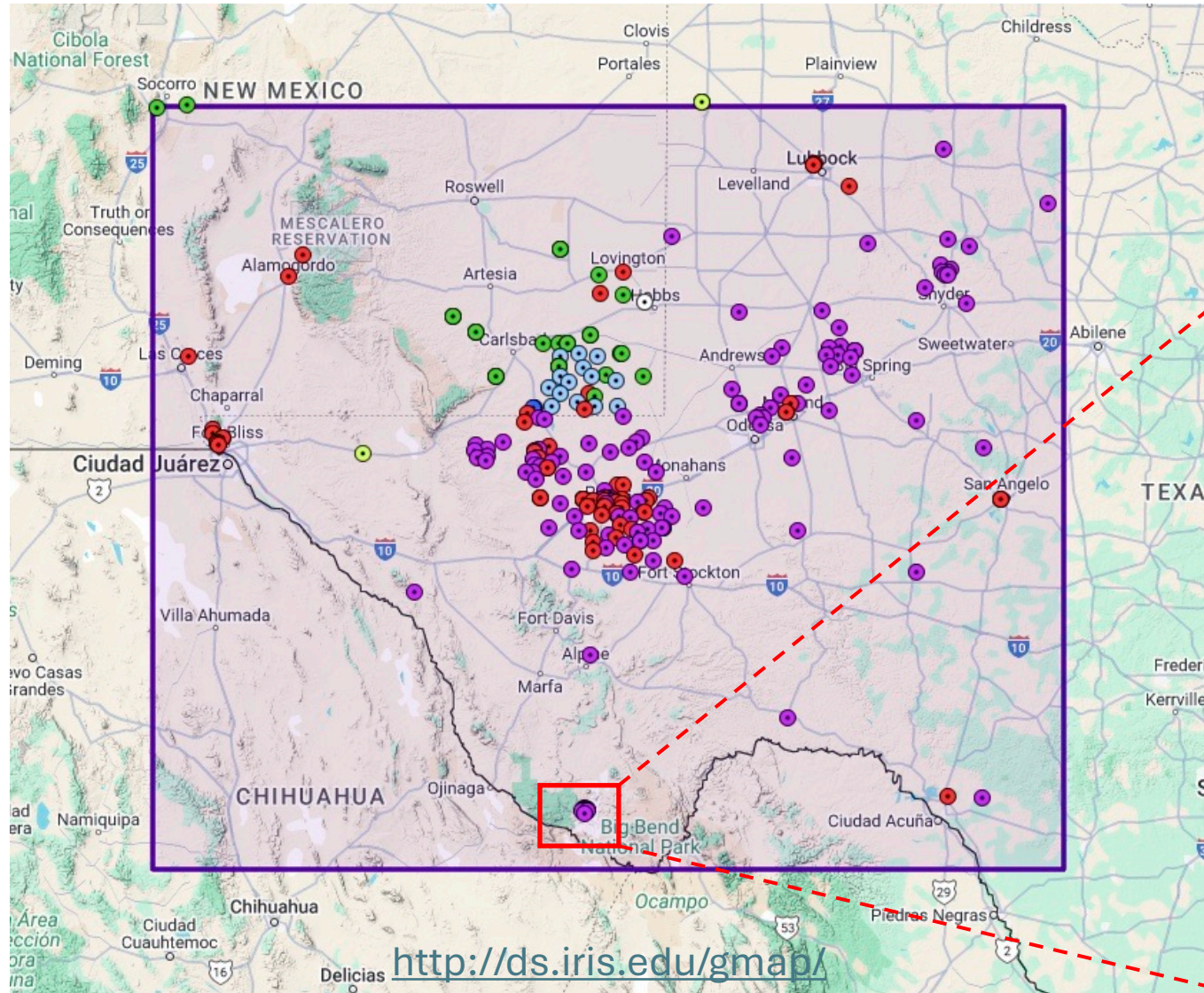


- Catalog QC: West Texas
- After event association
- (Left) new local earthquake 😊
- (Right) false detection 😞
 - *large distance (307 km) to nearest station; IM network?*

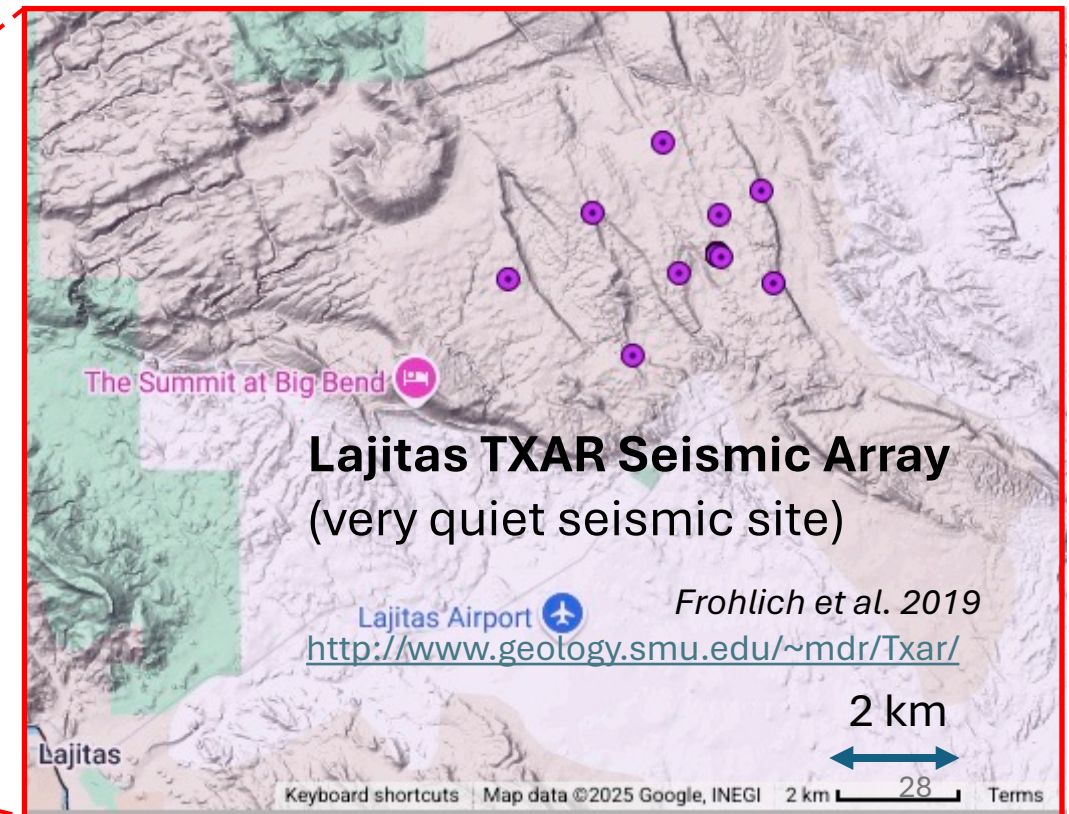
2-minute waveforms
 Vertical lines:

Origin time, P, S

Problem: Incorrect association/location due to **uneven station distribution**
(array far from rest of network & seismic sources)

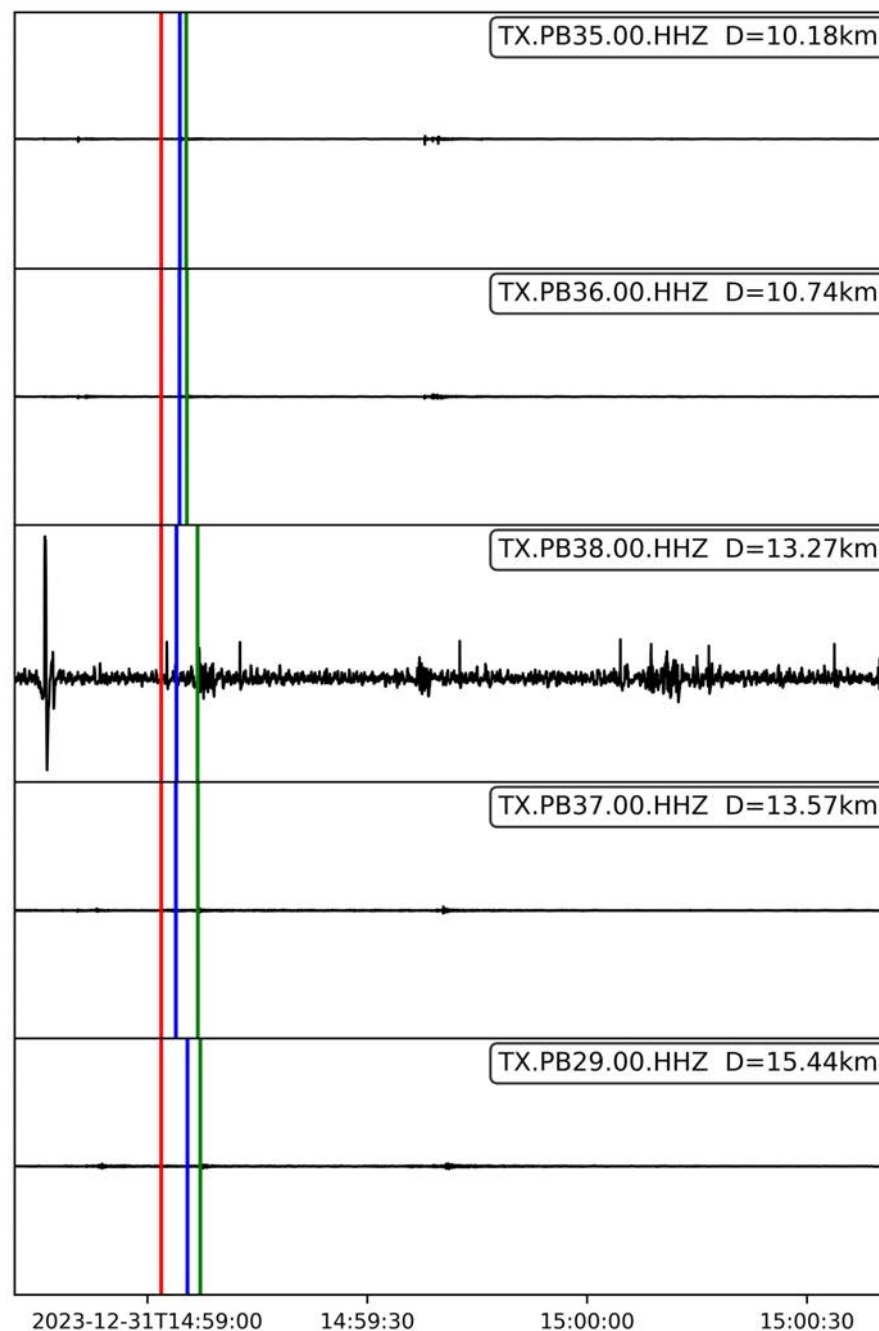
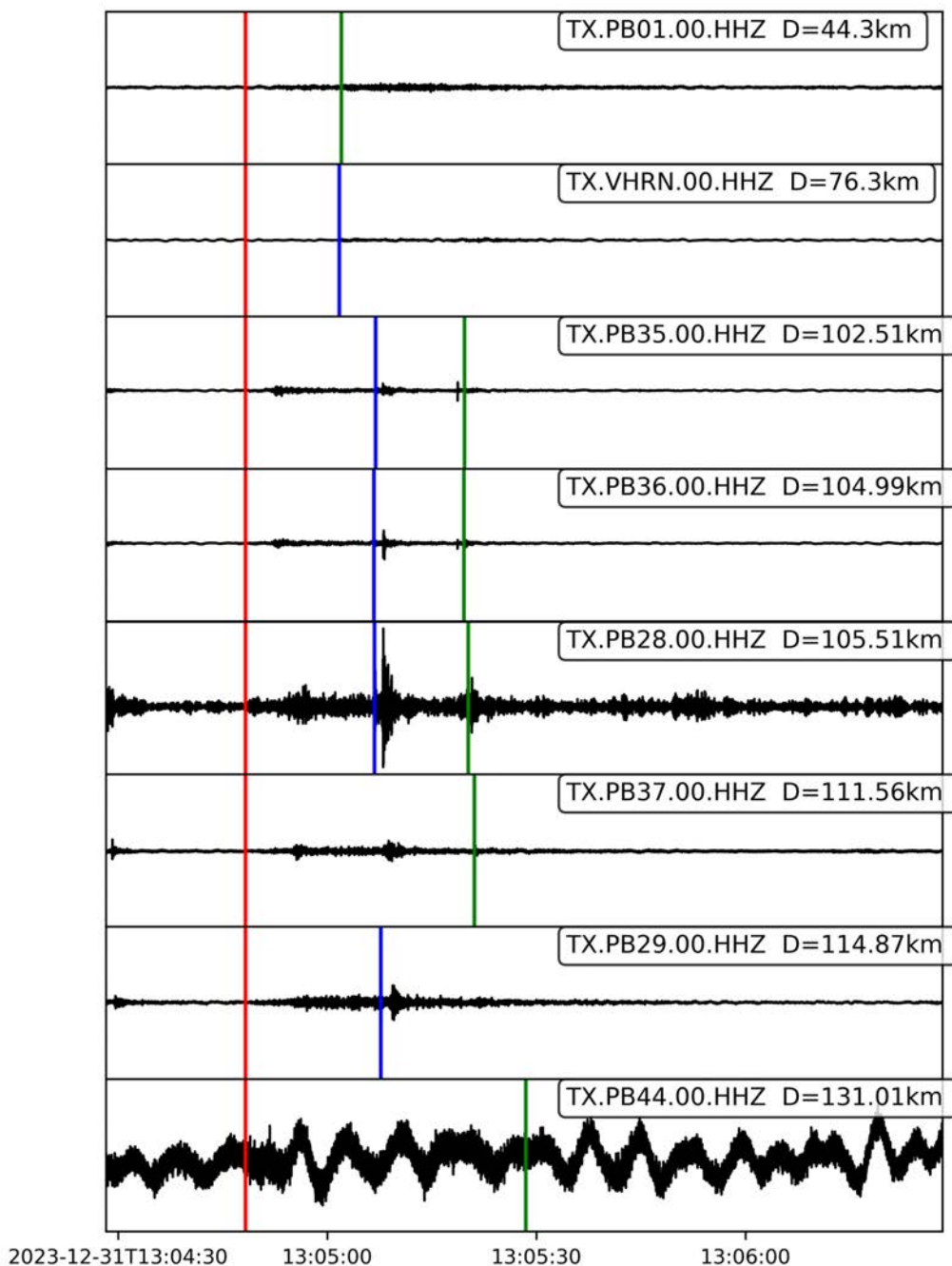


Solution: Use only one station from array (IM.TX01) for event association



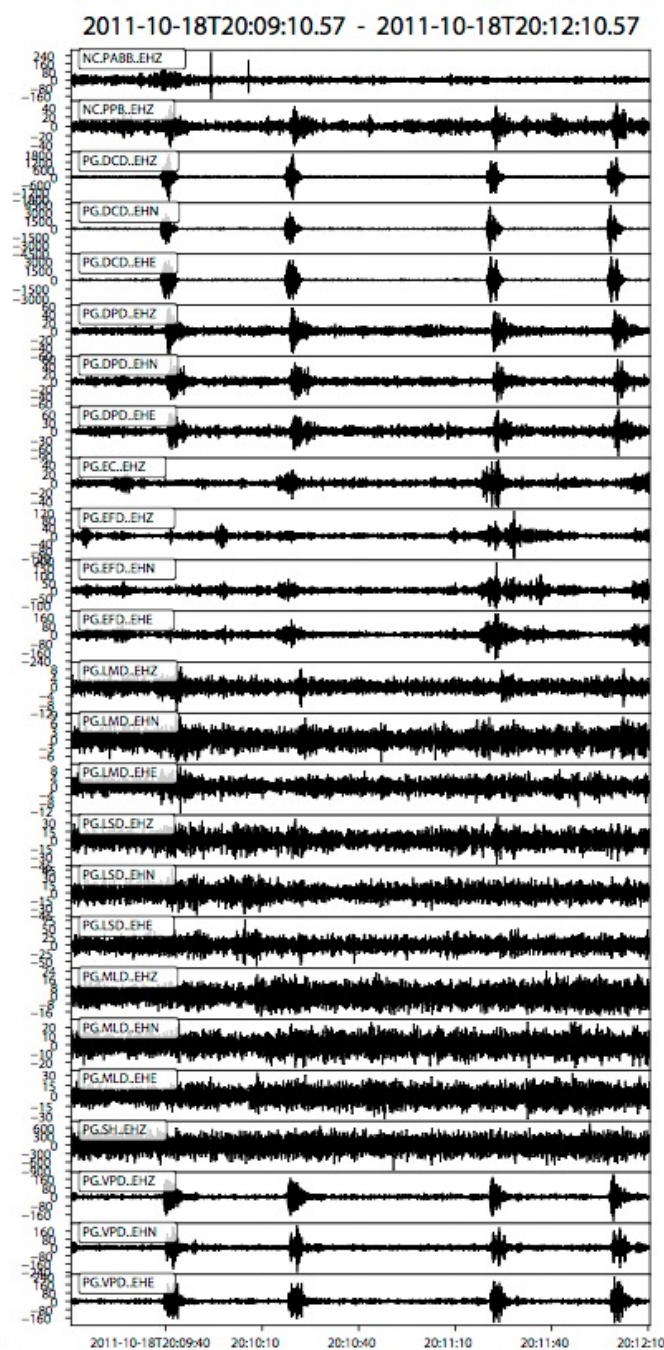
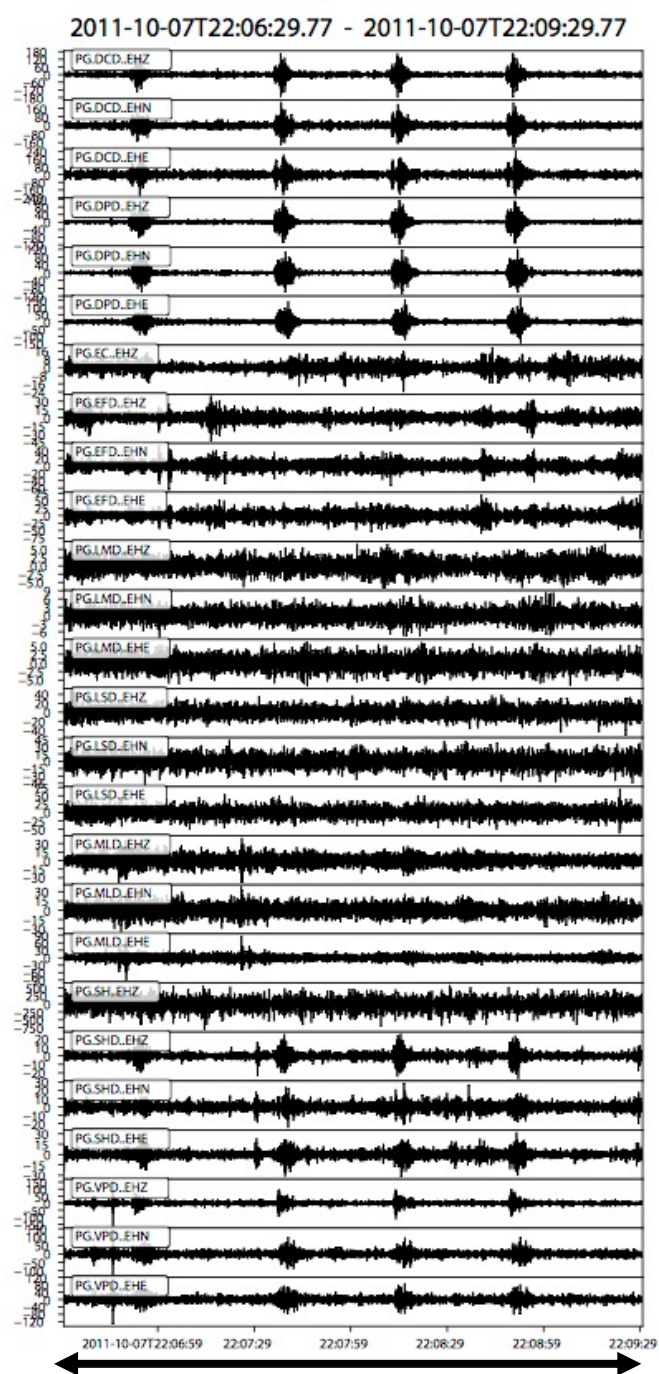
PyOcto Event #218 : 5/5/10 No Match to TexNet, Match to GaMMA, NSD<75.0
 Red: Event, Blue: P Pick, Green: S Pick


PyOcto Event #250 : 5/5/10 No Match to TexNet, Match to GaMMA, NSD<75.0
 Red: Event, Blue: P Pick, Green: S Pick



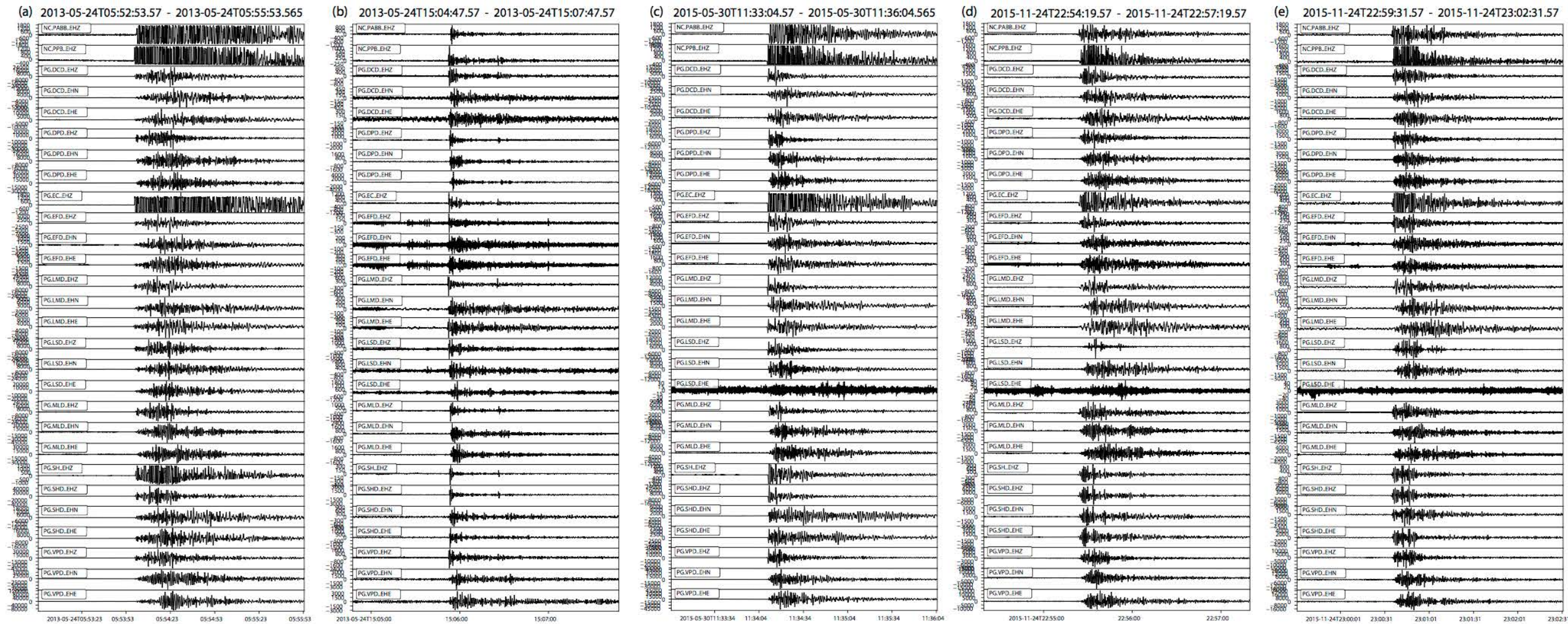
- Catalog QC: West Texas
- **Bad** event association 😞
- (Left) **P**, **S** picks from 2 different earthquakes associated as one event
- (Right) Noise picks associated as event → wrong moveout

2-minute waveforms
 Vertical lines:
 Origin time, **P**, **S**



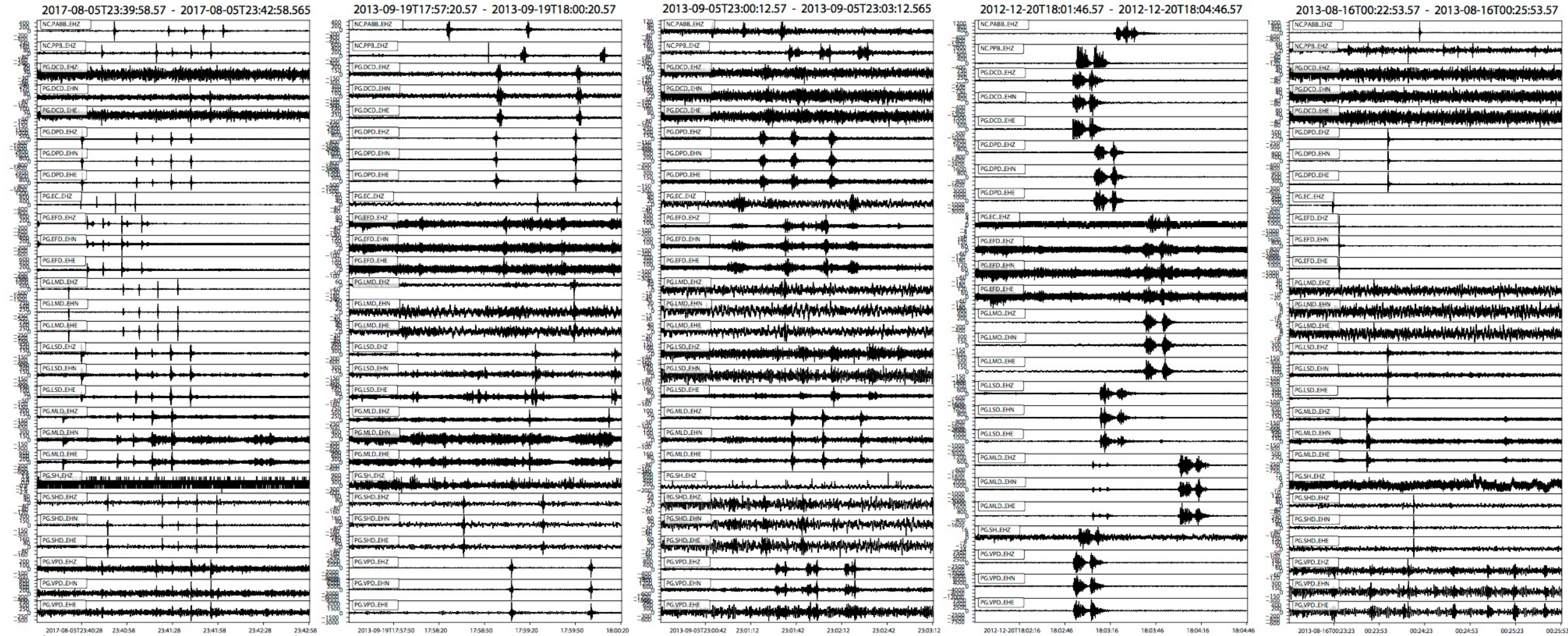
- Catalog QC: Central CA
- Filtered 3-12 Hz
- Non-earthquake signals: 

Deep teleseismic earthquakes (depth 600+ km) recorded in Central CA



Very little moveout (waves hit all stations at same time)

Infrasound signals recorded in Central CA



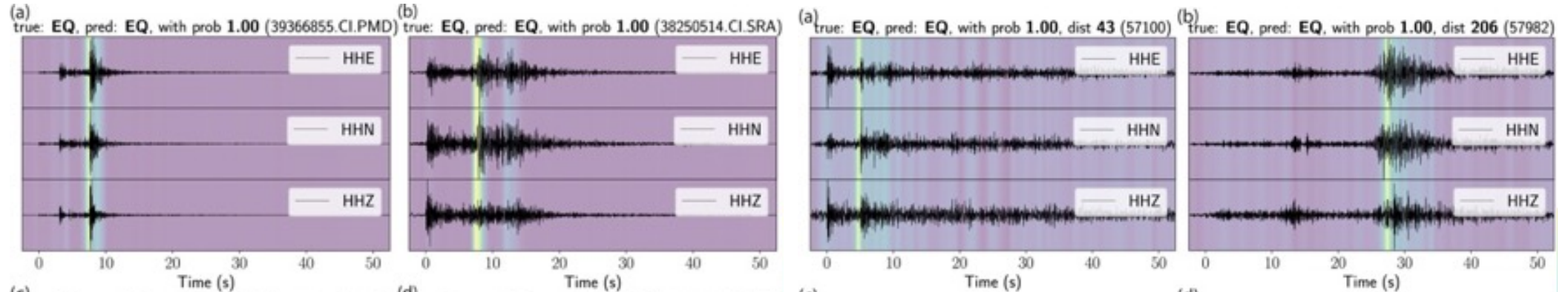
Slower moveout from sound waves

Quarry blasts: often in enhanced catalogs, but not 'interesting' to many seismologists

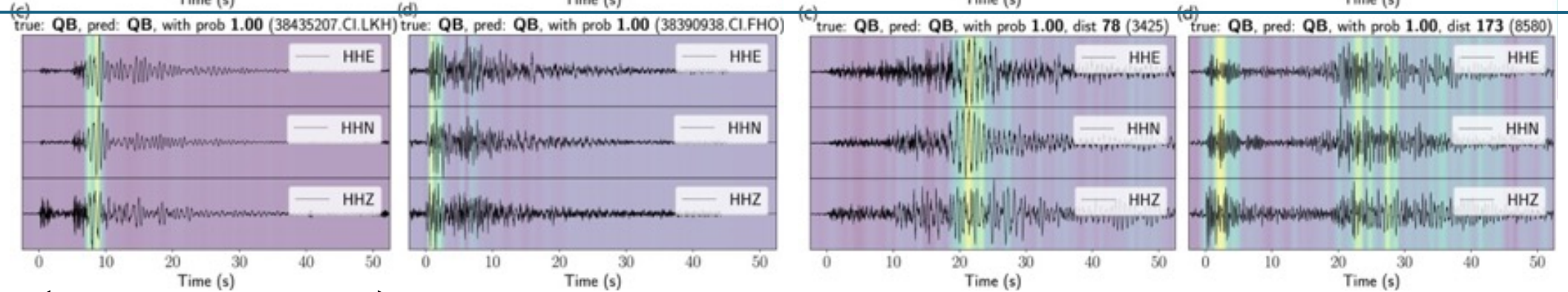


Check ComCat for blast events

earthquake



quarry blast



50 seconds

Check waveform plots:
lower frequency content?

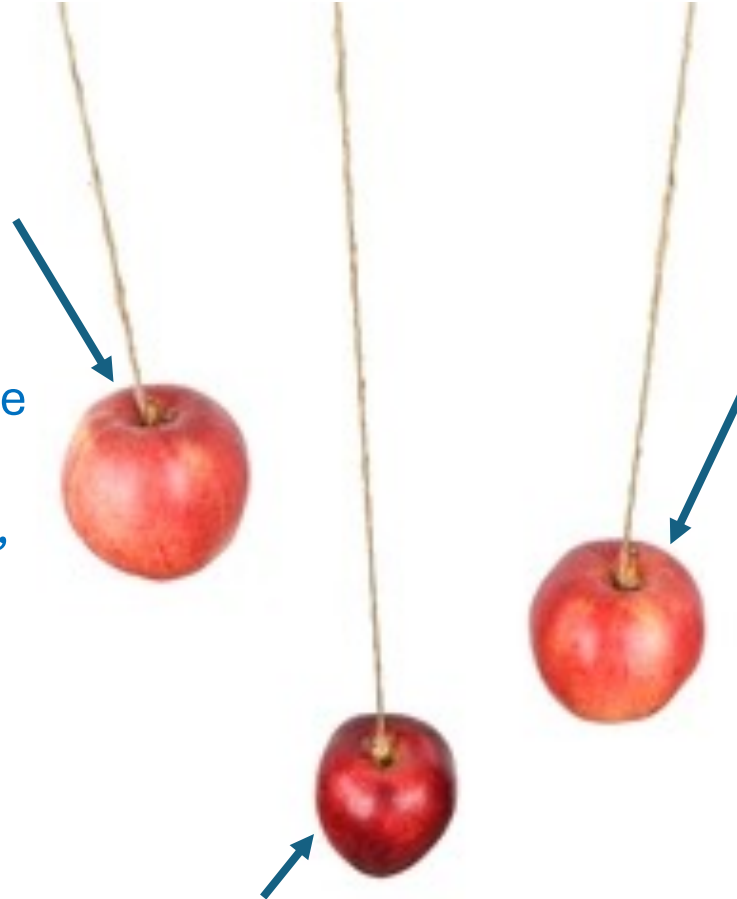
Events located near
known quarry?

Events occur only
in local daytime?

Automated post-processing strategies to remove false detections & unwanted signals from catalog

[Machine-learning]
classifiers: discriminate
different seismic signals

- Teleseismic earthquakes, quarry blasts, cultural noise sources, infrasound, ...
- Assemble training data set, create classifier model



Empirical algorithms: remove
specific false detections after
visual inspection

- e.g. false detections in coda of larger earthquakes
- may be specific to network geometry & source distribution

Simple thresholding & filtering to
discard non-earthquake signals

- easy & effective, but be aware of tradeoffs

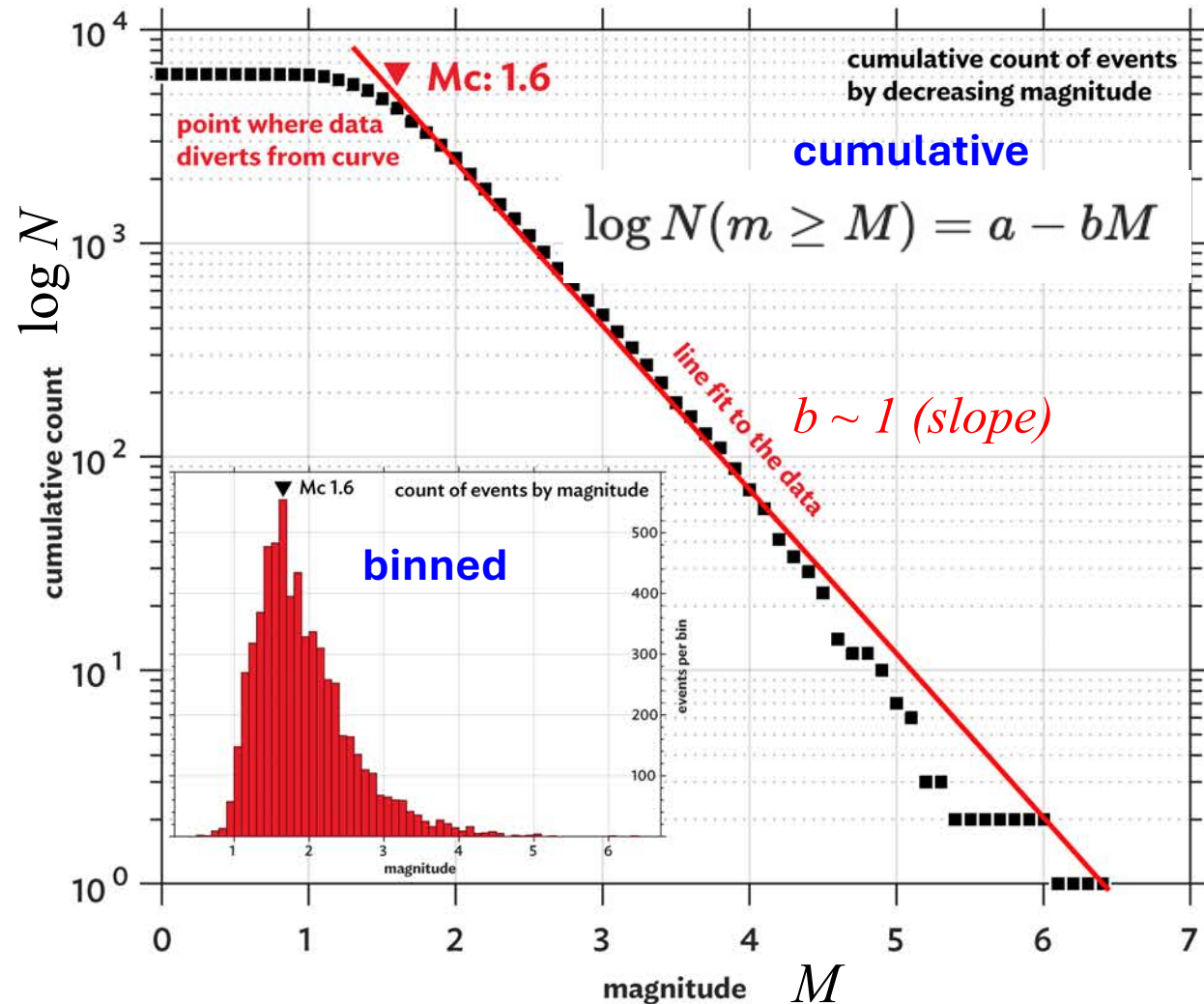
Simple thresholding & filtering to discard non-earthquake signals

- easy & effective, but be aware of tradeoffs

- Restrict locations to a certain region and/or depth range
- Set minimum threshold on:
 - Number of (P, S) picks
 - Number of stations per event
 - Output probabilities for P, S picks
- Set maximum bound on:
 - Distance to nearest station
 - Distance to farthest contributing station with picks
 - Azimuthal gap
 - RMS residual
 - Travel time residual for P, S pick at given station

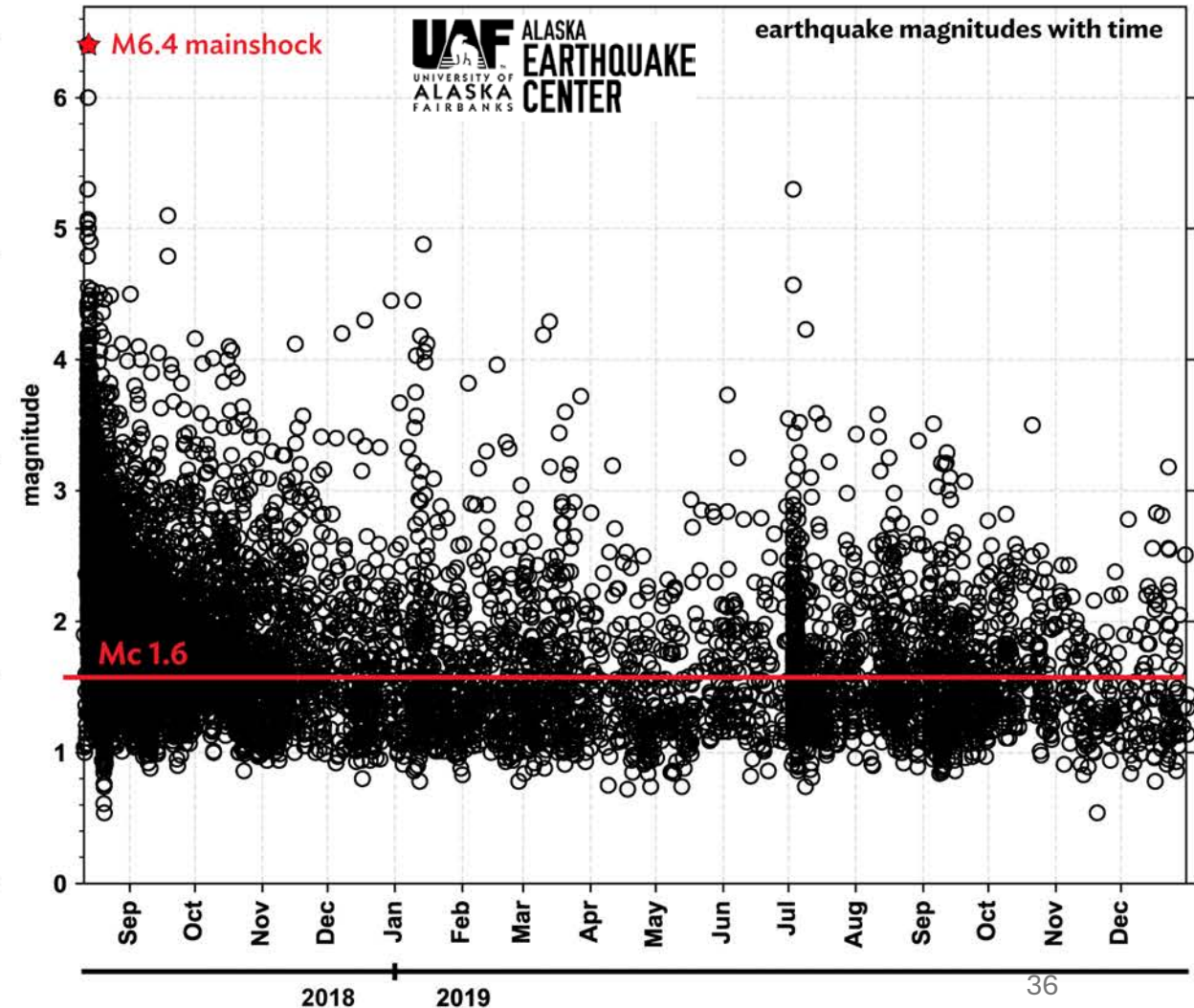
Magnitude of Completeness (M_c)

Magnitude-frequency distribution (MFD) (Gutenberg-Richter)



M_c : lowest magnitude above which the earthquake catalog is "complete"

Lower M_c is better (more complete)



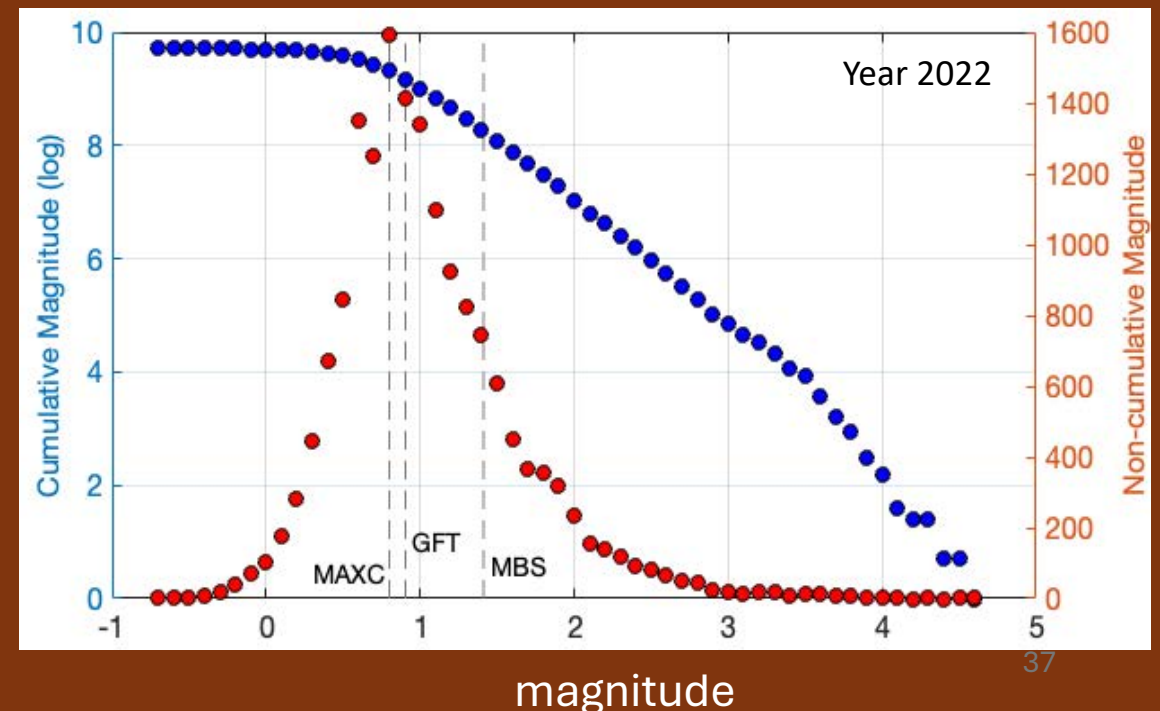
Many different methods to get catalog M_c

- **Maximum Curvature (MAXC):** max value of 1st derivative of MFD (typically, max of binned MFD)
 - * most reliable for small sample sizes (~50-100 events)
 - * works best with local datasets with fewer heterogeneities
- **Goodness-of-fit Test (GFT):** lowest magnitude cut-off where Gutenberg-Richter relation holds
- **Mc by b-value Stability (MBS):** first magnitude increment where $|b_{\text{avg}} - b| < \text{uncertainty of } b$
 - * tends to produce the highest (i.e., most conservative) M_c

* More info and original R codes from CORSSA: <https://corssa.org/>
<https://doi.org/10.5078/corssa-39071657>

* Matlab codes available on Github:
https://github.com/gtepp/research_codes

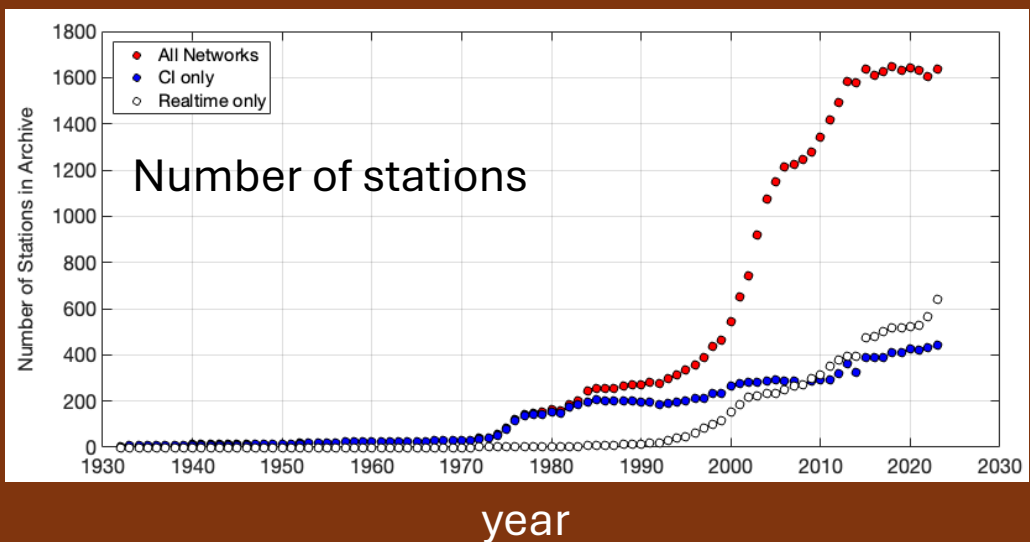
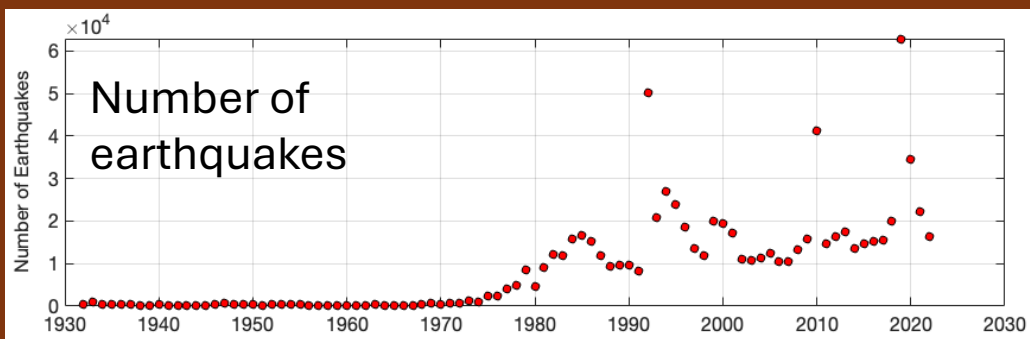
* Other methods to get M_c :
ZMAP (Wiemer, 2001)
b+ positive (van der Elst, 2021)



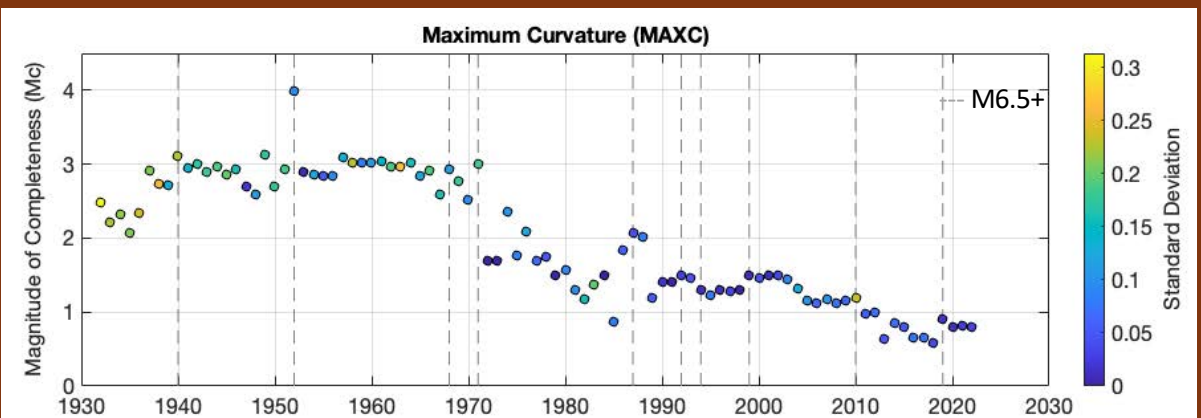
Temporal M_c changes

Past 90 years: SCSN M_c has decreased
from ~ 3 to ~ 1

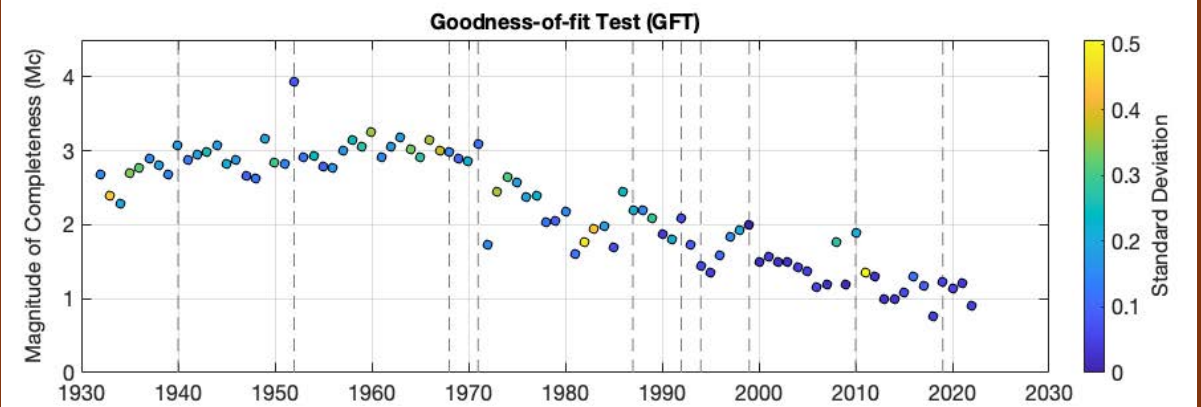
Full SCSN Catalog



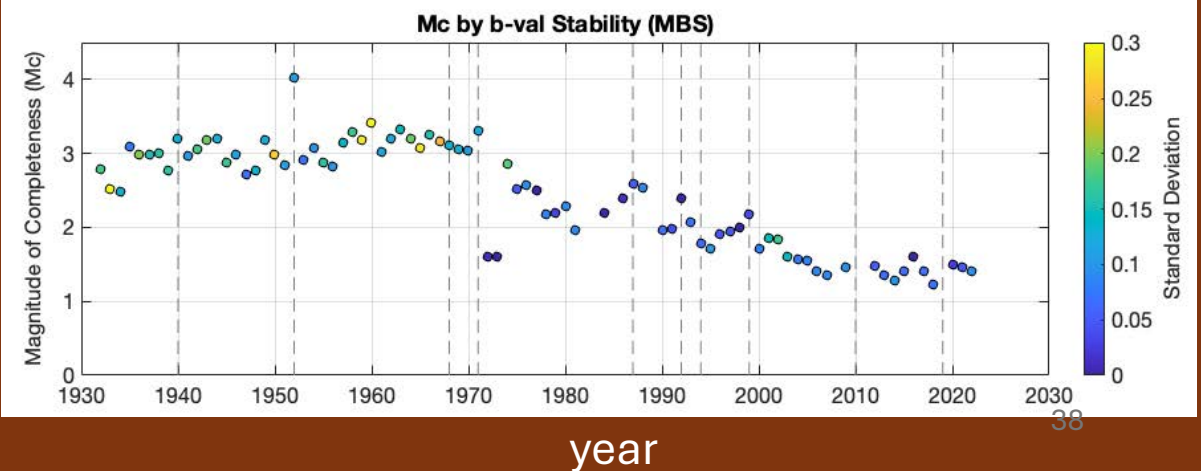
M_c



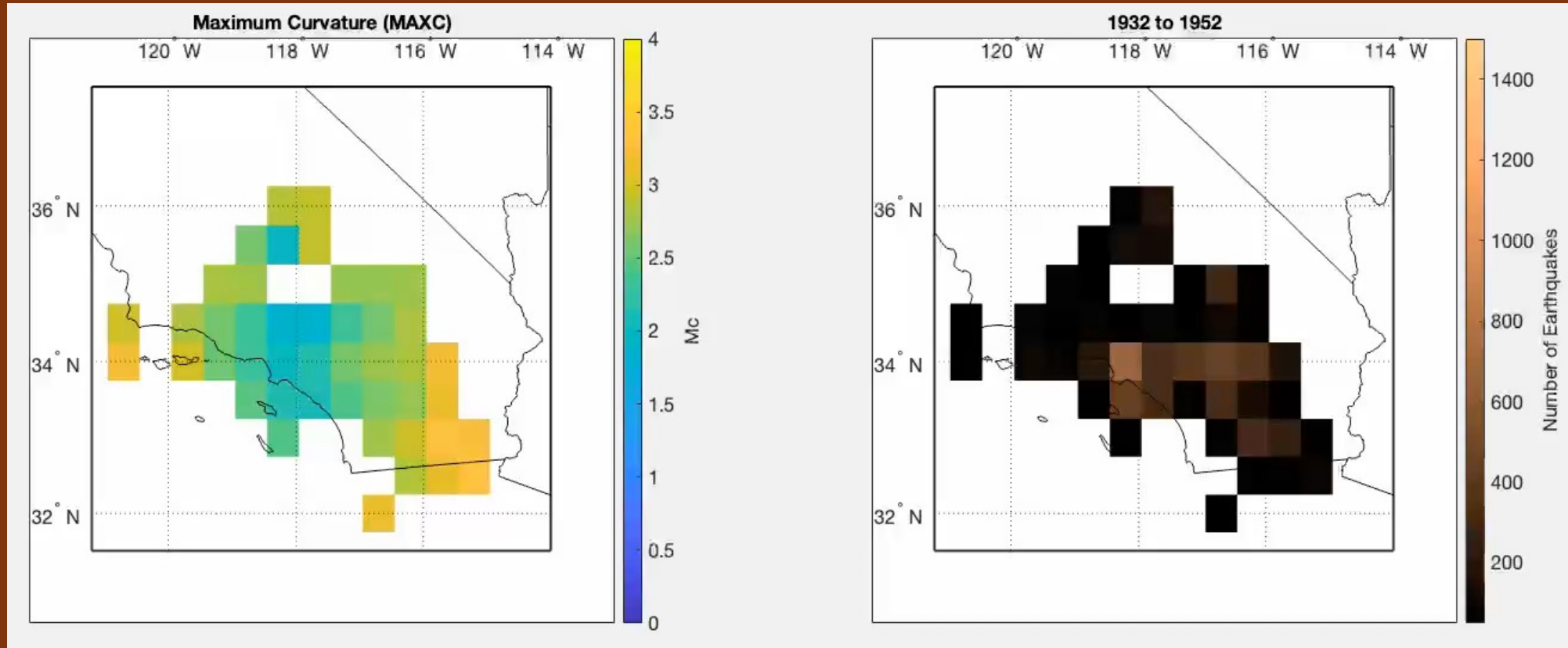
M_c



M_c

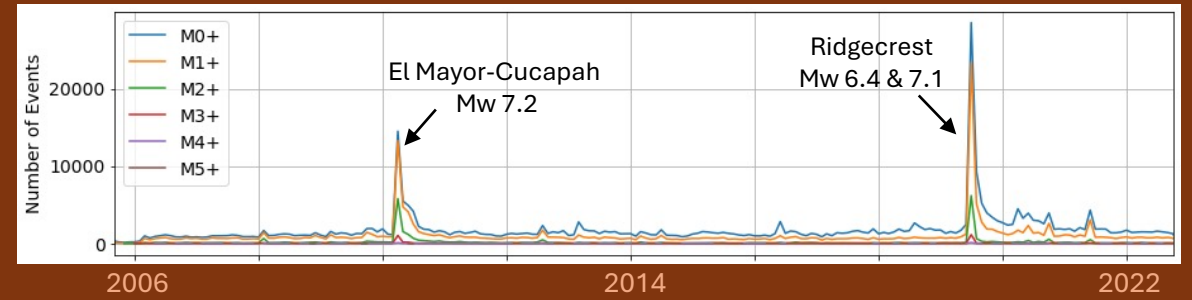
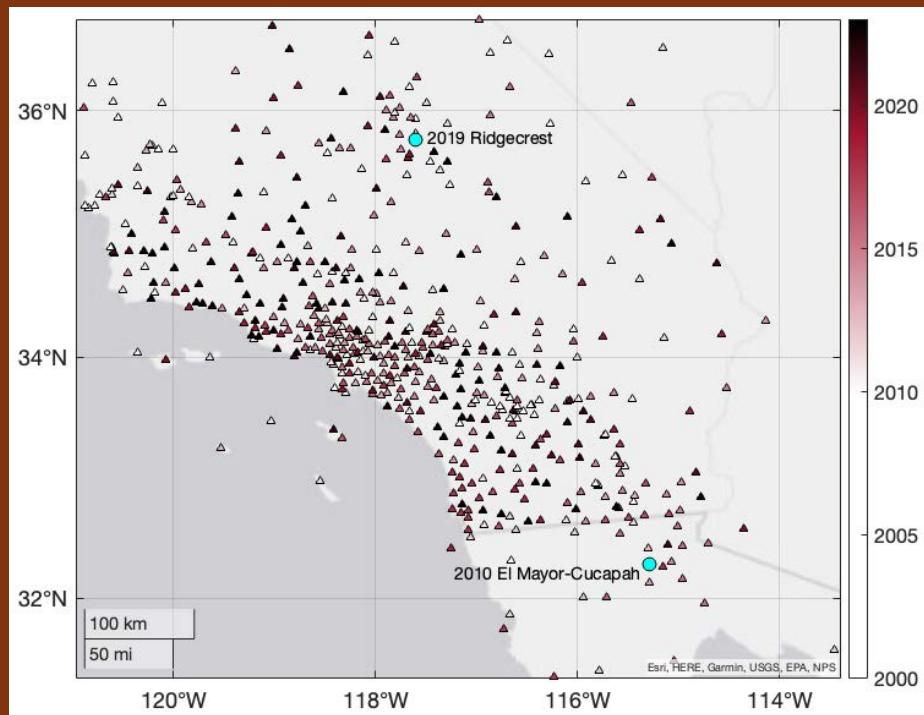


Spatial Mc Changes



- Bins: 0.5° (~ 50 km) squares
- 1932-1972: 20 years
- 1973-2022: 10 years
- Maximum curvature method
- Minimum 30 EQs/bin

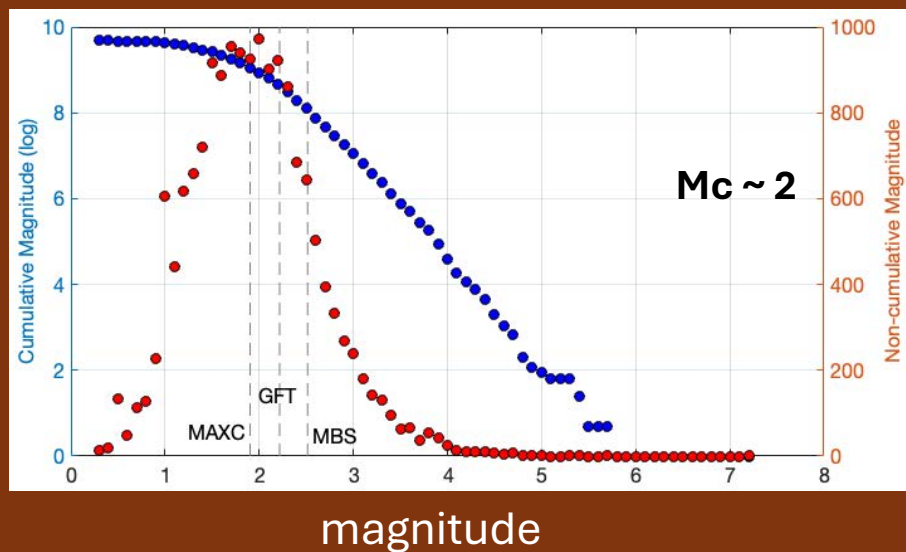
Past 90 years: SCSN Mc has decreased from ~ 3 to ~ 1 , but not uniformly throughout the region



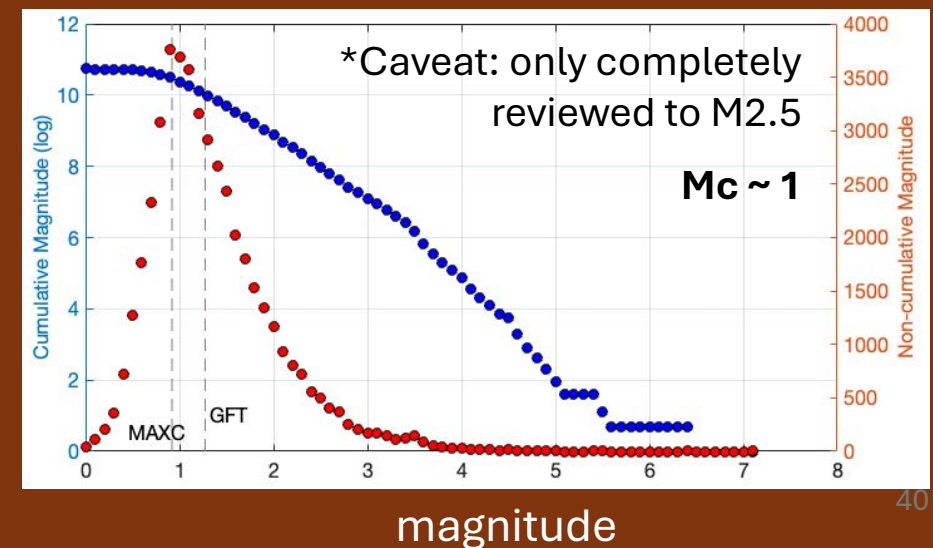
Expect higher M_c earlier in aftershock sequence

Ridgecrest M_c is ~ 1 mag. unit lower than El Mayor-Cucapah M_c

2010 M_w 7.2 El Mayor-Cucapah sequence

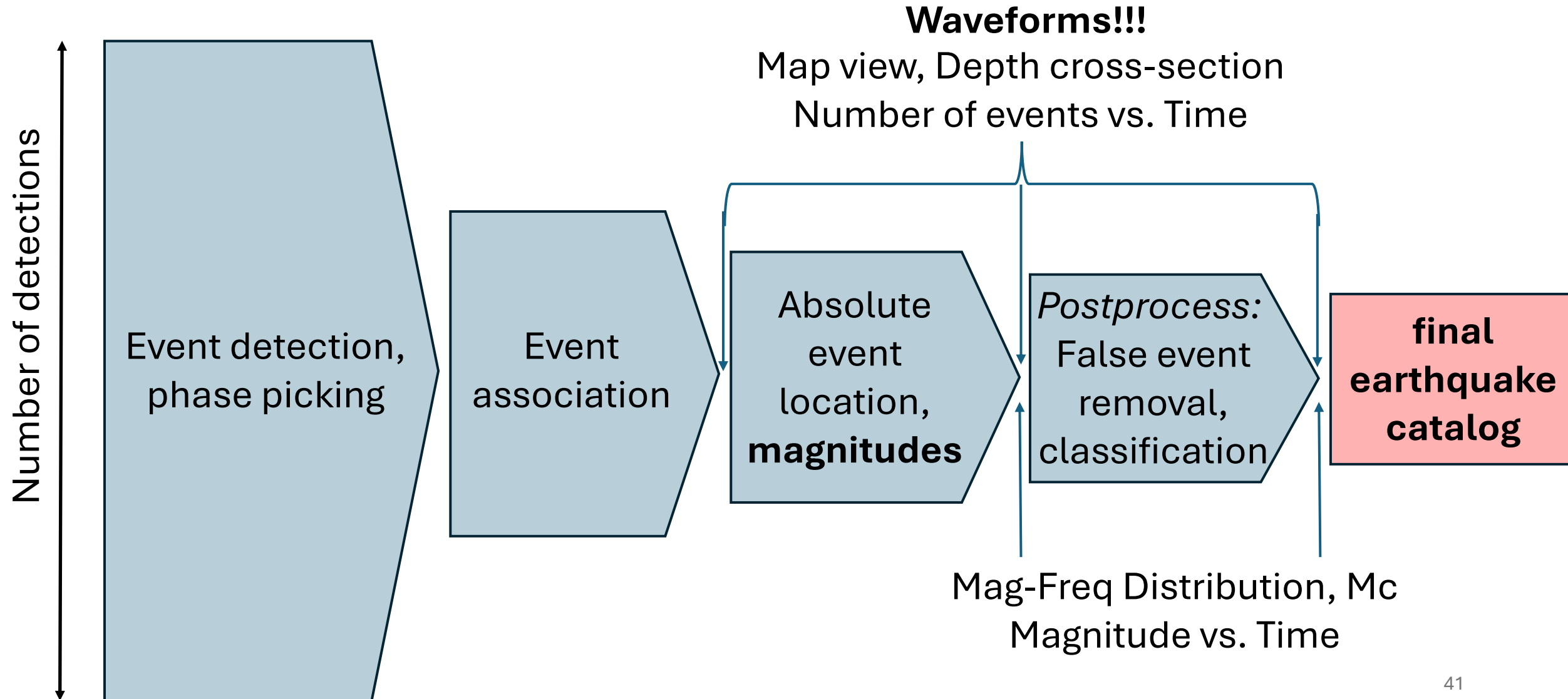


2019 M_w 6.4 & 7.1 Ridgecrest sequence



Catalog QC: Summary of things to check in plots

Compare against a reference catalog!



Extra slides

Earthquake Catalog QC General Tips

- Make plots - think like an analyst – use visual tools. Waveforms!
- Use seismology domain expertise
 - If enhanced catalog events don't look like real earthquakes, figure out why.
- Rule out all sources of noise (background or cultural). Be sure it's not an artifact.
- Check it's not a regional/teleaseismic earthquake.
- Check it's not a quarry blast, sonic boom, infrasound, active-source explosion, or other unwanted seismic signal.
- Only then have you discovered new earthquakes.

Reference Catalog: How to download?

- For bulk downloads of catalog events from ComCat
 - ComCat search API: <https://earthquake.usgs.gov/fdsnws/event/1/>
 - (I create the URL string, then call wget with a bash script)
 - Python libcomcat: <https://code.usgs.gov/ghsc/esi/libcomcat-python>
 - ObsPy get_events() with USGS Client (I save QuakeML files with picks): <https://docs.obspy.org/packages/obspy.clients.fdsn.html>
- For ANSS networks, all finalized local events are sent to ComCat. Can also download reference catalog from regional seismic network, which may contain extra events.
 - SCSN: Quarry blasts, teleseismic/regional events.
 - <https://scedc.caltech.edu/>
 - NCSN: Extra events – not all events are submitted to ComCat; catalogs available in other formats.
 - <https://www.ncedc.org/ncedc/catalog-search.html>

Alternative (non-authoritative) reference catalogs

- Regional seismic network websites (for ANSS networks) also have alternative (non-authoritative) earthquake catalogs and data sets for download, to use as reference catalogs
 - SCSN: template-matching, relocated, deep-learning, focal-mechanism, <https://scedc.caltech.edu/eq-catalogs/altcatalogs.html>, <https://scedc.caltech.edu/data/deeplearning.html>
 - NCSN: real-time double-difference. <https://www.ncedc.org/ncedc/catalog-search.html>

MATCH events: location, magnitude comparison

- Compare enhanced vs. reference catalog
 - Location (latitude, longitude, depth, origin-time)
 - Magnitude
 - Distribution of residuals (ideally zero) – are they low enough for you?
- It's difficult to exactly reproduce locations & magnitudes from ComCat (or other reference catalog)
 - Differences in monitoring software, velocity model, manual pick times/weights, filtering before amplitudes, ...
 - Just check that these values are close enough for your needs.

Reference Catalog Tip: Select boundaries carefully

- Carefully select the bounding box or circular radius for downloading your reference catalog, before comparing it to enhanced catalog. It should include:
 - Entire seismogenic region of interest to you
 - Check past seismicity locations on web map: <http://ds.iris.edu/ieb/> or <https://earthquake.usgs.gov/earthquakes/map/>
 - All seismic stations you plan to use for catalog generation
 - Check station locations and availability on web map: <http://ds.iris.edu/gmap/> or <https://www.fdsn.org/networks/>
- **NEW events** in your enhanced catalog might be from reference catalog events outside your selected bounding region
 - Check waveforms: could be a regional distance earthquake if higher duration and amplitude

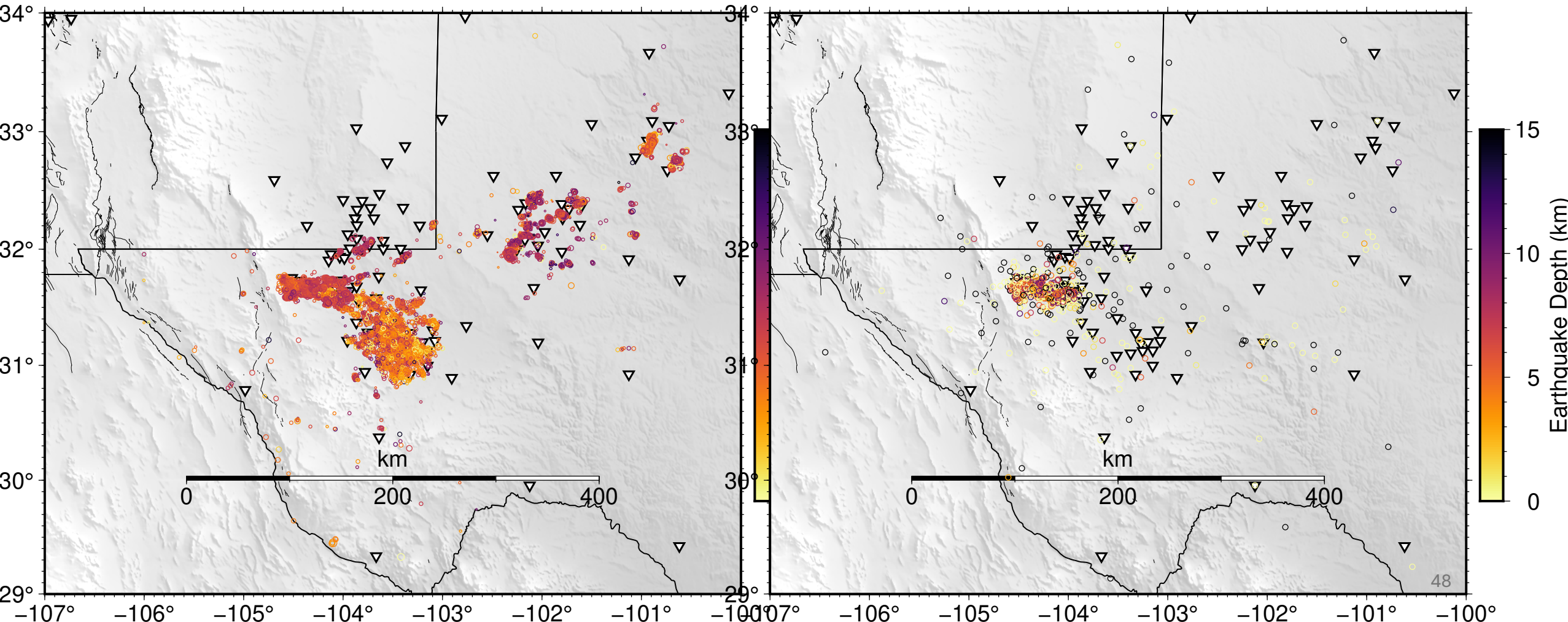
Catalog QC, **locations in map view**: TexNet



Many scattered event locations –
possible false detections?

Reference catalog:
TexNet (2015-2024: 9 years)

Enhanced catalog: 2023-12-31 (1 day)
GaMMA associator (3 P, 3 S, 6 total picks)



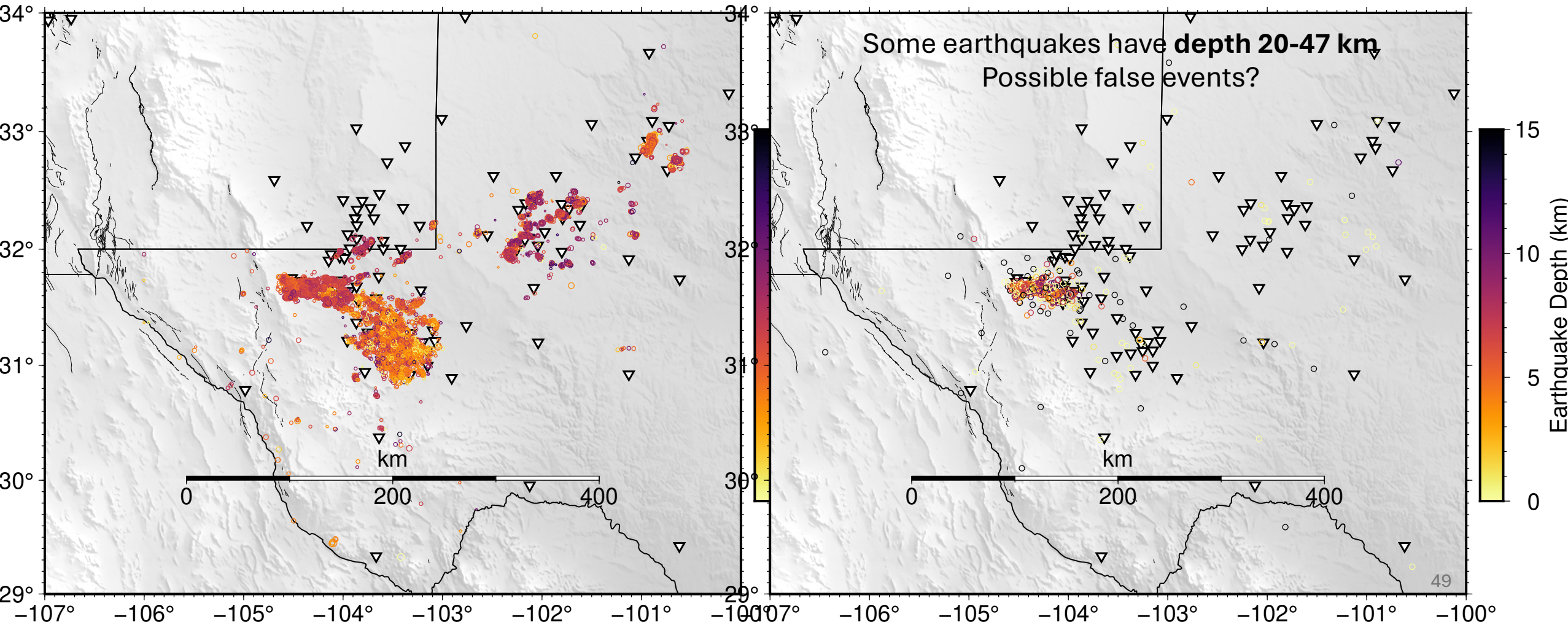
Catalog QC, **locations in map view**: TexNet



Removed false detections (fewer scattered events)
with stricter association criteria

Reference catalog:
TexNet (2015-2024: 9 years)

Enhanced catalog: 2023-12-31 (1 day)
GaMMA associator (4 P, 4 S, 8 total picks)

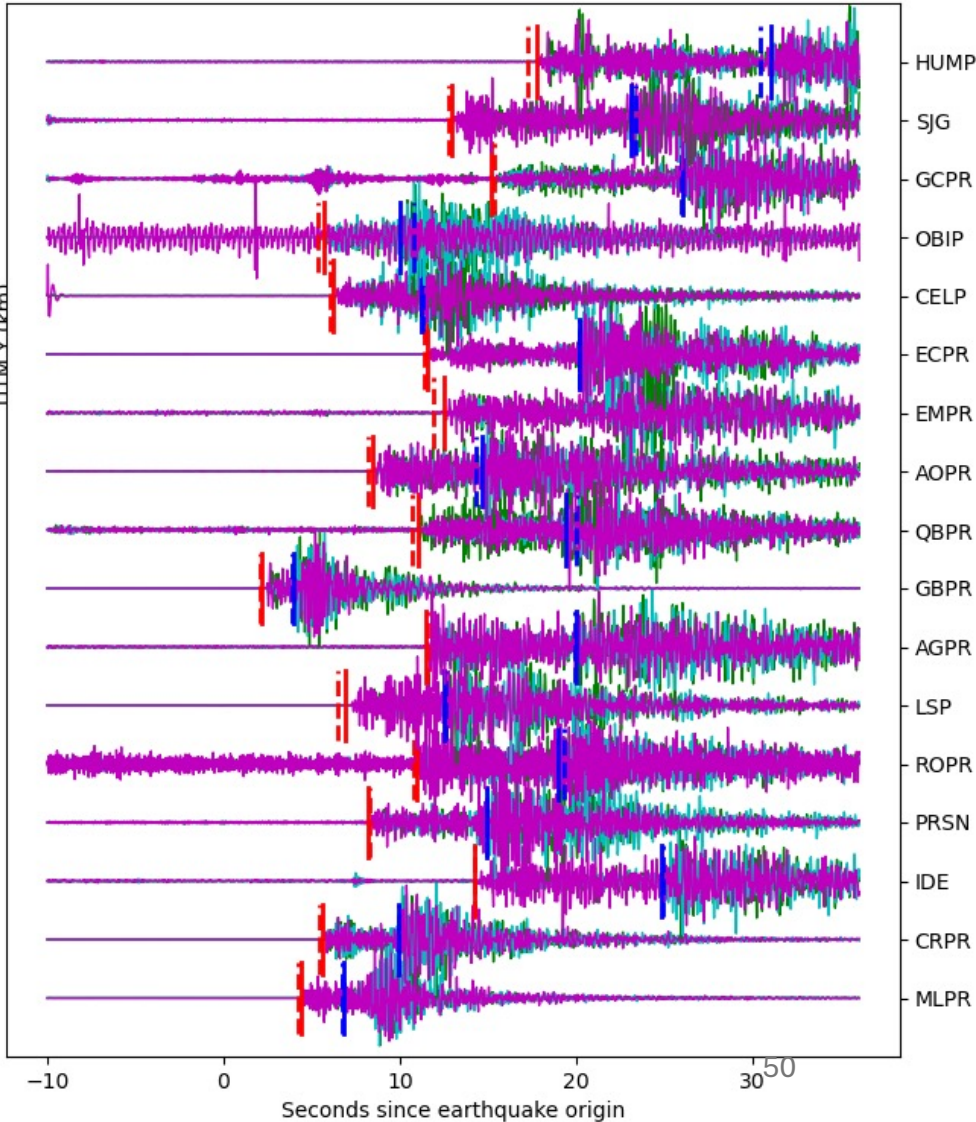
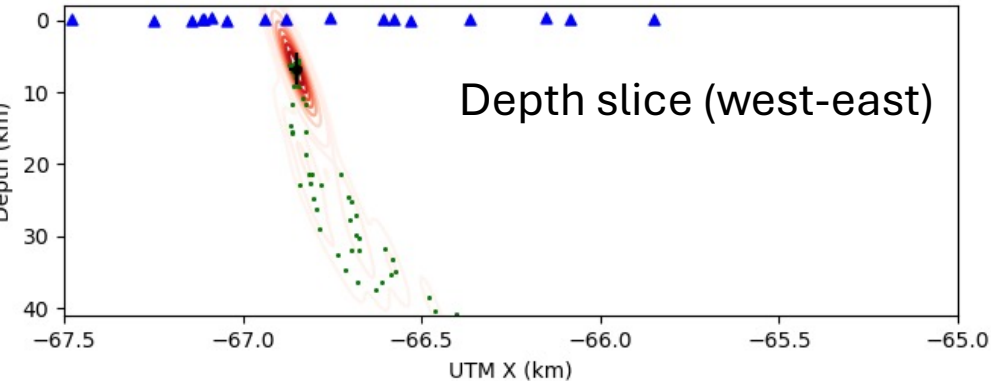
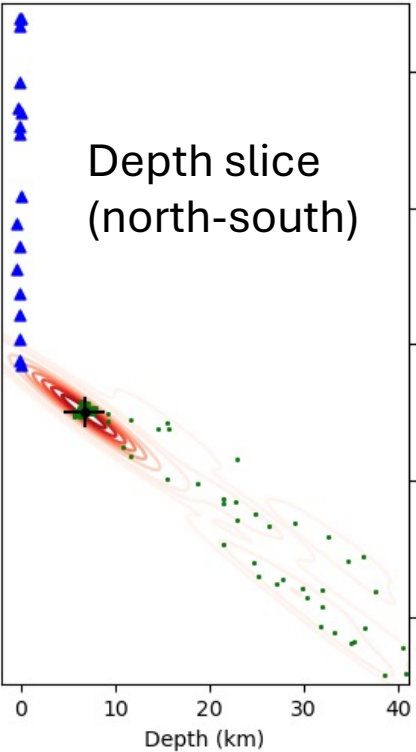
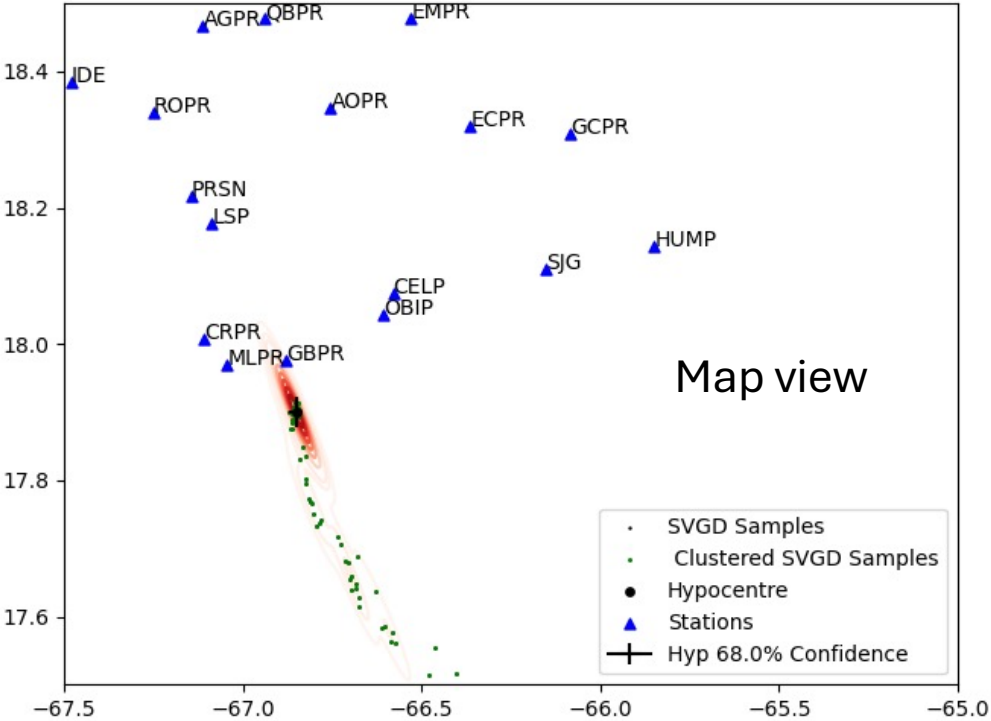


Catalog QC: 2020 M2.8 earthquake, southwest Puerto Rico



Earthquake 2020-01-07 05:59:47.516874763 +/- 0.16s
Hyp=[-66.85,17.90,6.72] - Hyp Uncertainty (km) +/- [1.99,2.48,2.14]

Event waveforms with **P**, **S** picks
Solid: auto-picks, Dashed: predicted arrivals



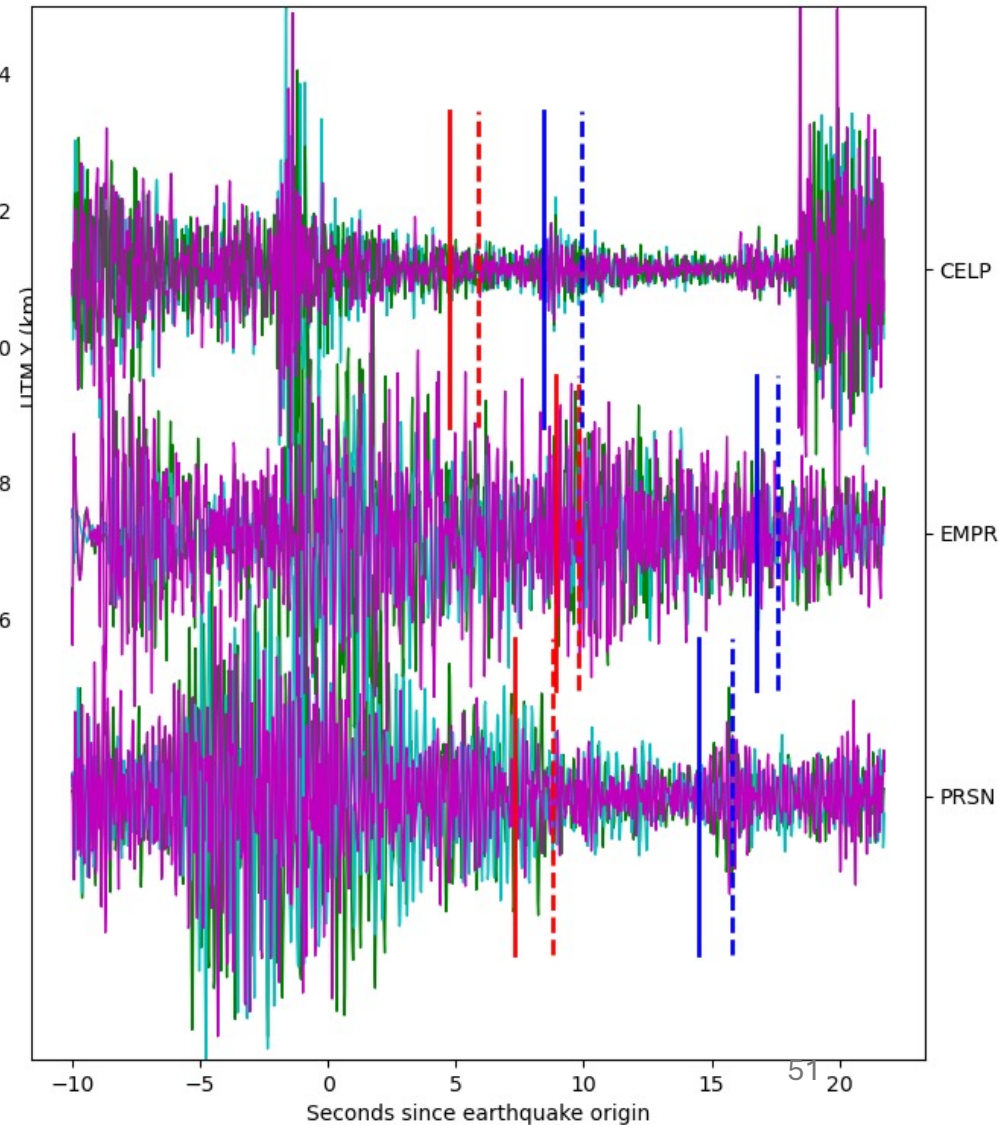
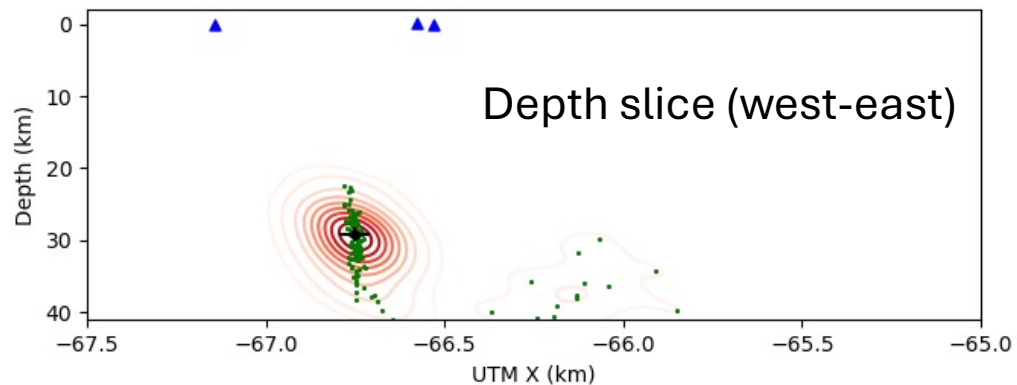
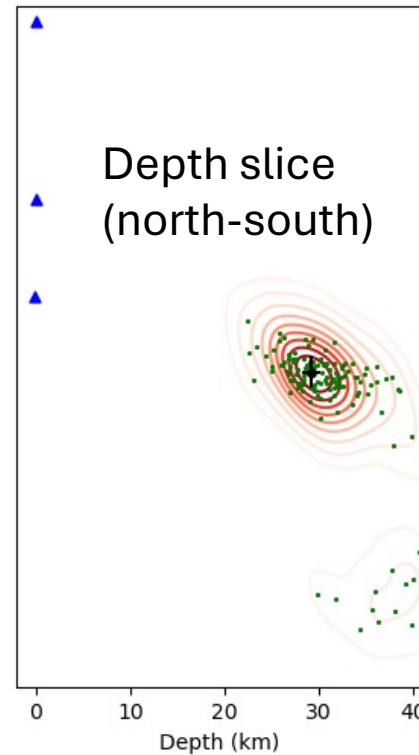
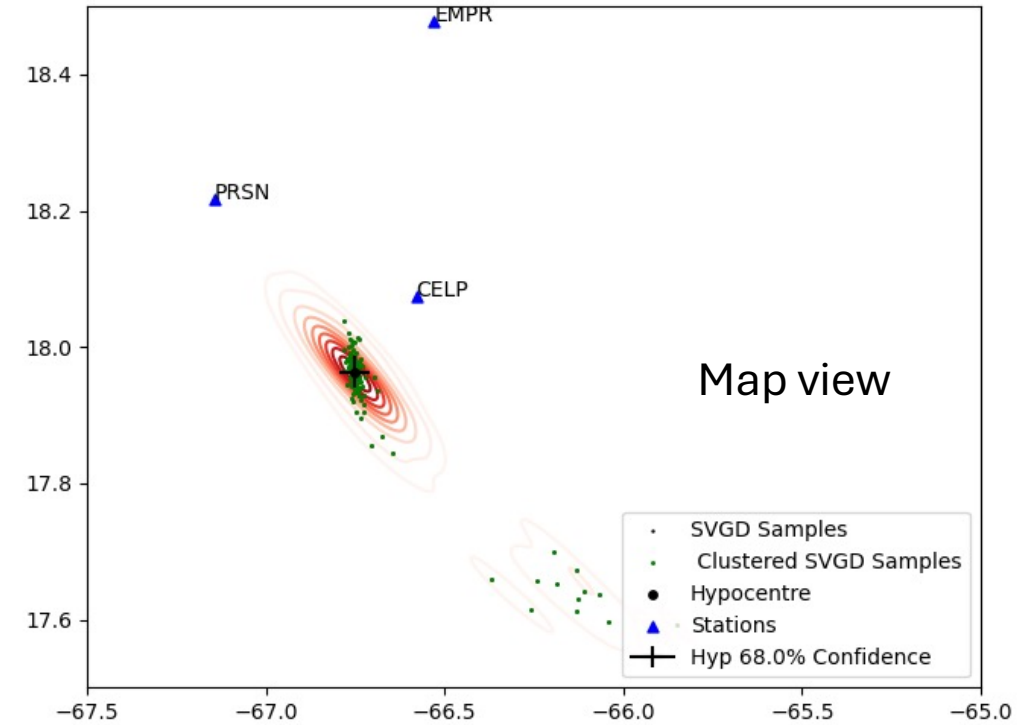
Catalog QC: False detection in coda of larger earthquake (also, too deep?)

2020-01-07T08:29:36.930000 30576.930000 17.963500 -66.752333 **29.160000** **4.480000** 1000330

Earthquake 2020-01-07 08:29:36.925860007 +/- 1.21s
Hyp=[-66.75,17.96,29.16] - Hyp Uncertainty (km) +/- [4.69,2.51,0.93]

Event waveforms with **P**, **S** picks

Solid: auto-picks, Dashed: predicted arrivals



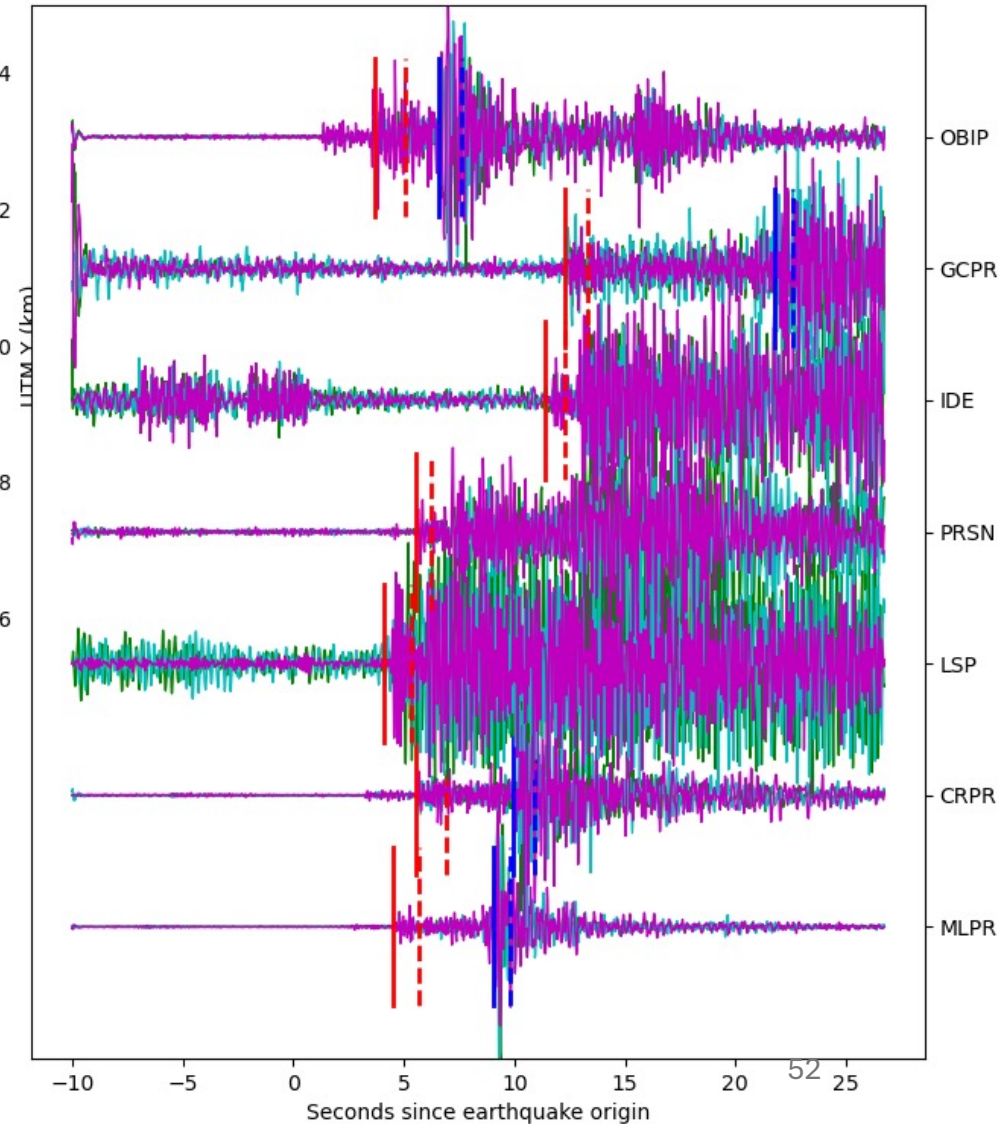
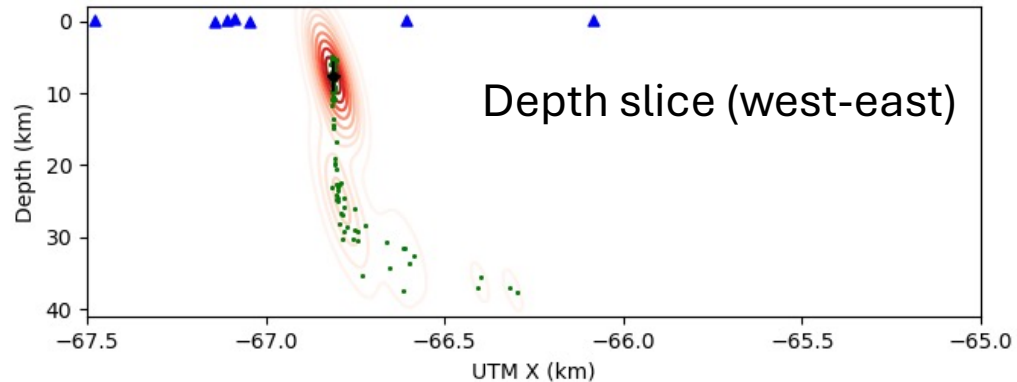
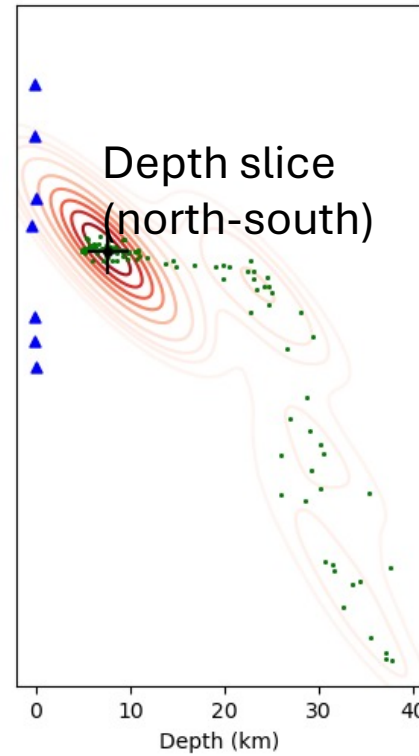
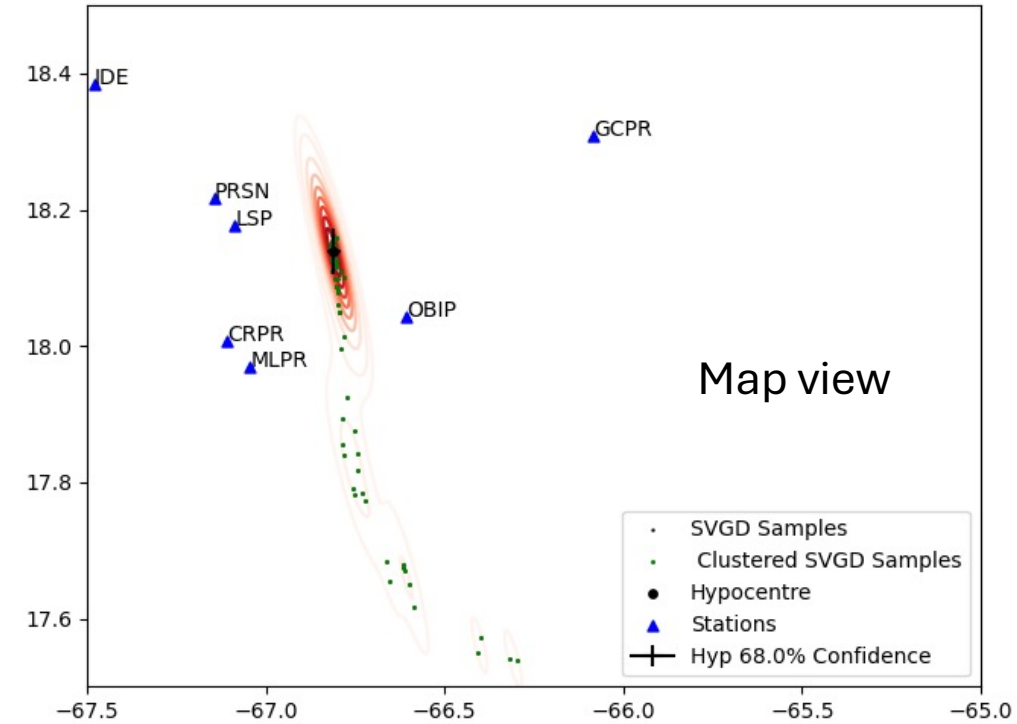
Catalog QC: Real (newly detected) earthquake

2020-01-07T08:29:18.320000 30558.320000 18.139667 -66.810000 **7.660000** **4.440000** 1000328

Earthquake 2020-01-07 08:29:18.318609738 +/- 1.01s
Hyp=[-66.81,18.14,7.66] - Hyp Uncertainty (km) +/- [2.00,3.72,2.12]

Event waveforms with **P**, **S** picks

Solid: auto-picks, Dashed: predicted arrivals



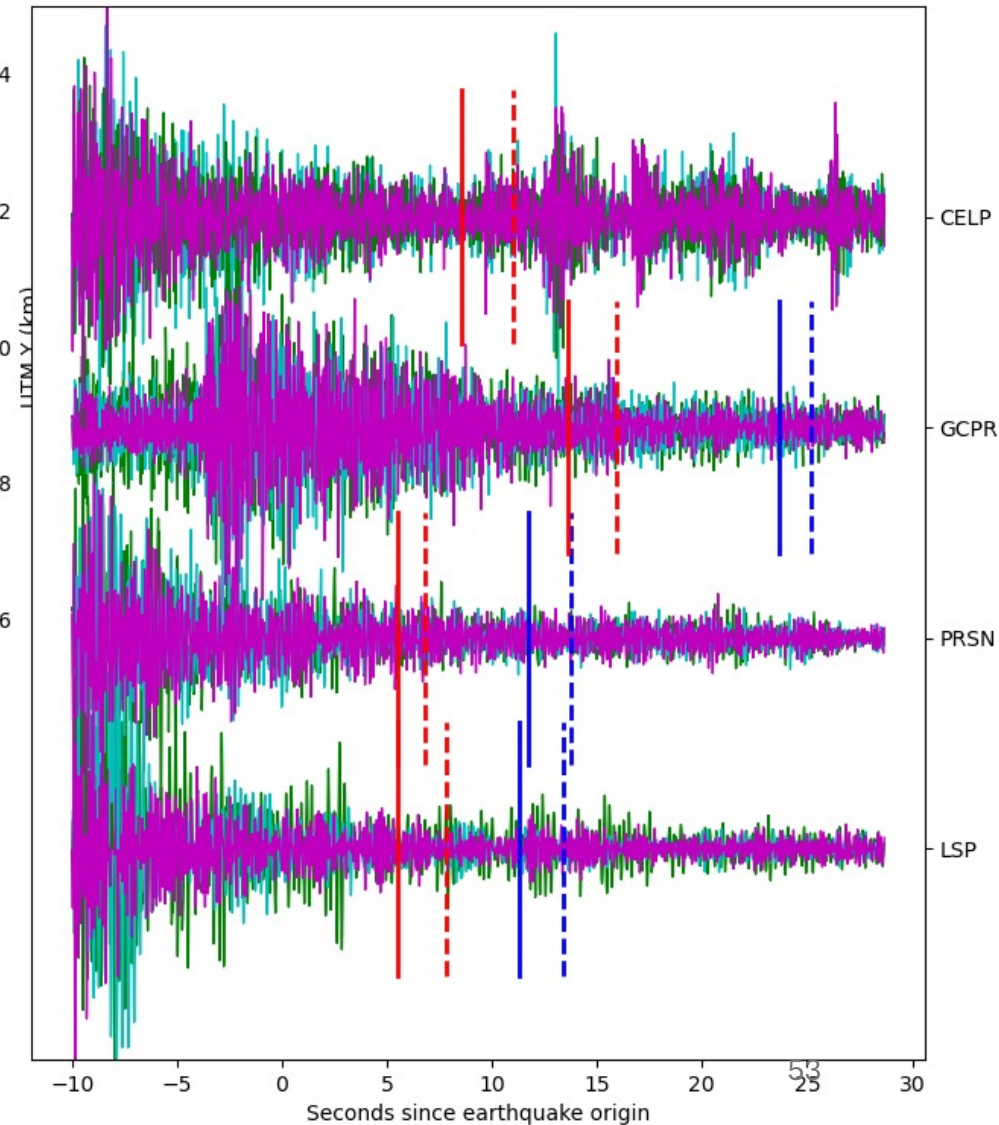
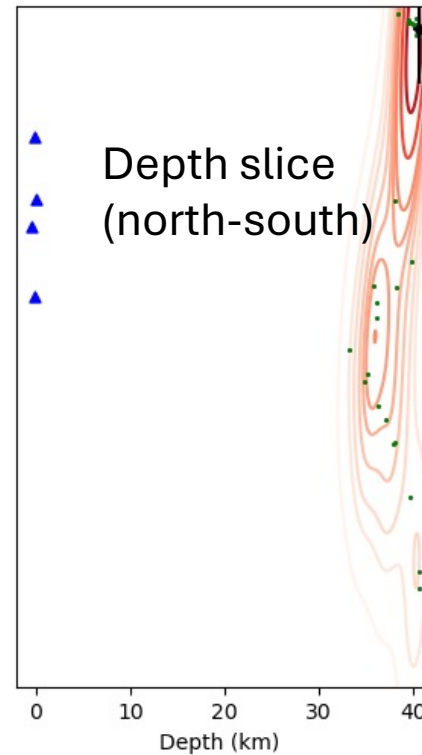
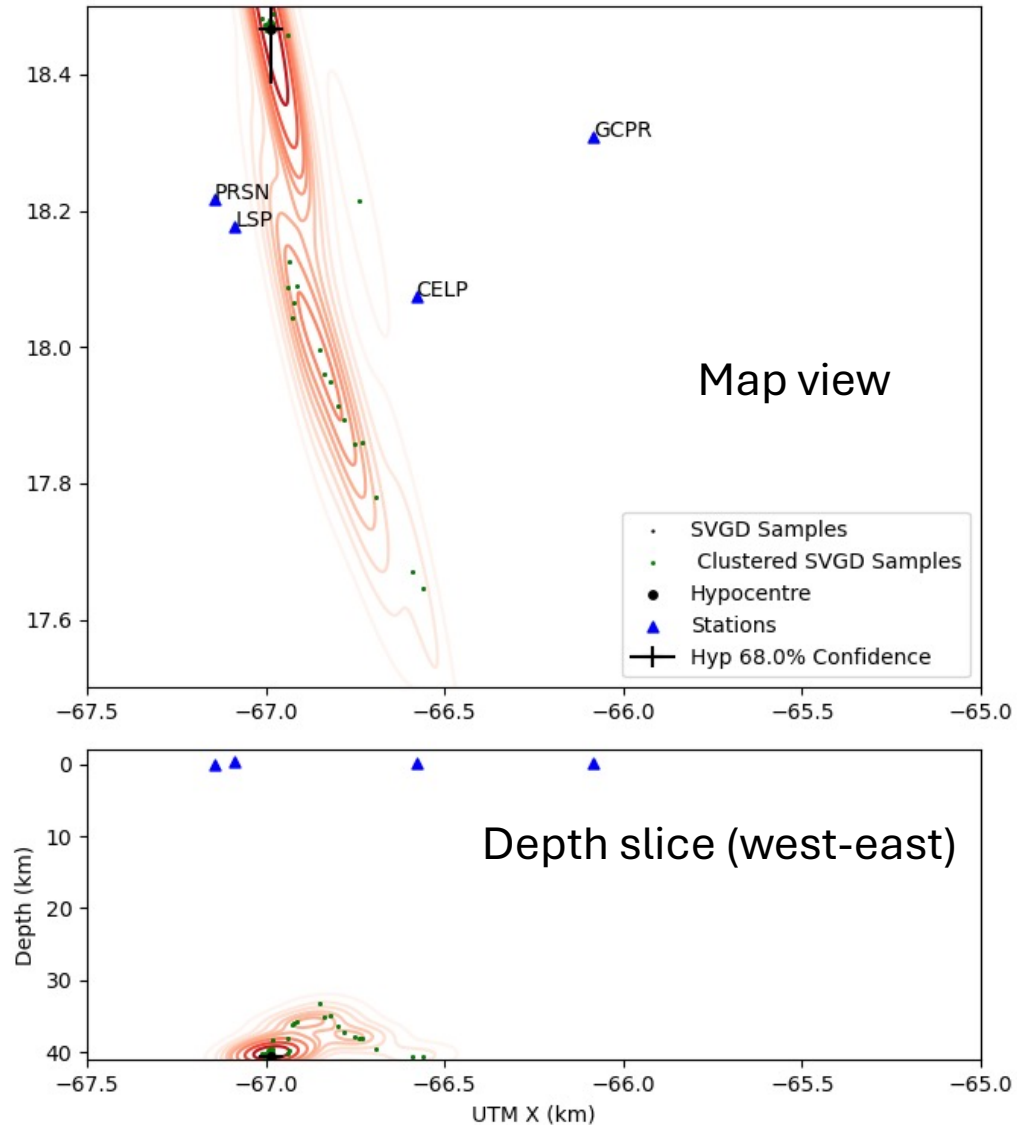
Catalog QC: False detection in coda of larger earthquake (also, too deep?)

2020-01-07T22:21:53.870000 80513.870000 18.466000 -66.986833 **40.580000** **2.170000** 1001466

Earthquake 2020-01-07 22:21:53.871442969 +/- 2.08s
Hyp=[-66.99,18.47,40.58] - Hyp Uncertainty (km) +/- [3.84,8.85,0.59]

Event waveforms with **P**, **S** picks

Solid: auto-picks, Dashed: predicted arrivals



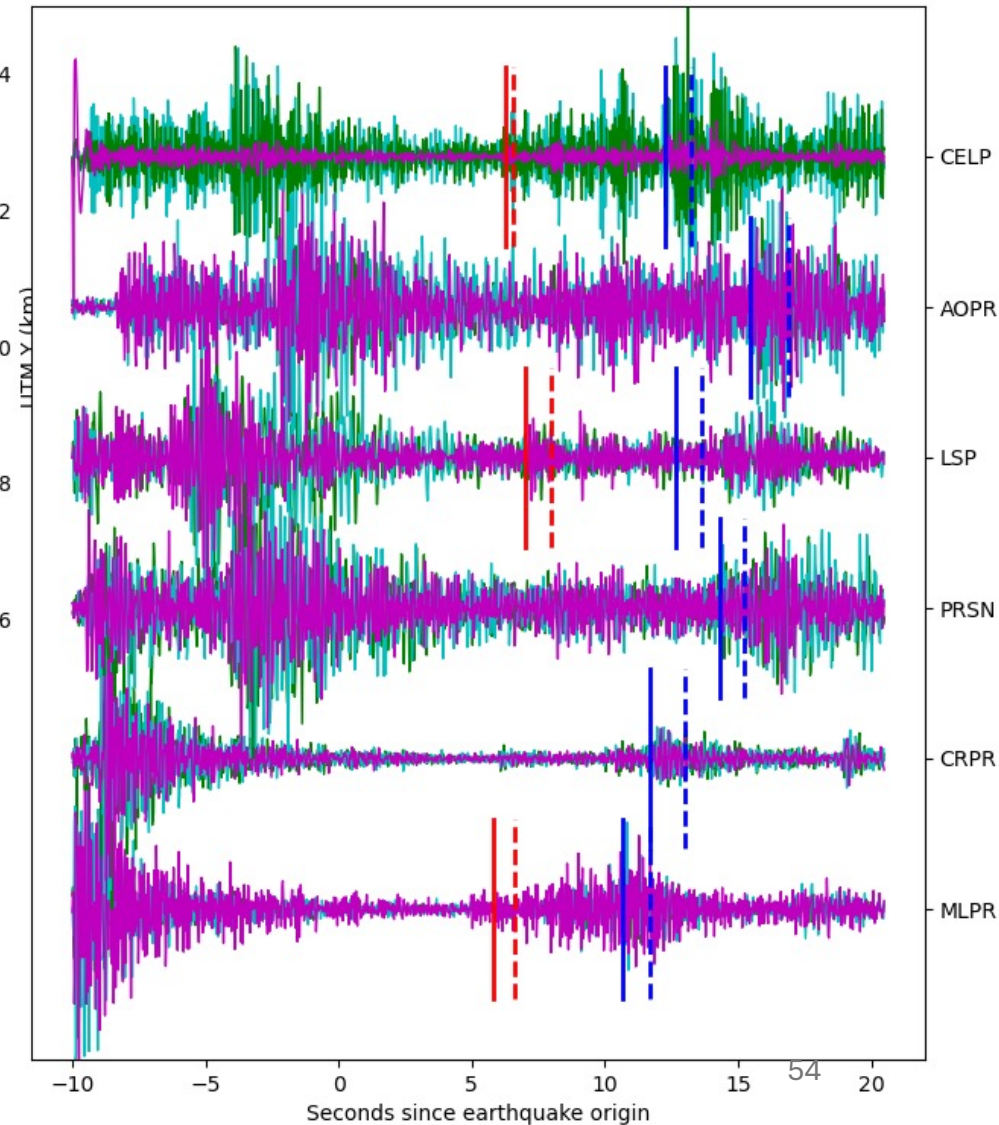
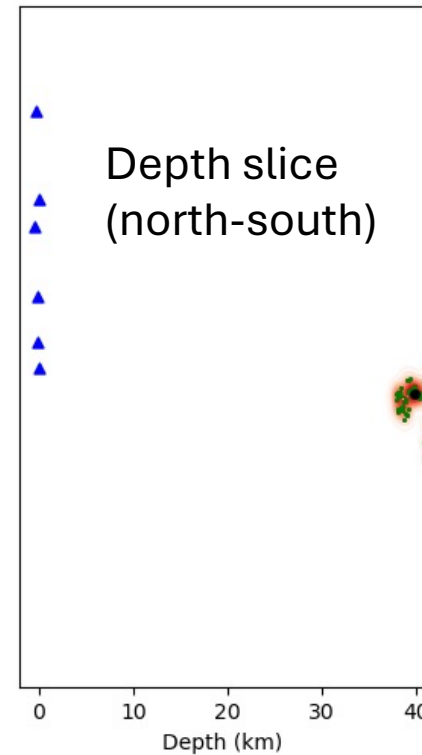
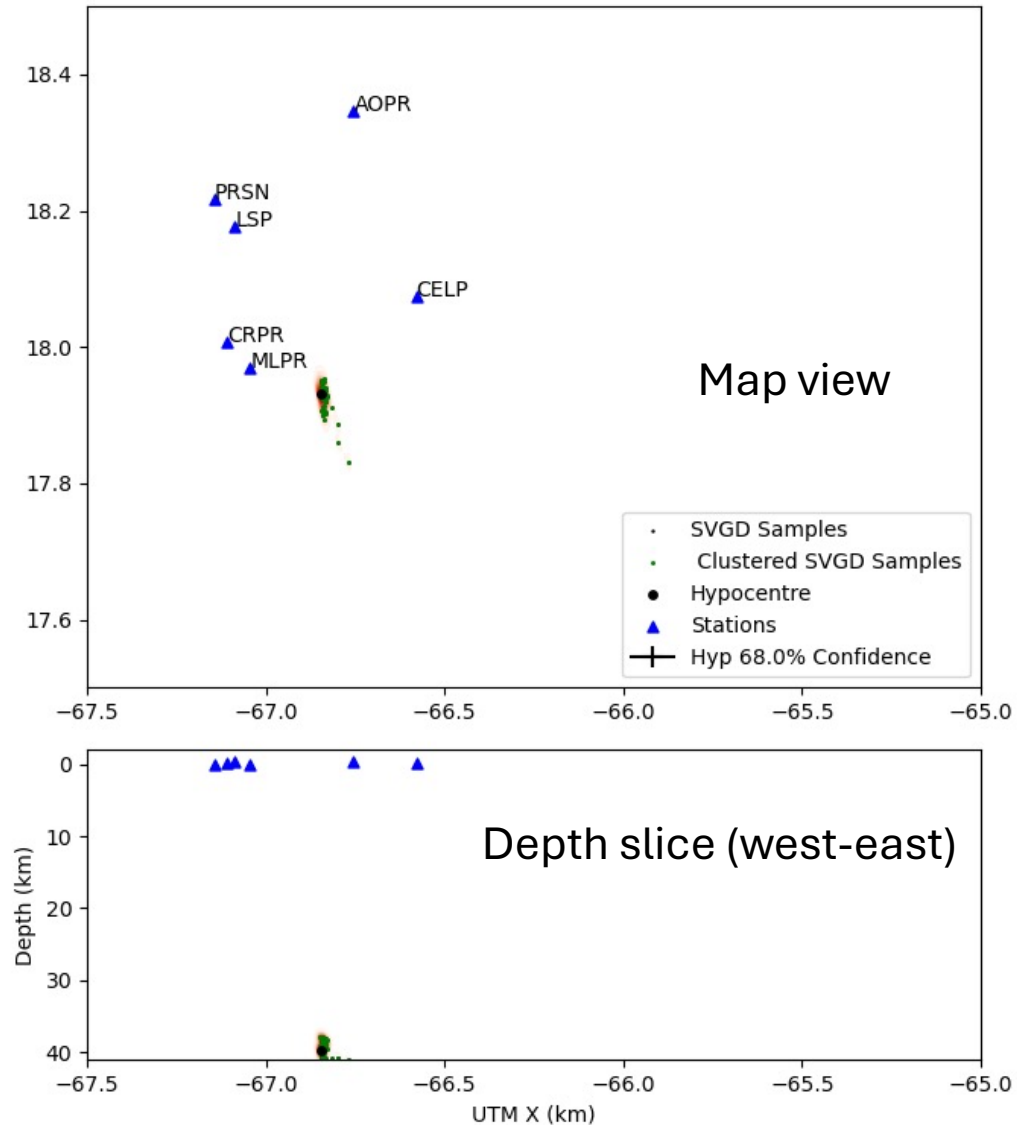
Catalog QC: False detection in coda of larger earthquake (also, too deep?)

2020-01-07T20:27:22.970000 73642.970000 17.931000 -66.846500 **39.800000** **1.560000** 1001320

Earthquake 2020-01-07 20:27:22.965268426 +/- 0.97s
Hyp=[-66.85,17.93,39.80] - Hyp Uncertainty (km) +/- [0.36,0.59,0.24]

Event waveforms with **P**, **S** picks

Solid: auto-picks, Dashed: predicted arrivals



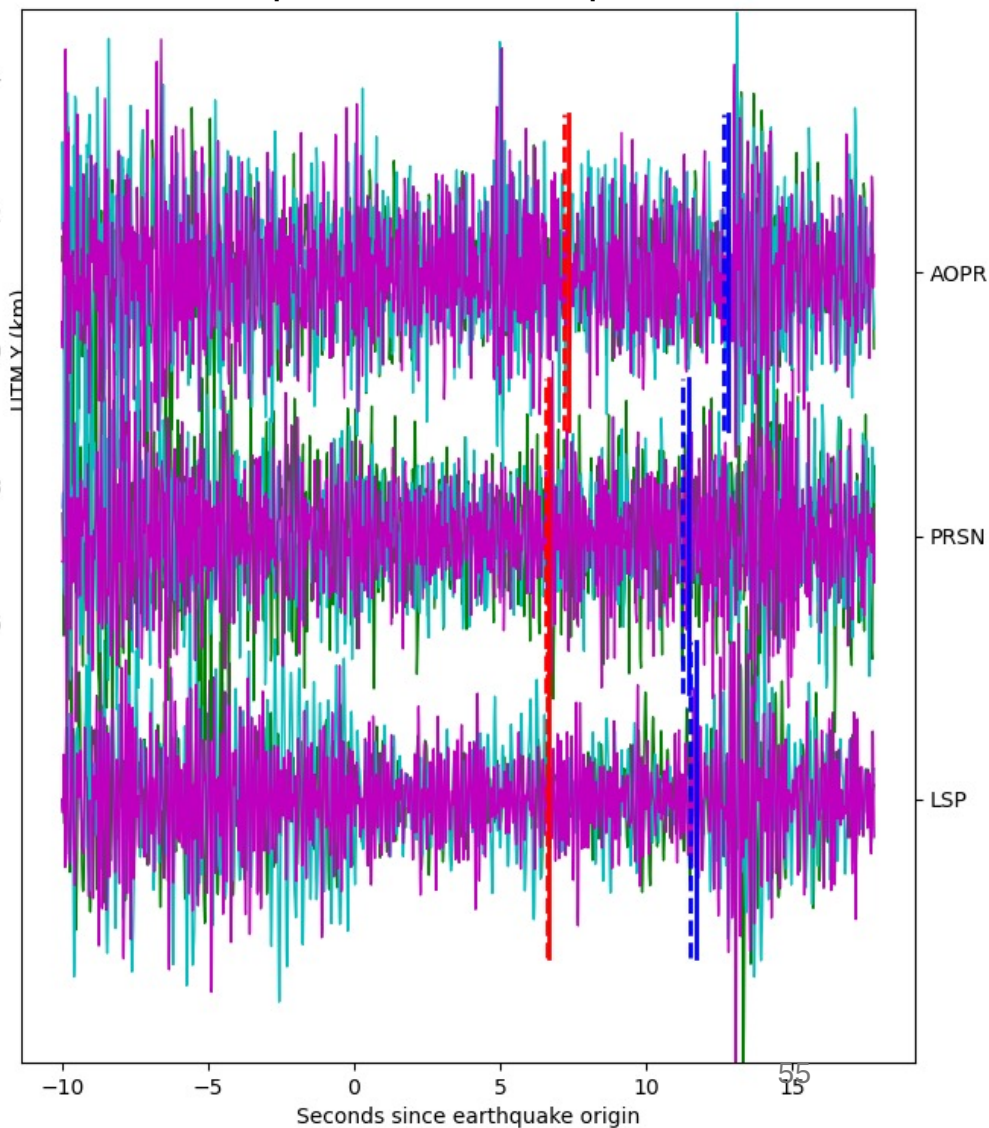
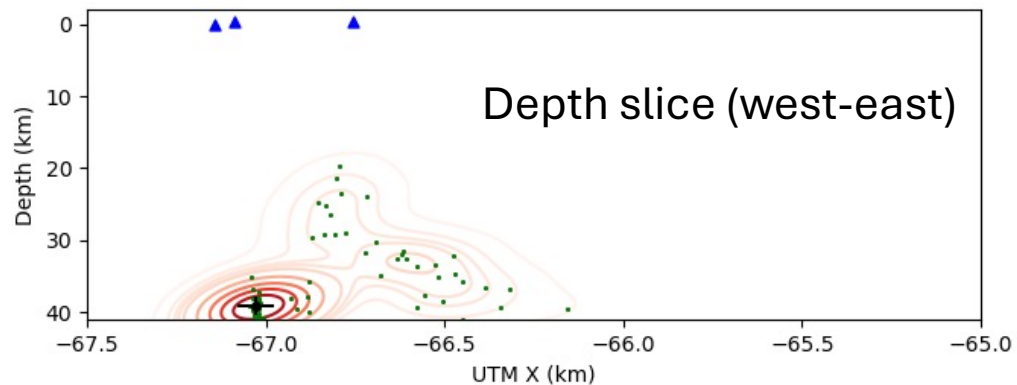
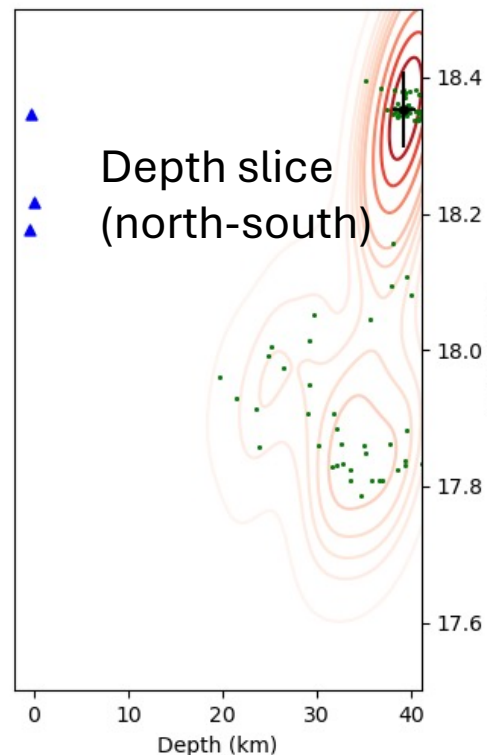
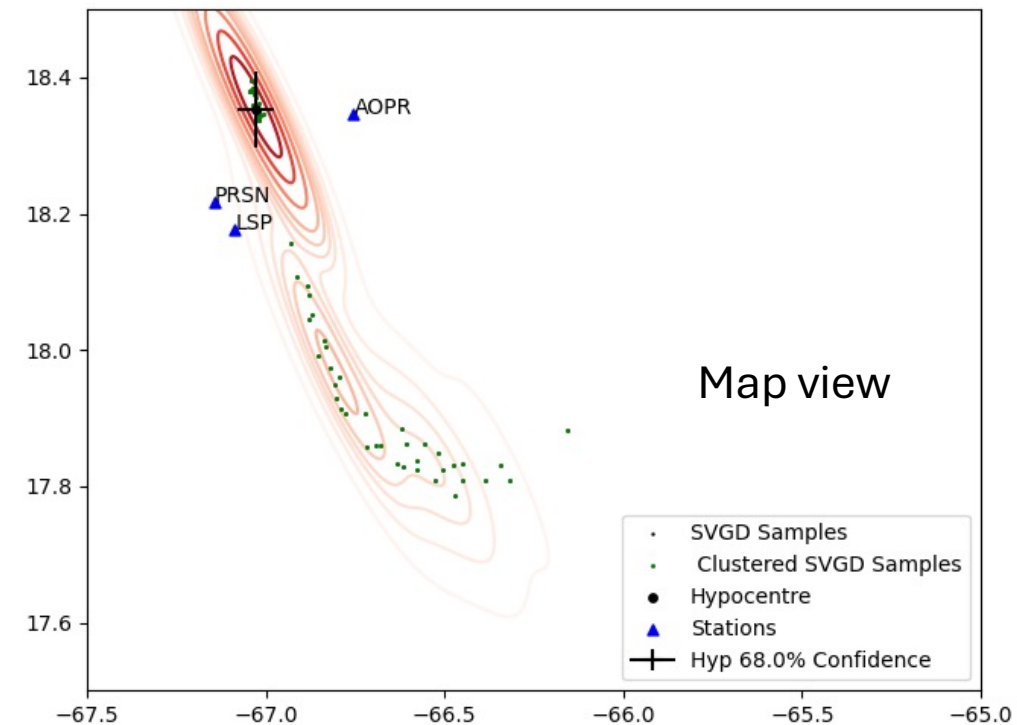
Catalog QC: False detection in noise (also, too deep?)

2020-01-07T15:58:52.370000 57532.370000 18.353667 -67.028000 **39.160000 1.310000** 1000980

Earthquake 2020-01-07 15:58:52.367385814 +/- 0.17s
Hyp=[-67.03,18.35,39.16] - Hyp Uncertainty (km) +/- [5.86,6.19,1.22]

Event waveforms with **P**, **S** picks

Solid: auto-picks, Dashed: predicted arrivals



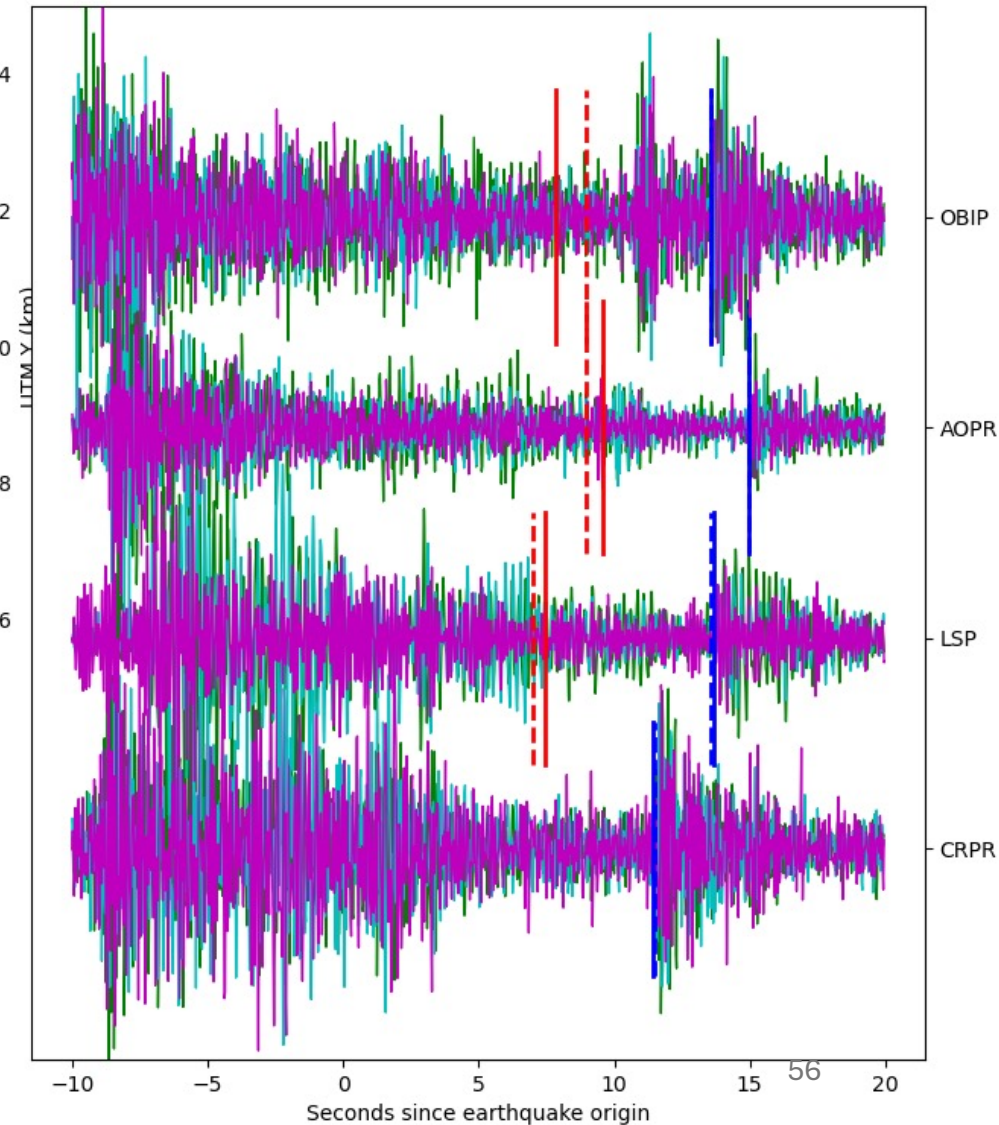
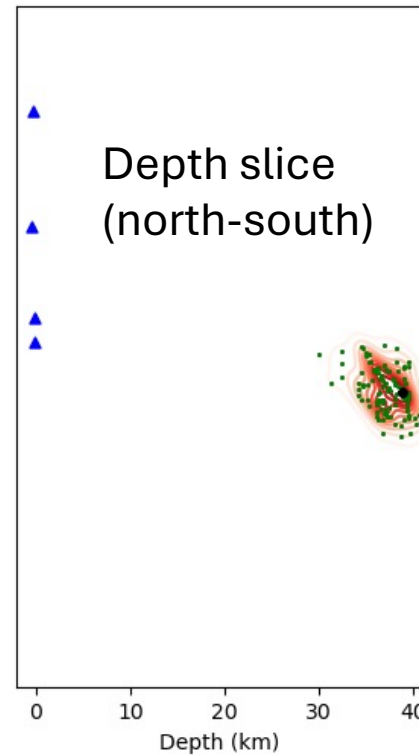
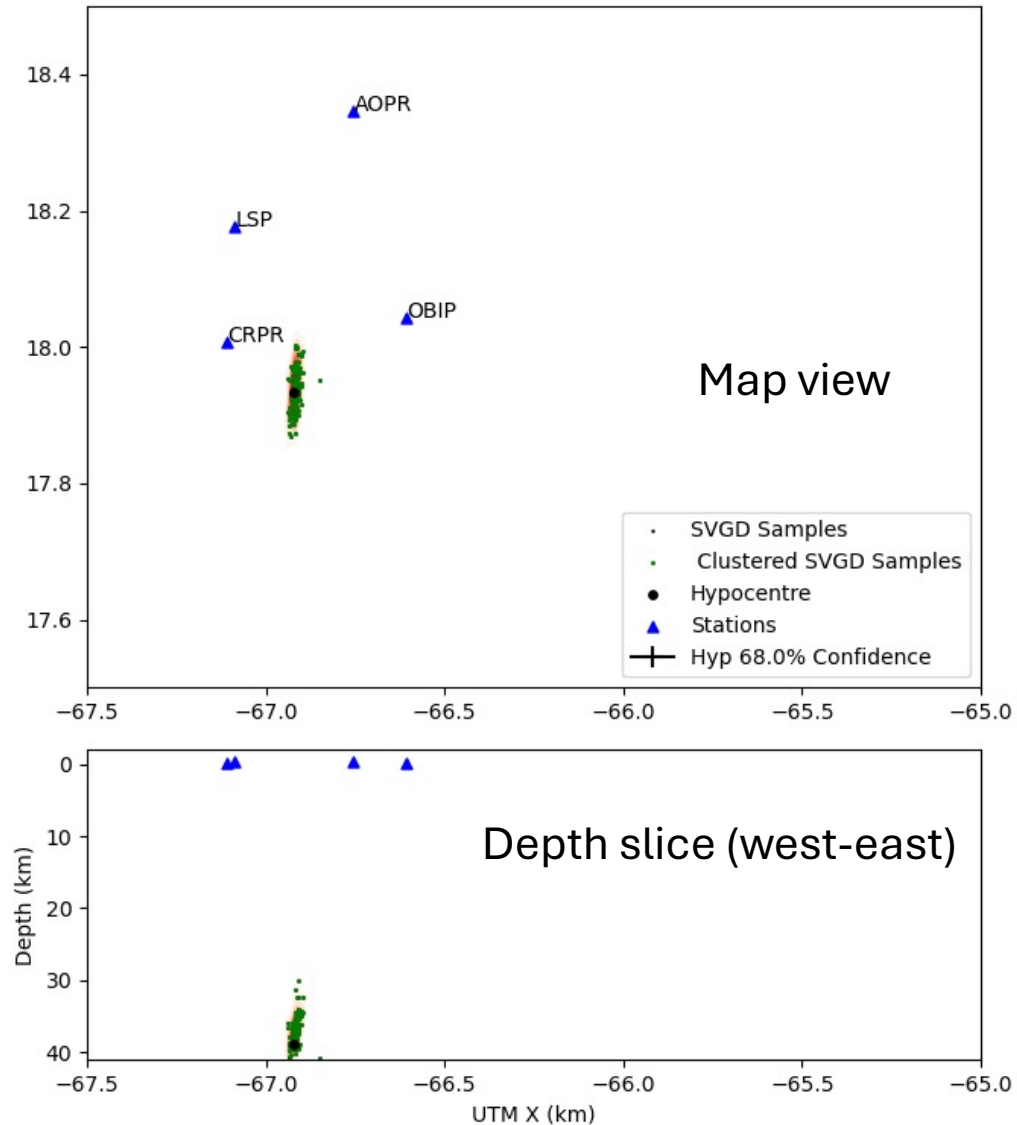
Catalog QC: False detection in coda of larger earthquake (also, too deep?)

2020-01-07T20:43:26.700000 74606.700000 17.933500 -66.923333 **38.810000** **1.570000** 1001342

Earthquake 2020-01-07 20:43:26.699477316 +/- 0.08s
Hyp=[-66.92,17.93,38.81] - Hyp Uncertainty (km) +/- [0.27,0.82,0.48]

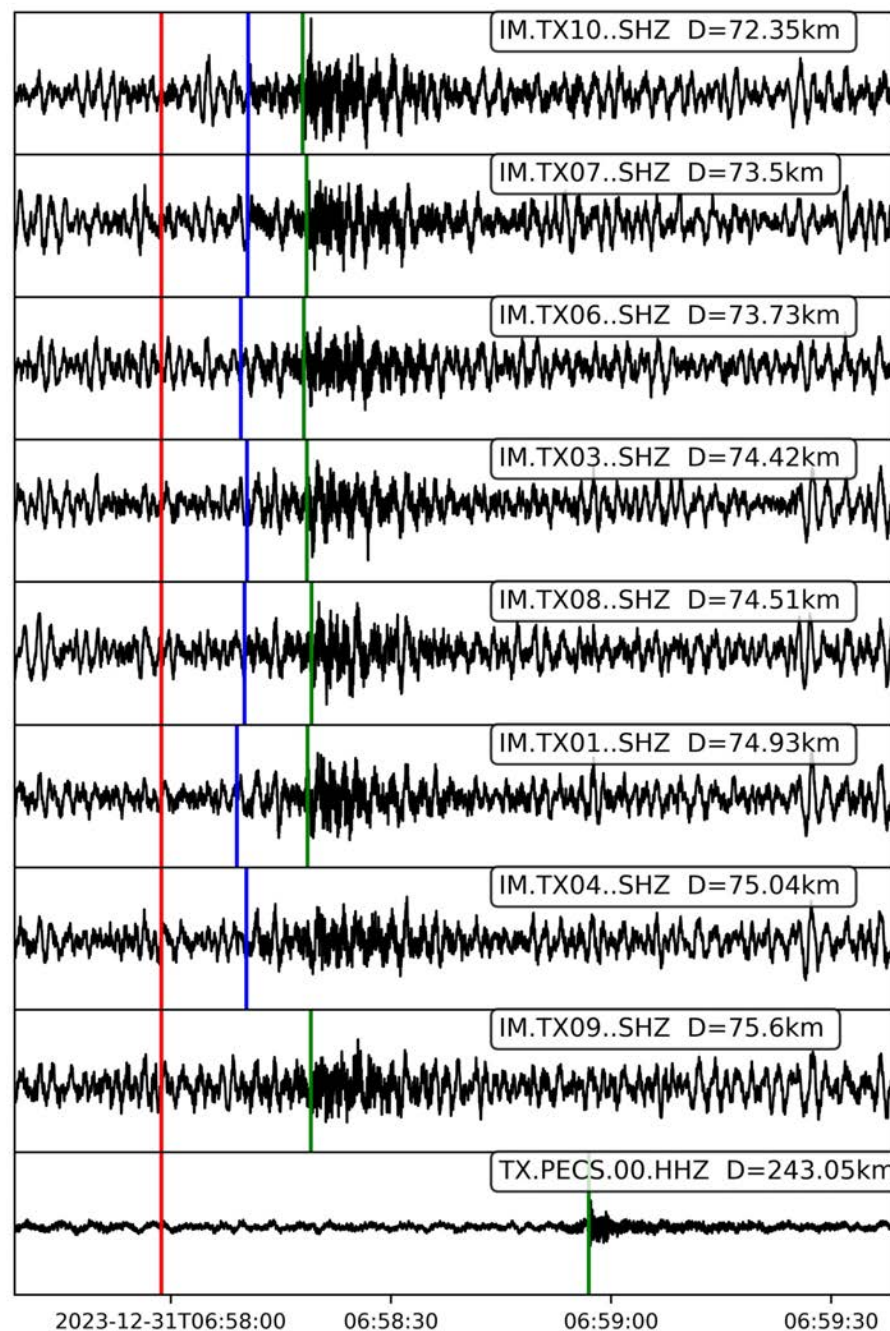
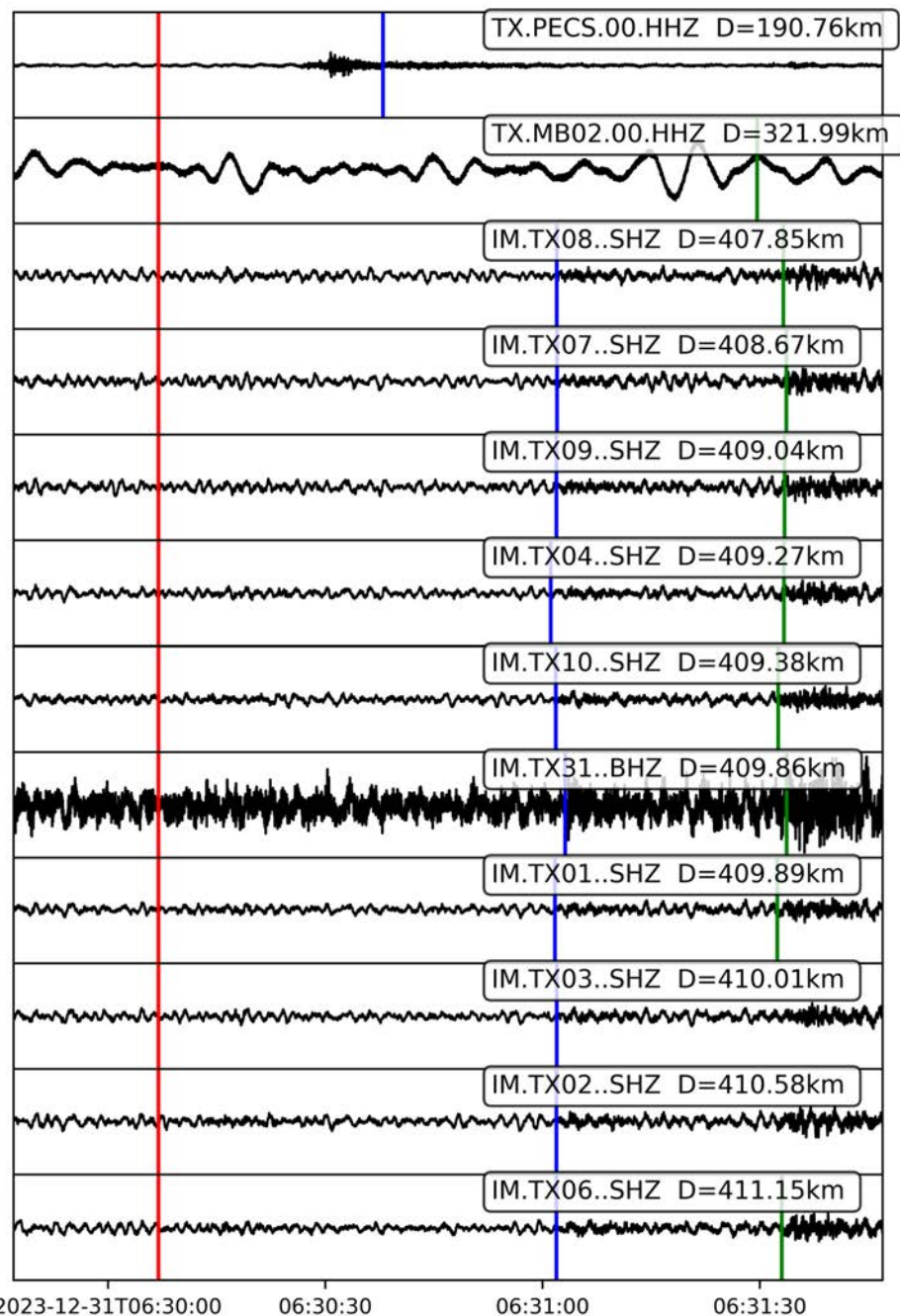
Event waveforms with **P**, **S** picks

Solid: auto-picks, Dashed: predicted arrivals



PyOcto Event #79 : 6/6/12 No Match to IRIS or GaMMA
Red: Event, Blue: P Pick, Green: S Pick

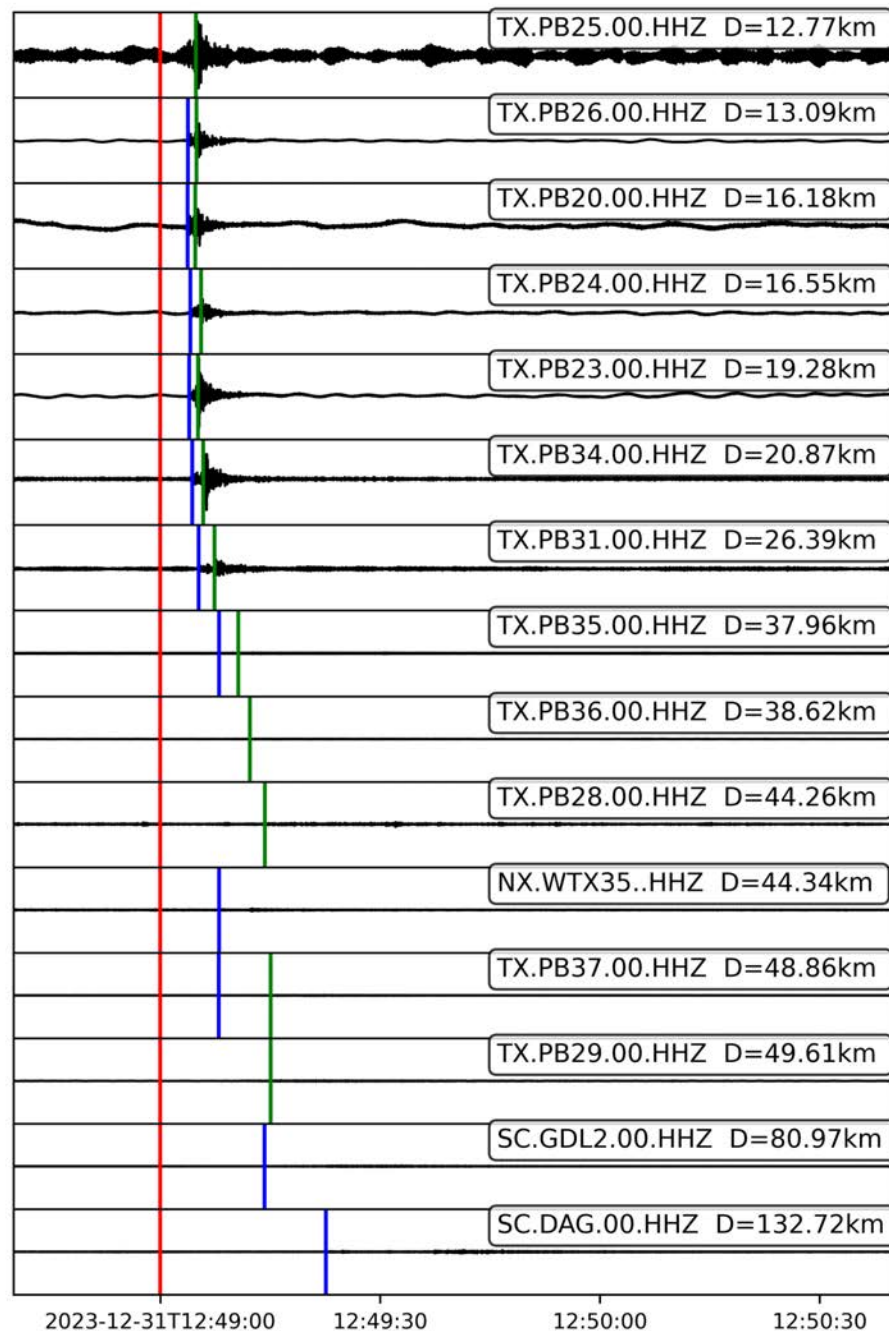
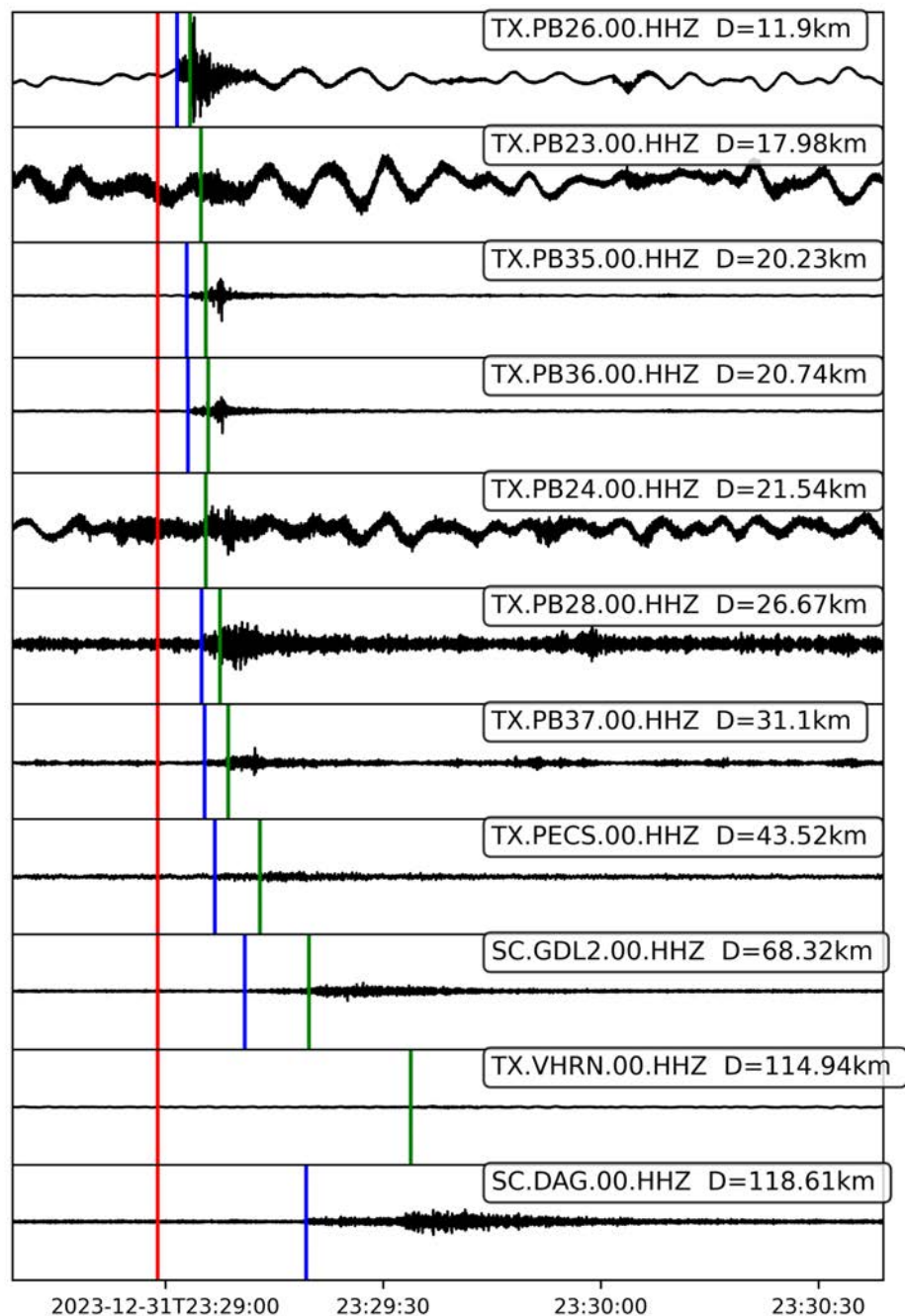
PyOcto Event #83 : 6/6/12 No Match to IRIS or GaMMA
Red: Event, Blue: P Pick, Green: S Pick



- Catalog QC: West Texas
- Bad event associations – dominated by array
- Notice large distance to nearest station

2-minute waveforms

Vertical lines:
Origin time, P, S



- Catalog QC: West Texas

- **Good** event association examples

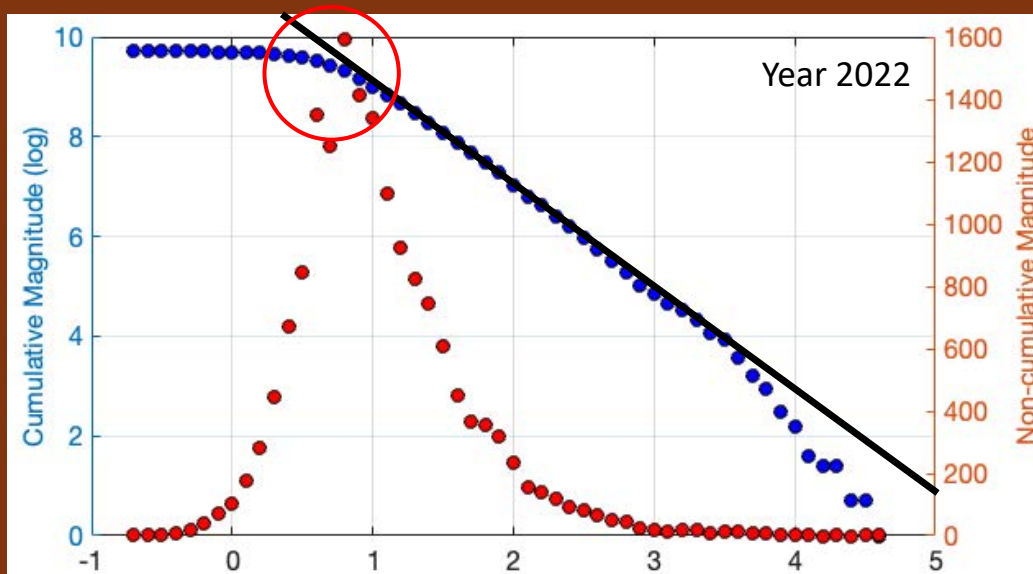
- P, S picks on earthquake signals with expected moveout

2-minute waveforms

Vertical lines:
 Origin time, P, S

Magnitude of Completeness

- lowest magnitude at which the catalog is “complete” (based on Gutenberg-Richter relation)
- examine frequency-magnitude distribution obtained by binning earthquake magnitudes
- one approach for evaluating the quality of a catalog



Can consider:

- temporal changes of M_c
- spatial variations of M_c across region

* More info and original R codes from CORSSA: [doi:10.5078/corssa-0018080](https://doi.org/10.5078/corssa-0018080)

* Matlab codes available on Github:
https://github.com/gtepp/research_codes

Calculated with different methods, including

- Maximum Curvature: max value of 1st derivative of FMD (typically, max bin of non-cum FMD)
 - * most reliable for small sample sizes (<~50-100 events)
 - * works best with local datasets with fewer heterogeneities
- Goodness-of-fit Test: lowest magnitude cut-off where Gutenberg-Richter relation holds
- Mc by b-value Stability: first magnitude increment where $|b_{\text{avg}} - b| < \text{uncertainty of } b$
 - * tends to produce the highest (i.e., most conservative) Mc

Bootstrapping

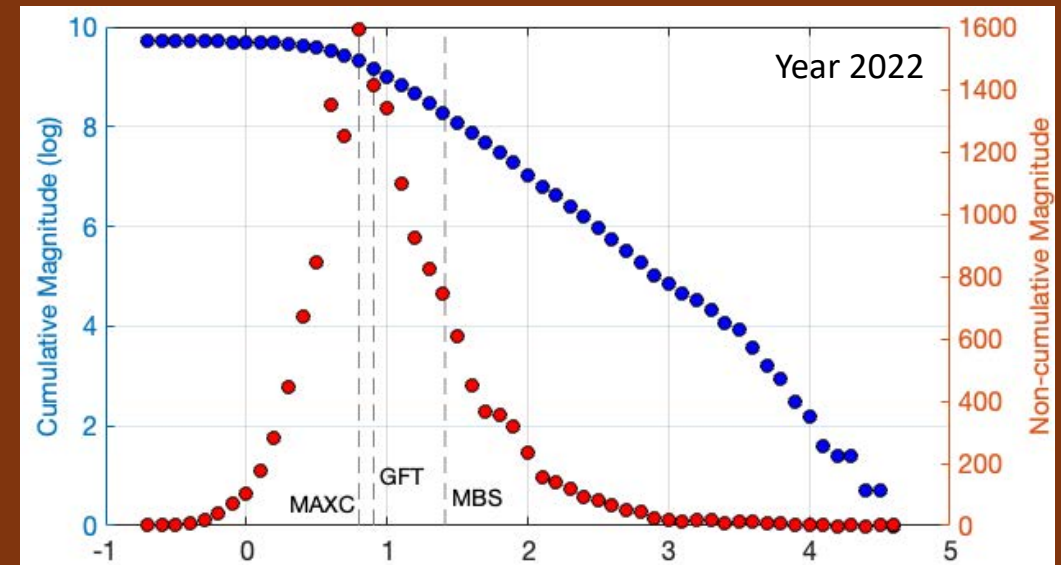
- gives sense of variation from sampling
- 200 sets with resampling
 - mean and st. dev.

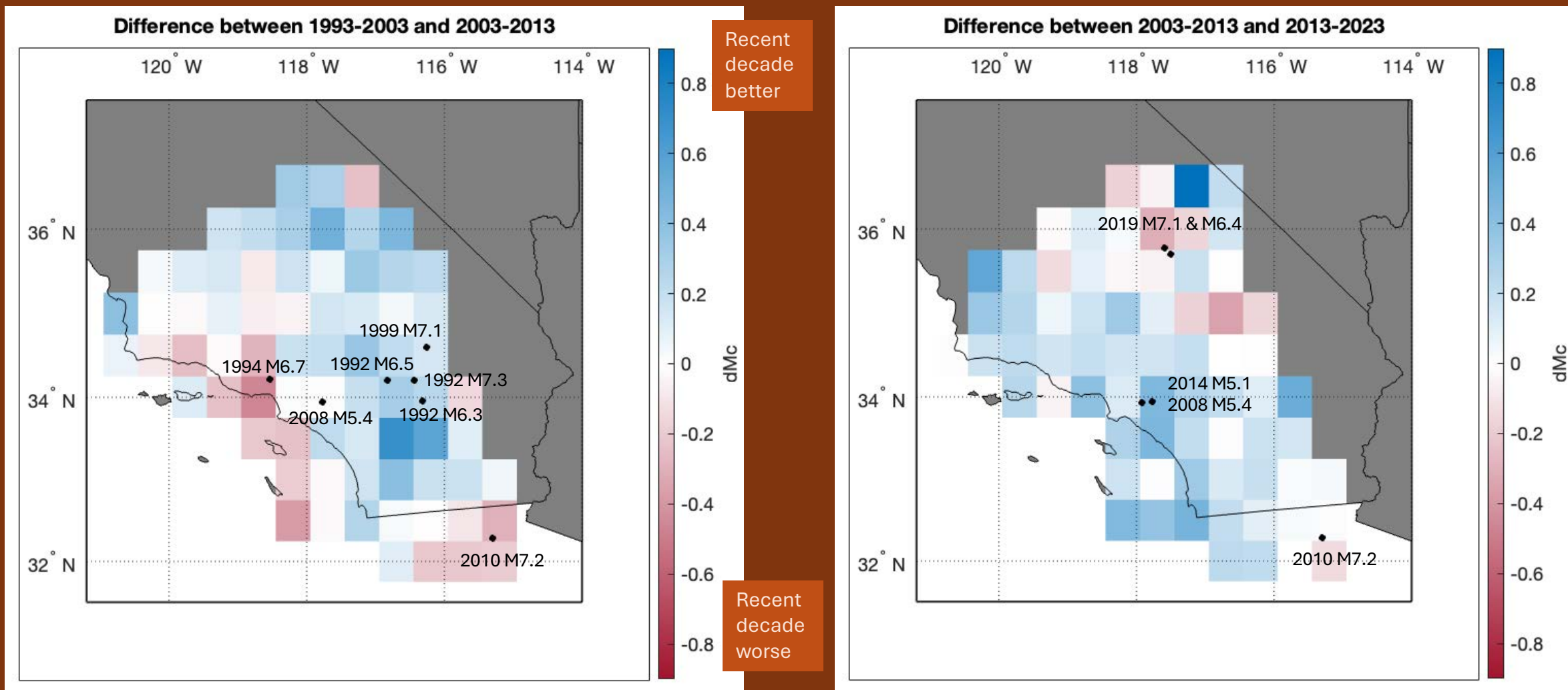
original: 1, 1, 1, 2, 2, 3

sample A: 1, 2, 2, 2, 3, 3

sample B: 1, 1, 1, 1, 2, 2

sample C: 1, 1, 2, 2, 2, 3



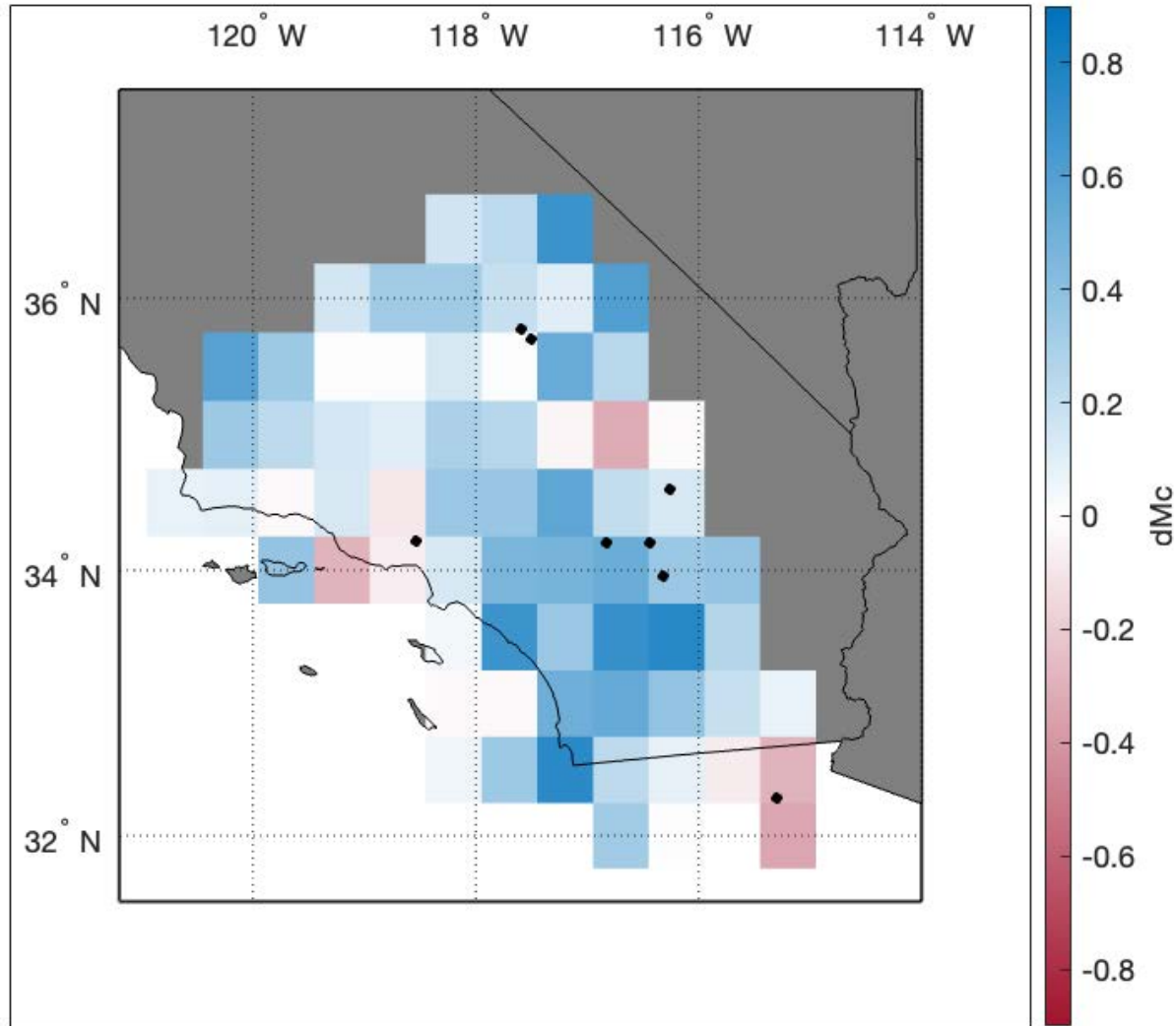


Better: 1994 Northridge

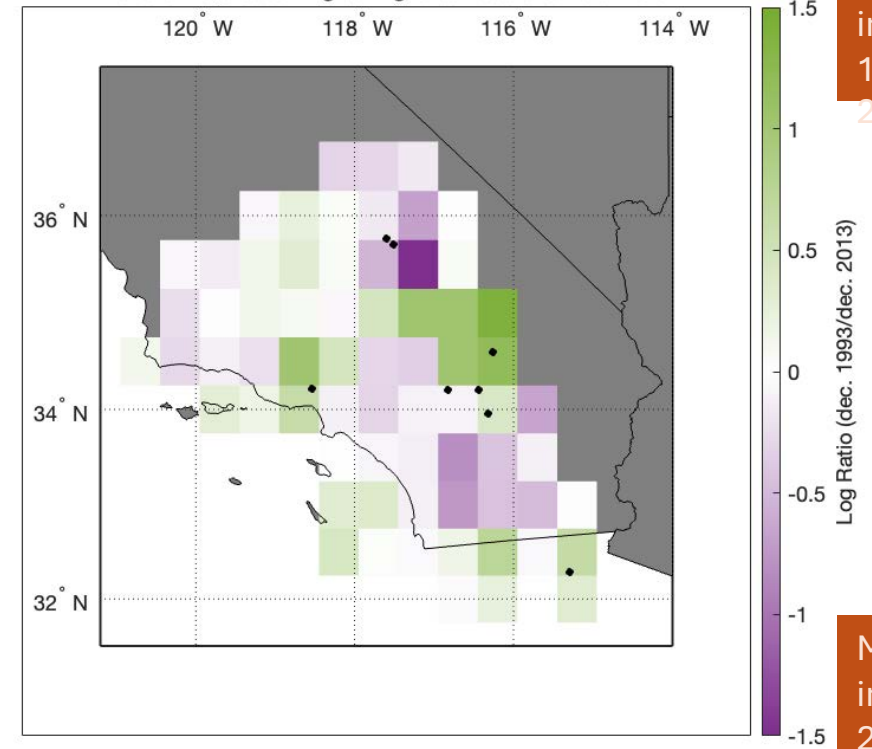
Worse: 2010 El Mayor, 2019 Ridgecrest*, 1992 Landers-Big Bear & Joshua Tree, 1999 Hector Mine

*incomplete

Difference between 1993-2003 and 2013-2022



Ratio of EQs occurring during 1993-2003 and 2013-2023



More EQs
in
1993-
2003

More EQs
in
2013-
2023

-> small improvement (~0.2-0.7 m.u.) in most areas over past 30 years

-> regions with worse Mc typically have fewer earthquakes