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

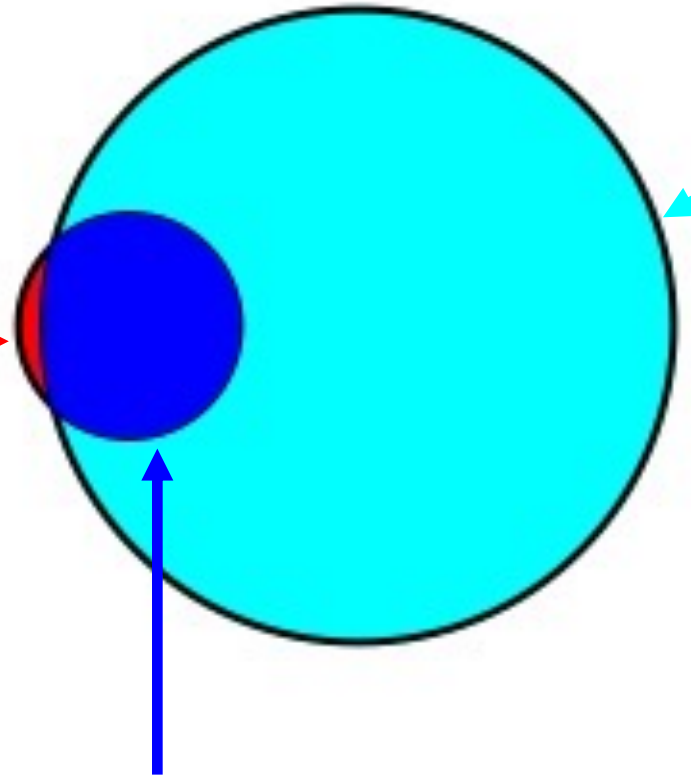
https://www.usgs.gov/programs/earthquake-hazards/national-earthquake-information-center-neic

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?)*



NEW events (only in enhanced catalog)

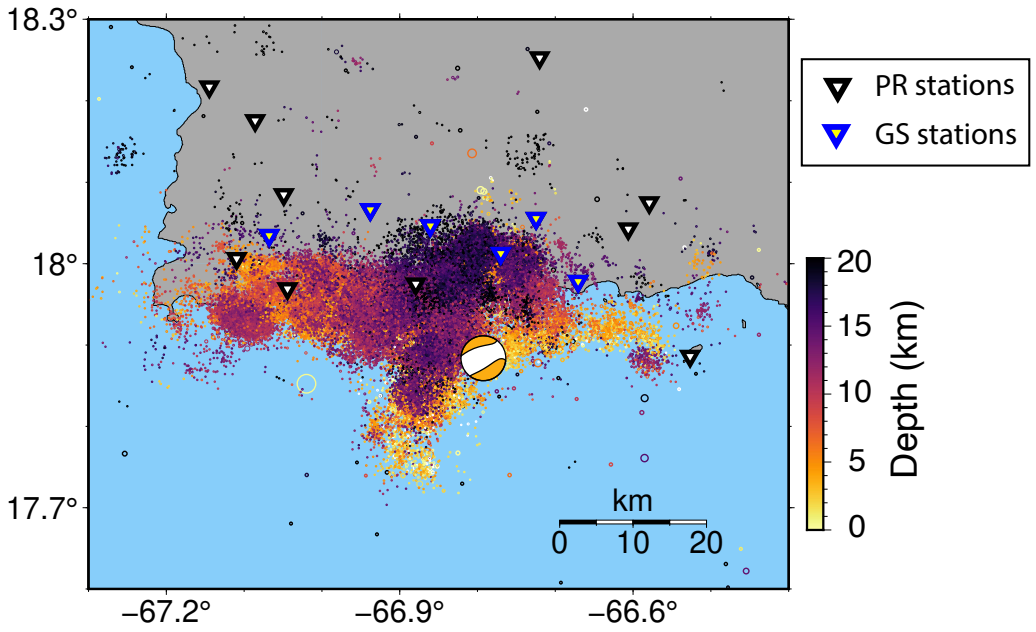
- *Newly detected small local earthquakes?* 😀
- *False detections from noise?* ☹️
- *Other signals (quarry blasts, sonic booms, explosions)?* 🧐
- *Regional or teleseismic earthquakes?* 🧐

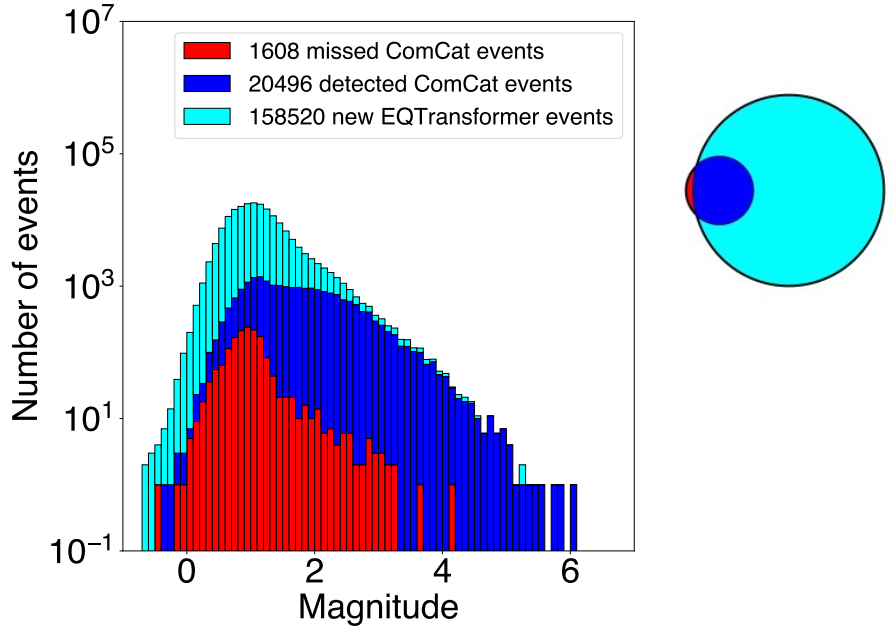
MATCH events (common to both catalogs)

- *Origin times match within X~5 seconds*
- *Hypocenters match within Y~25 km*

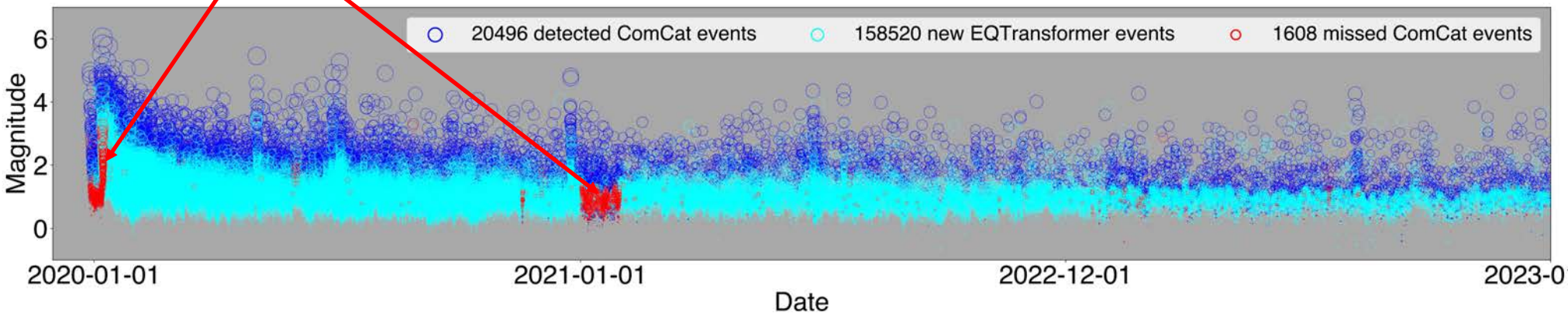
Event comparison example: SW Puerto Rico sequence

Yoon et al. (2023), BSSA



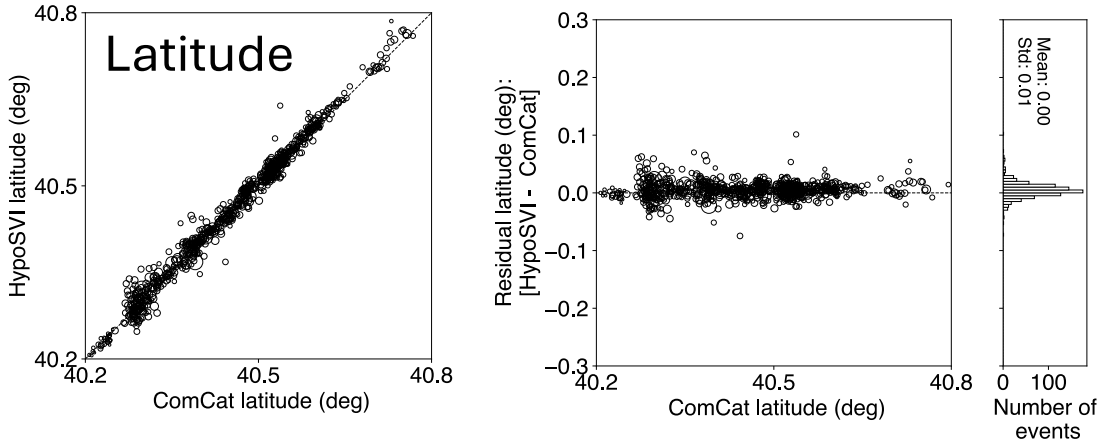


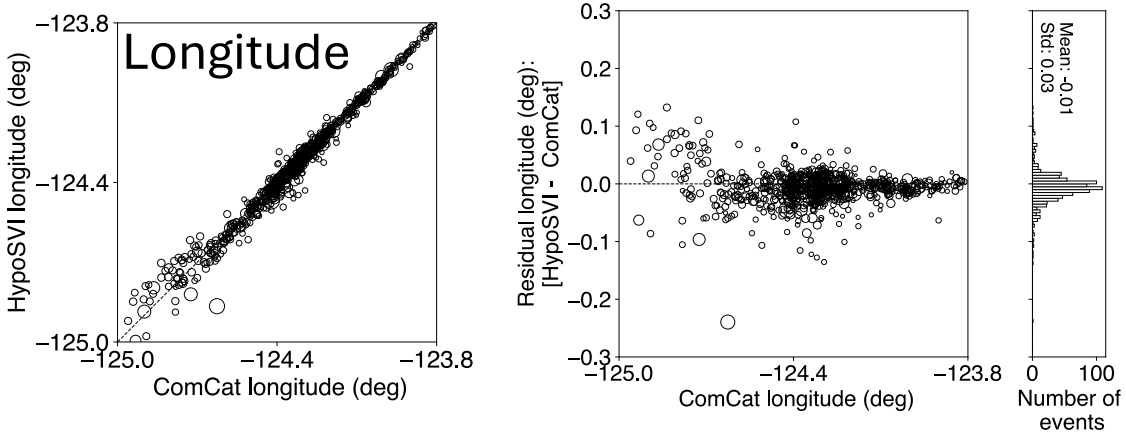
Times with many missed ComCat earthquakes: PRSN* analyst review complete

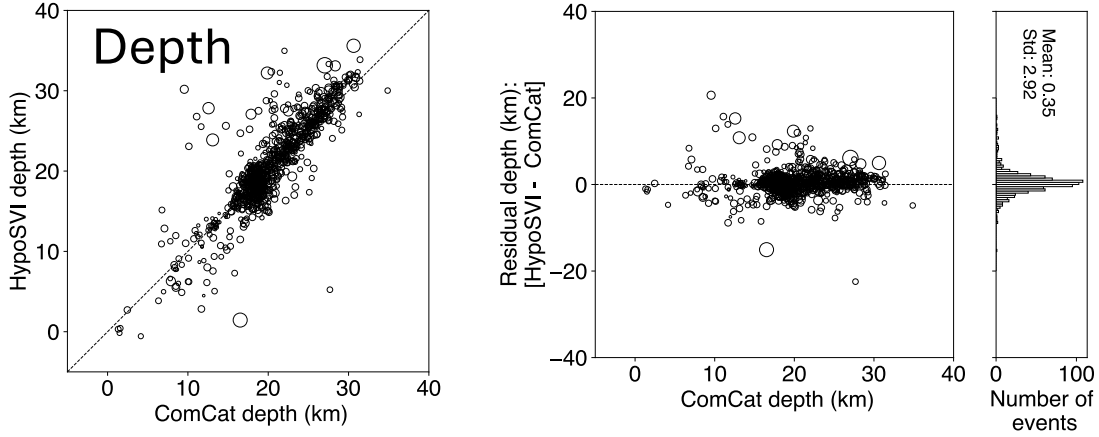


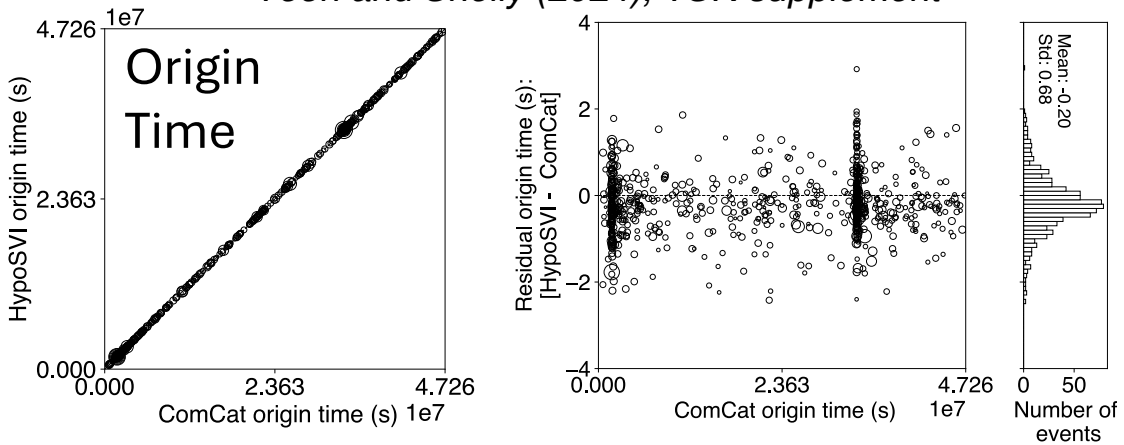
*PRSN: Puerto Rico Seismic Network

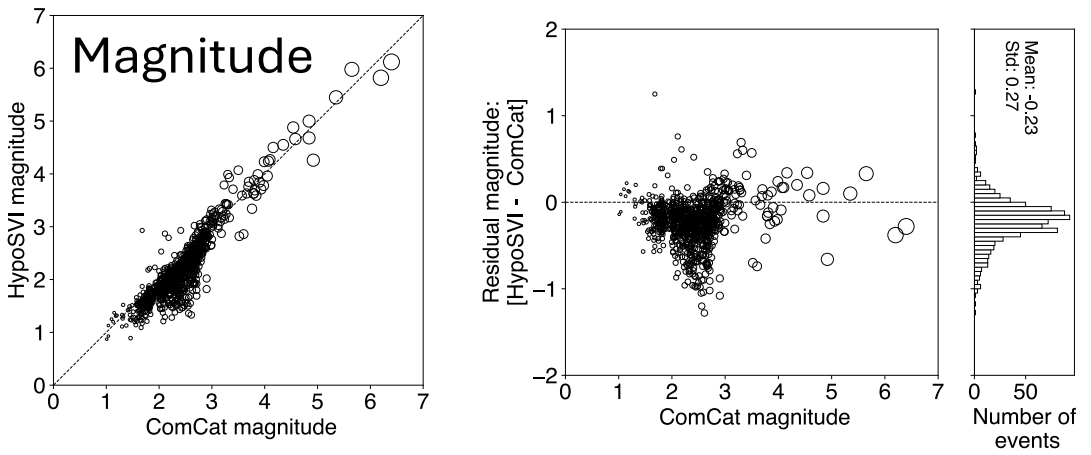
Yoon and Shelly (2024), TSR supplement











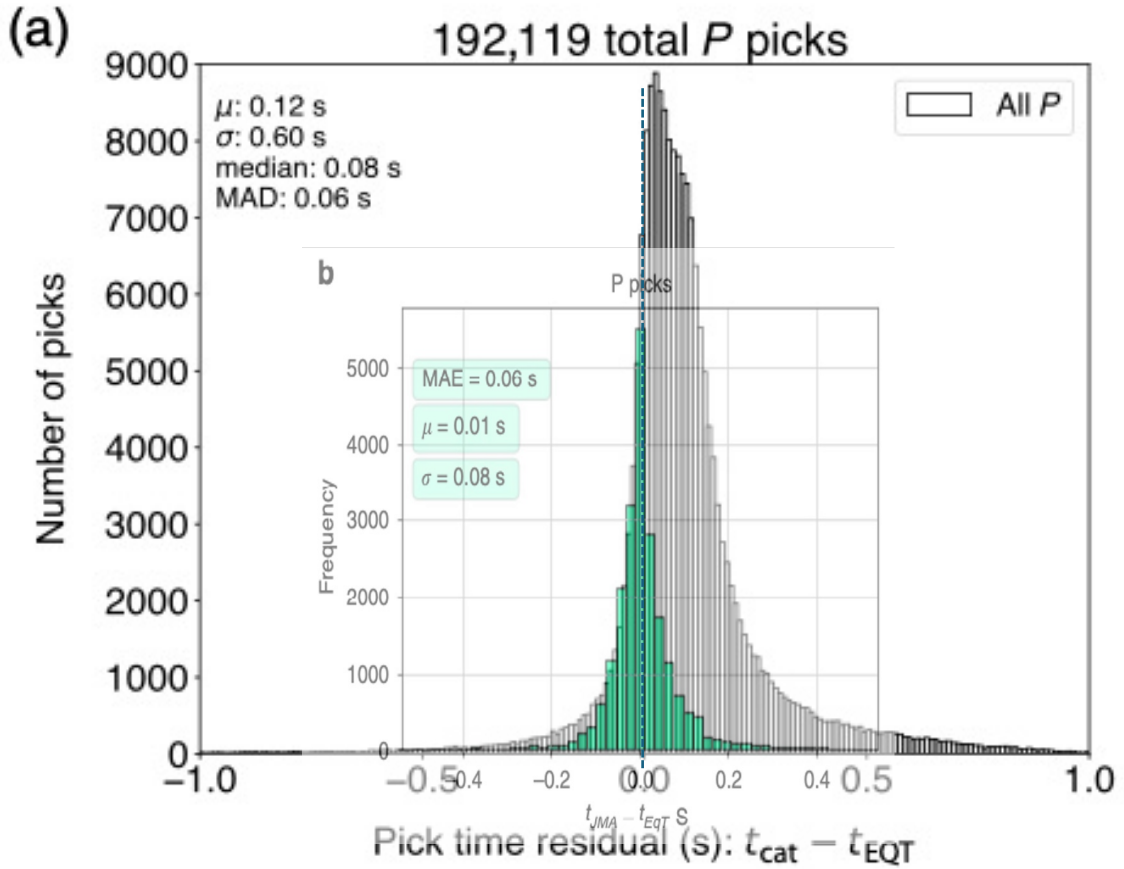
MATCH events: **location, magnitude** comparison

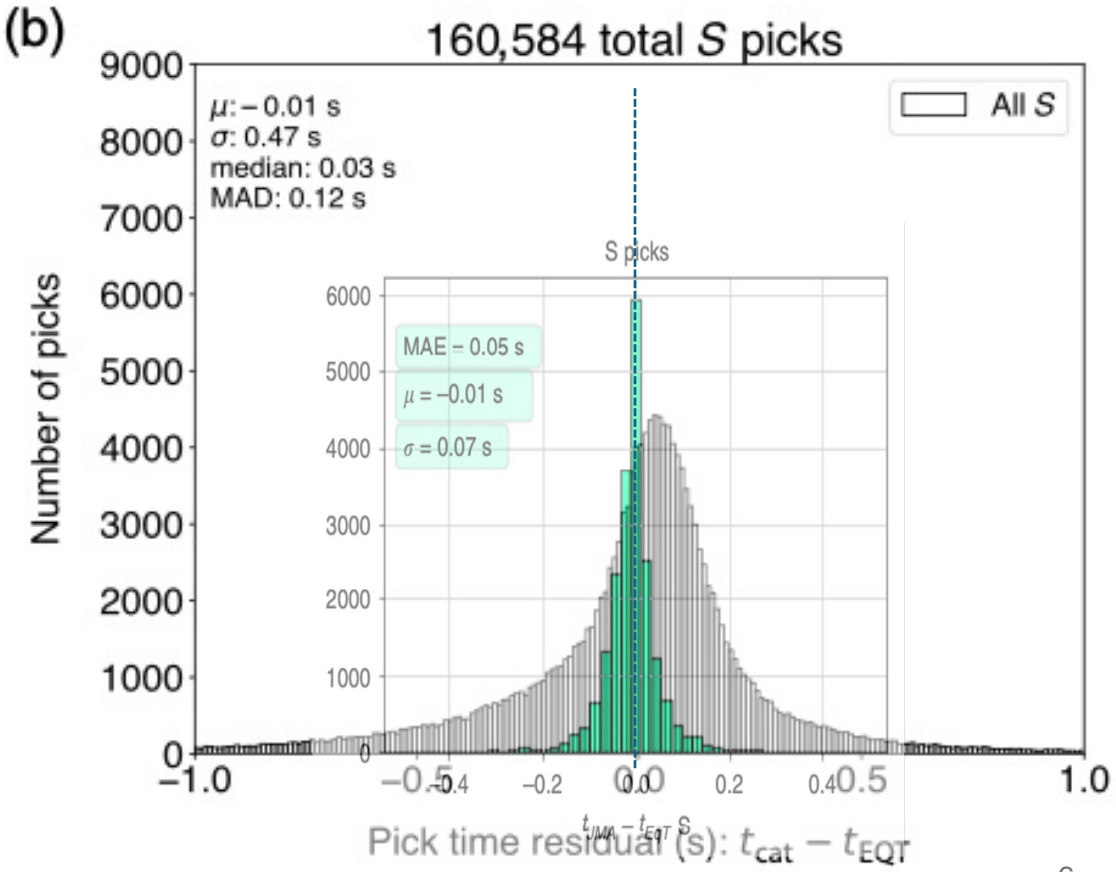
Horizontal axis: ComCat (reference catalog)

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

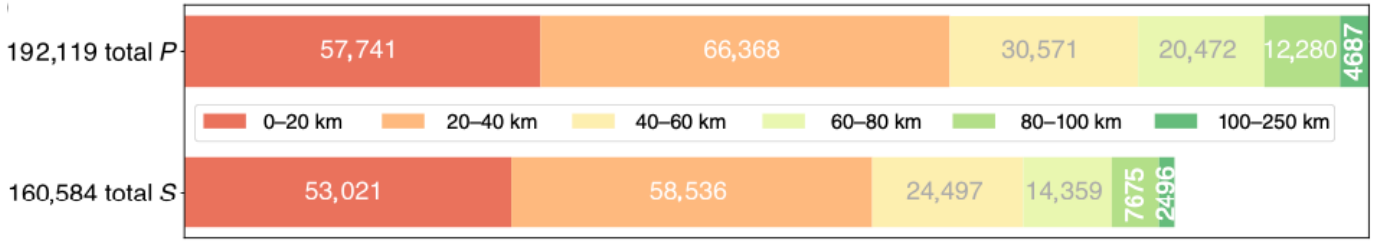




Yoon et al. (2023), BSSA

"Ground truth" t_{cat}: ComCat analyst picks

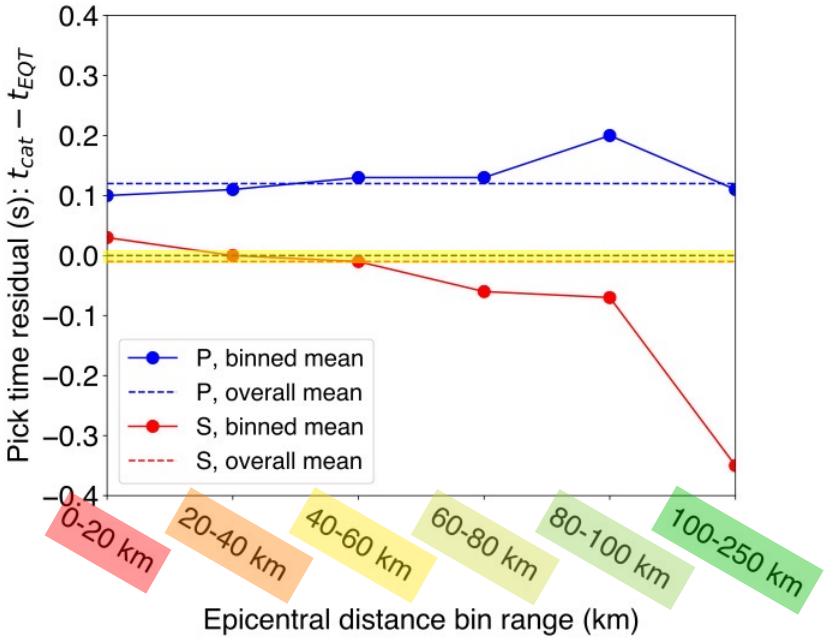
EQT **pick quality** (number of picks, pick accuracy) **degrades** with **increasing** event-station **distance**



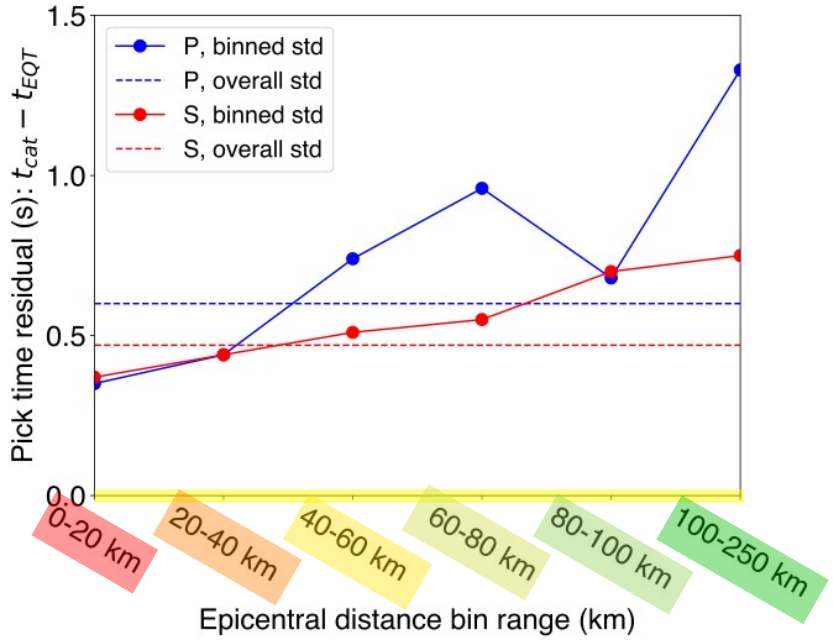
Yoon et al. (2023), BSSA

In practice: use only picks out to a maximum distance (~100 km?) for location

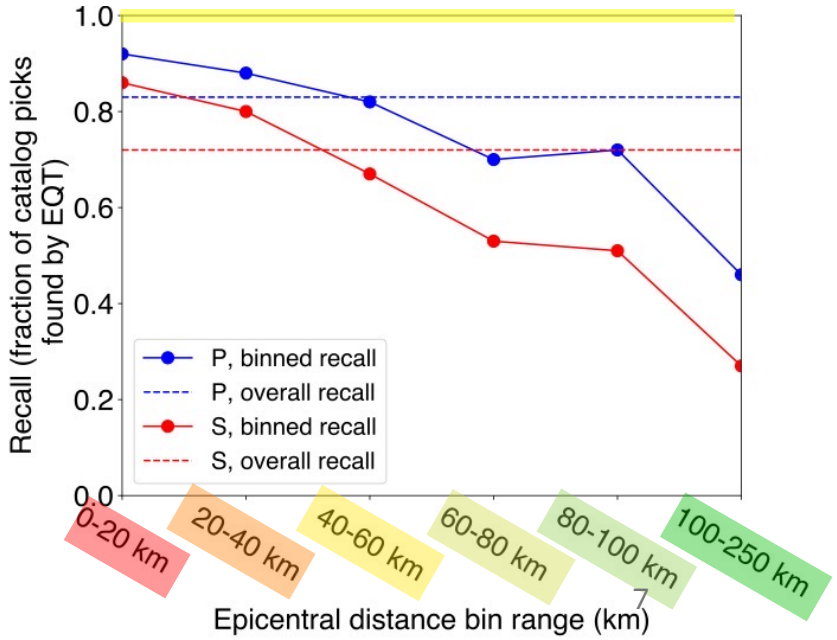
Mean $\mu[t_{cat} - t_{EQT}]$
of pick time residual



Stdev $\sigma[t_{cat} - t_{EQT}]$
of pick time residual



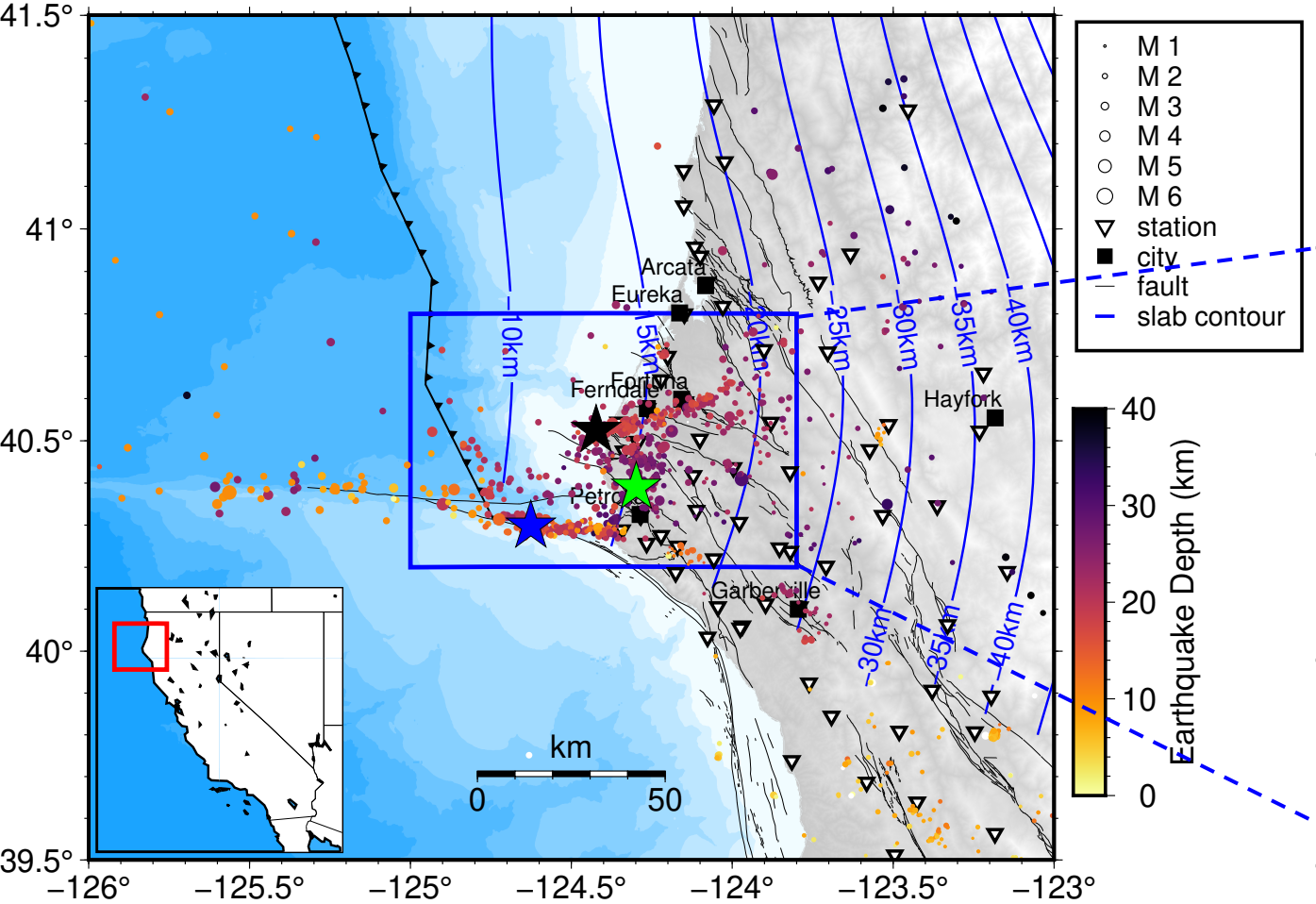
$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} = \frac{N_{\text{cat,EQT}}}{N_{\text{cat,EQT}} + N_{\text{cat,only}}},$$



Reference Catalog Tip: Select boundaries carefully

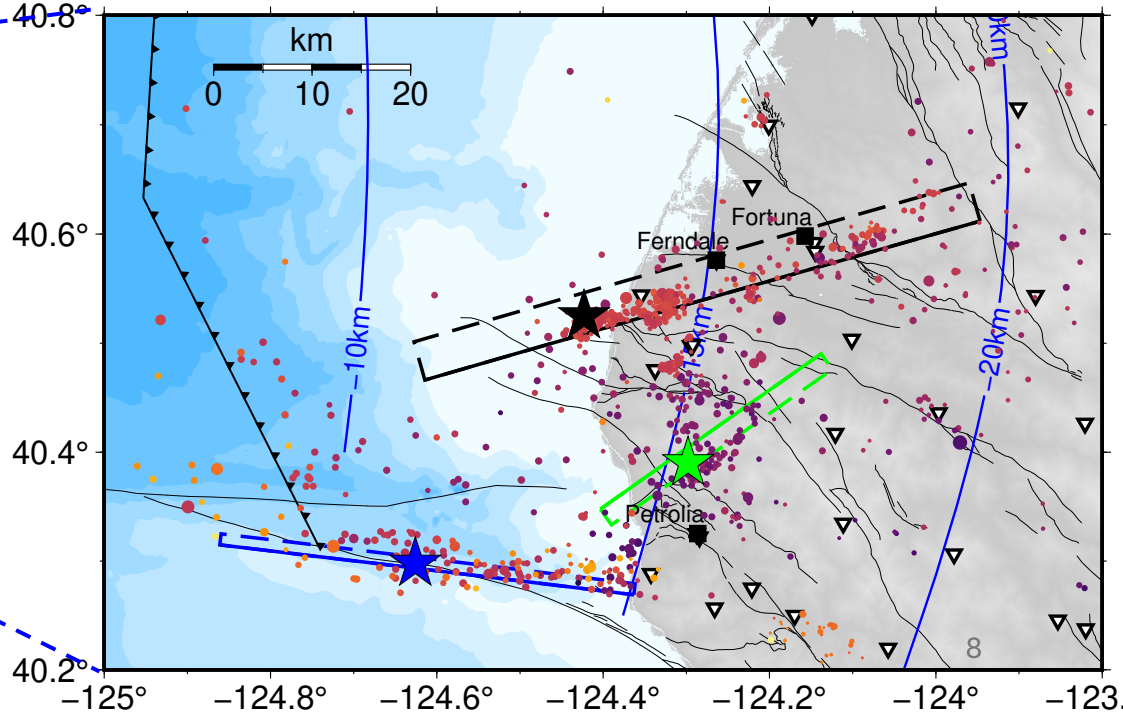
Example: Mendocino Triple Junction, CA

Region big enough: includes most seismicity and all stations used



Yoon and Shelly (2024), TSR

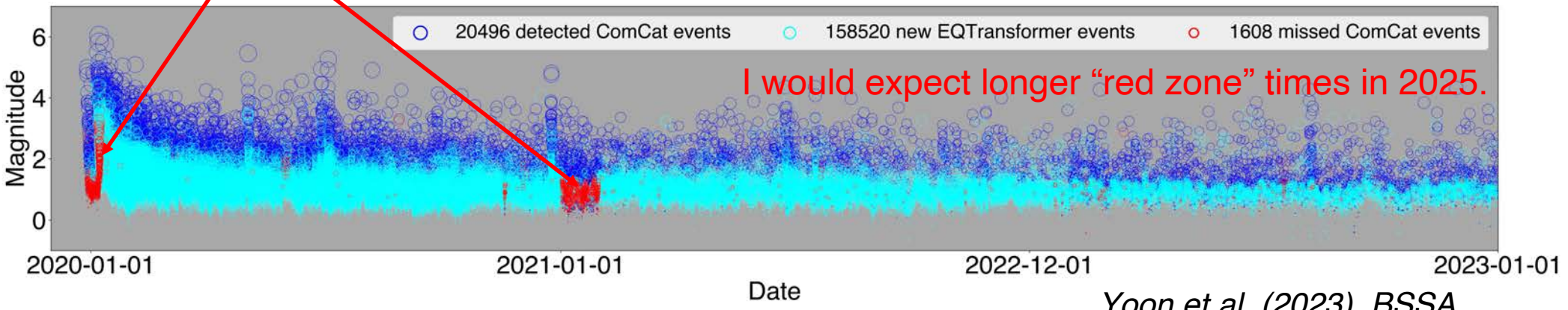
Region too small: missing some regions with seismicity; not all stations



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/

Times with many missed ComCat earthquakes: PRSN* analyst review complete (back in 2023)



Yoon et al. (2023), BSSA

Reference Catalog Tip: It changes over time

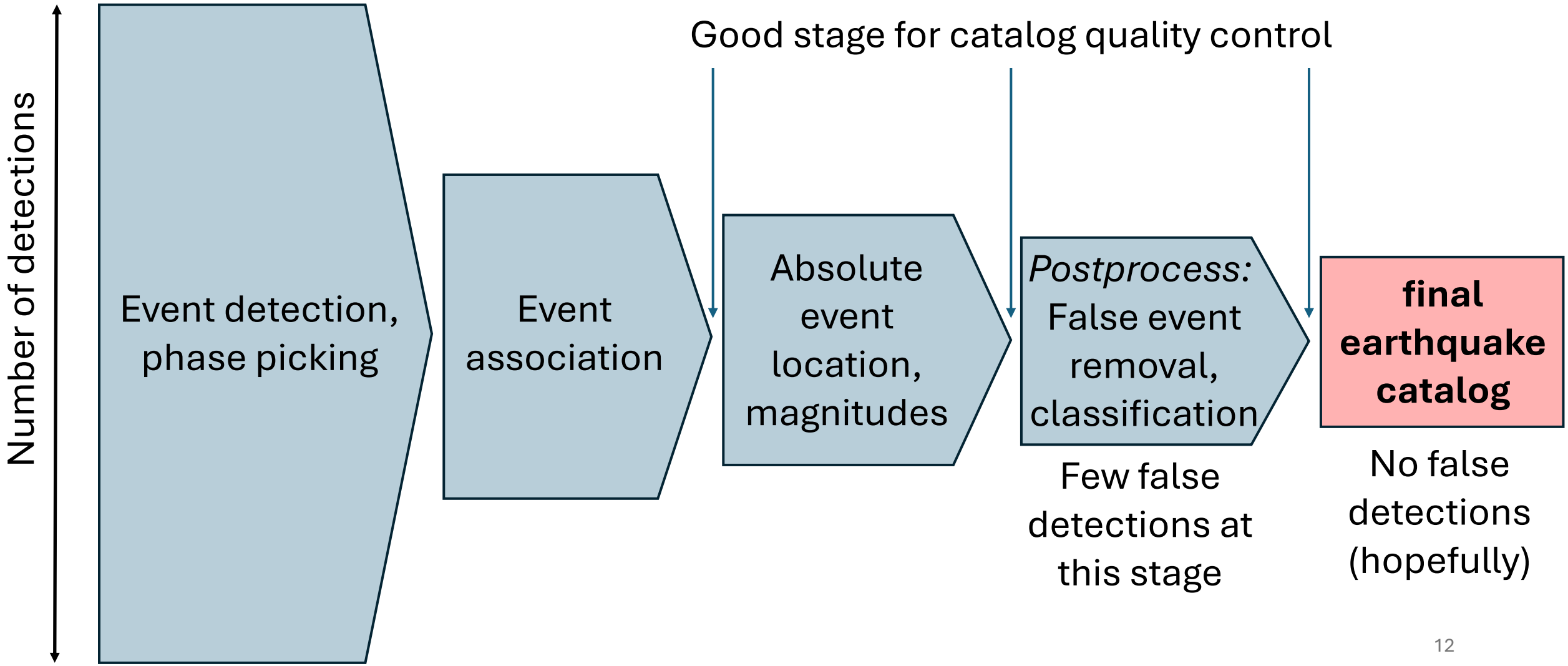
- **Save the reference catalog file itself, with filename including the date you downloaded it.**
  - When reference catalog changes in the future, you still have the original version, so you can easily reproduce your results.

Earthquake Catalog: Quality Control (QC)

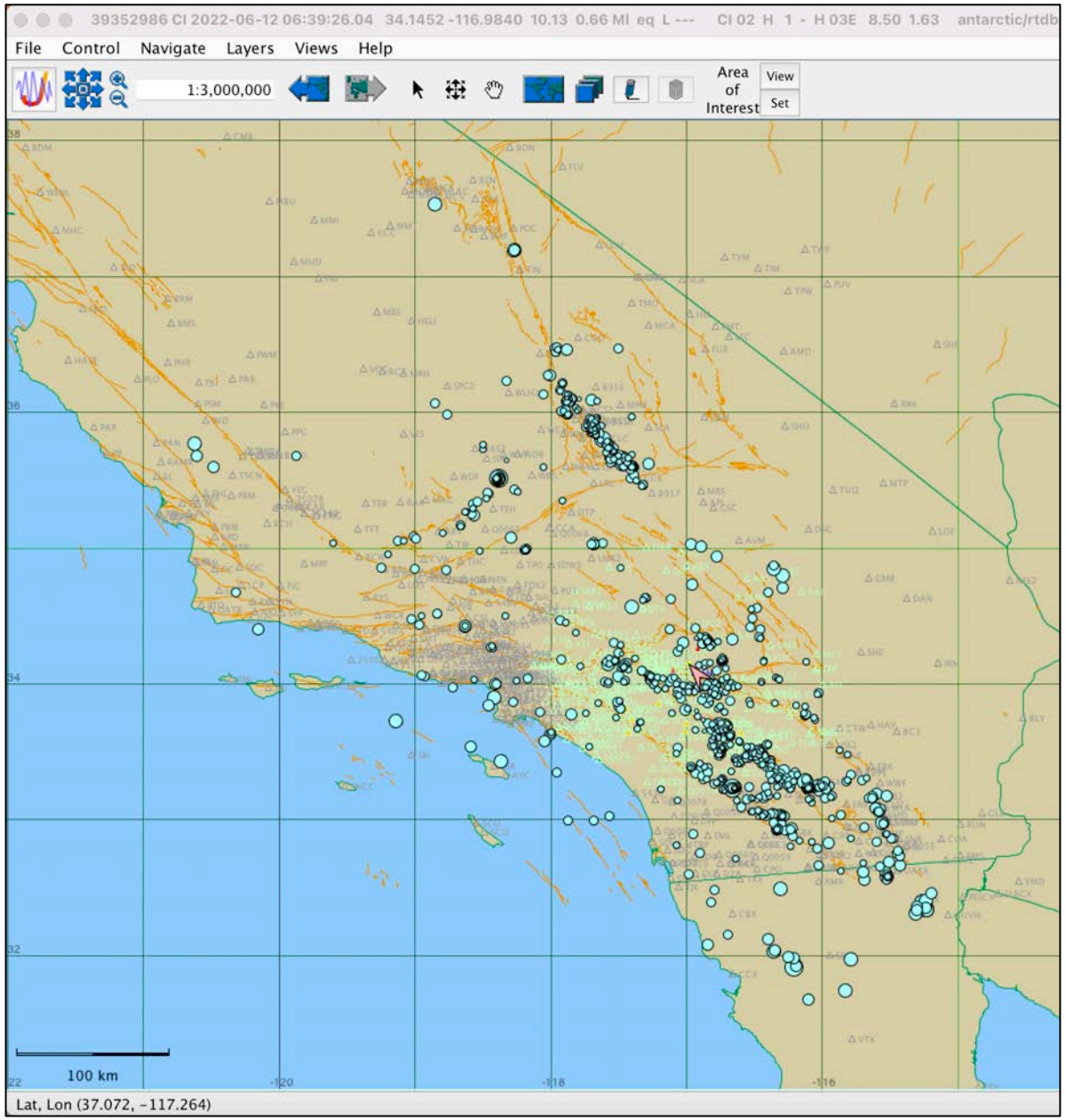
- I created a custom (deep-learning) enhanced earthquake catalog automatically. How good is it?
  - No ground truth available

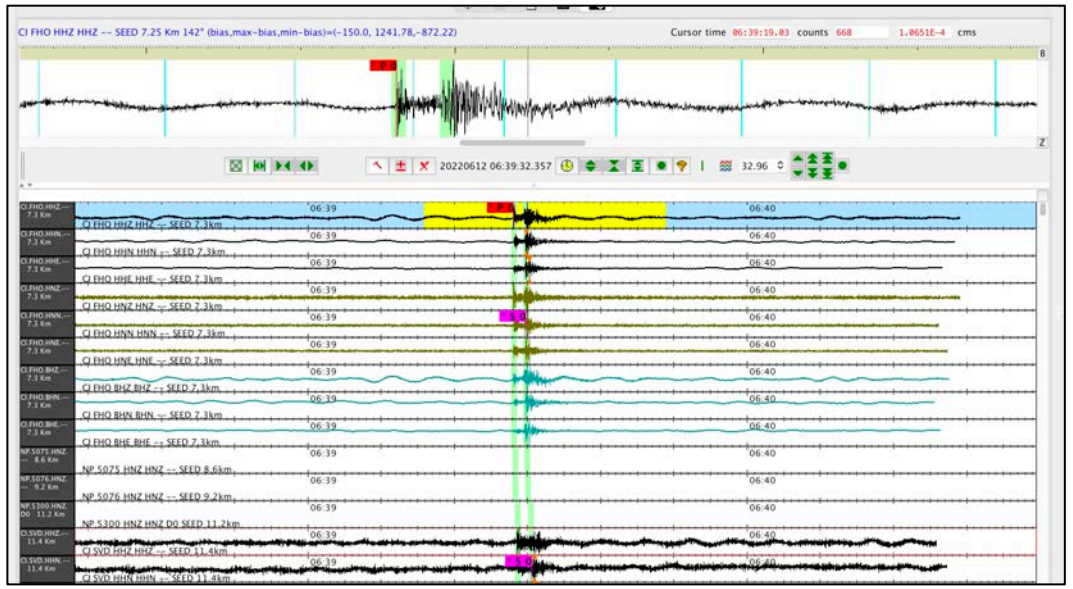
- Before using or interpreting your enhanced catalog, do this first:
  - Compare with a **reference catalog**
  - **Data visualization** (make plots!) to check for quality

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

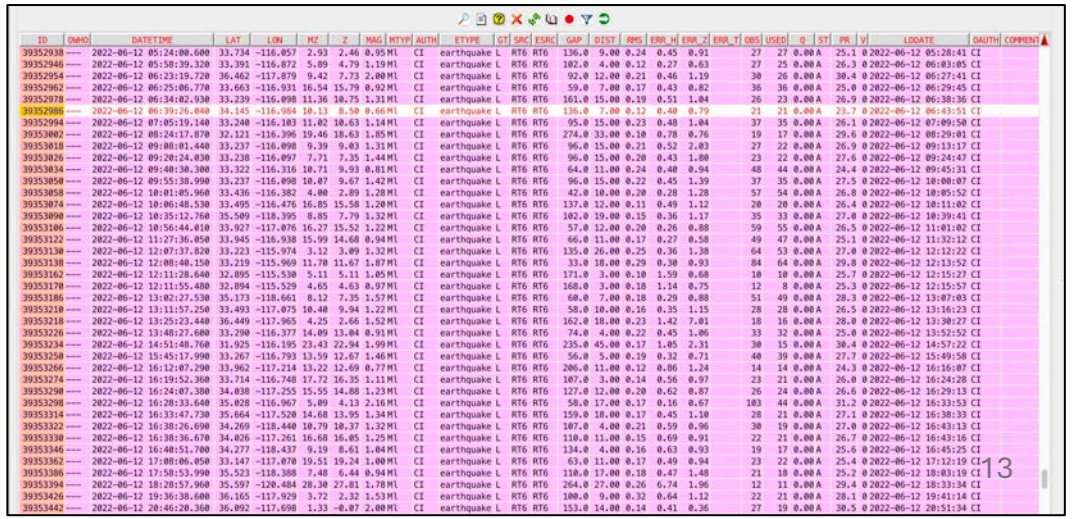


Earthquake Catalog QC: Think like an analyst





AQMS Jiggle manual review and picking interface at SCSN



Earthquake Catalog QC Tips: Check Parameters

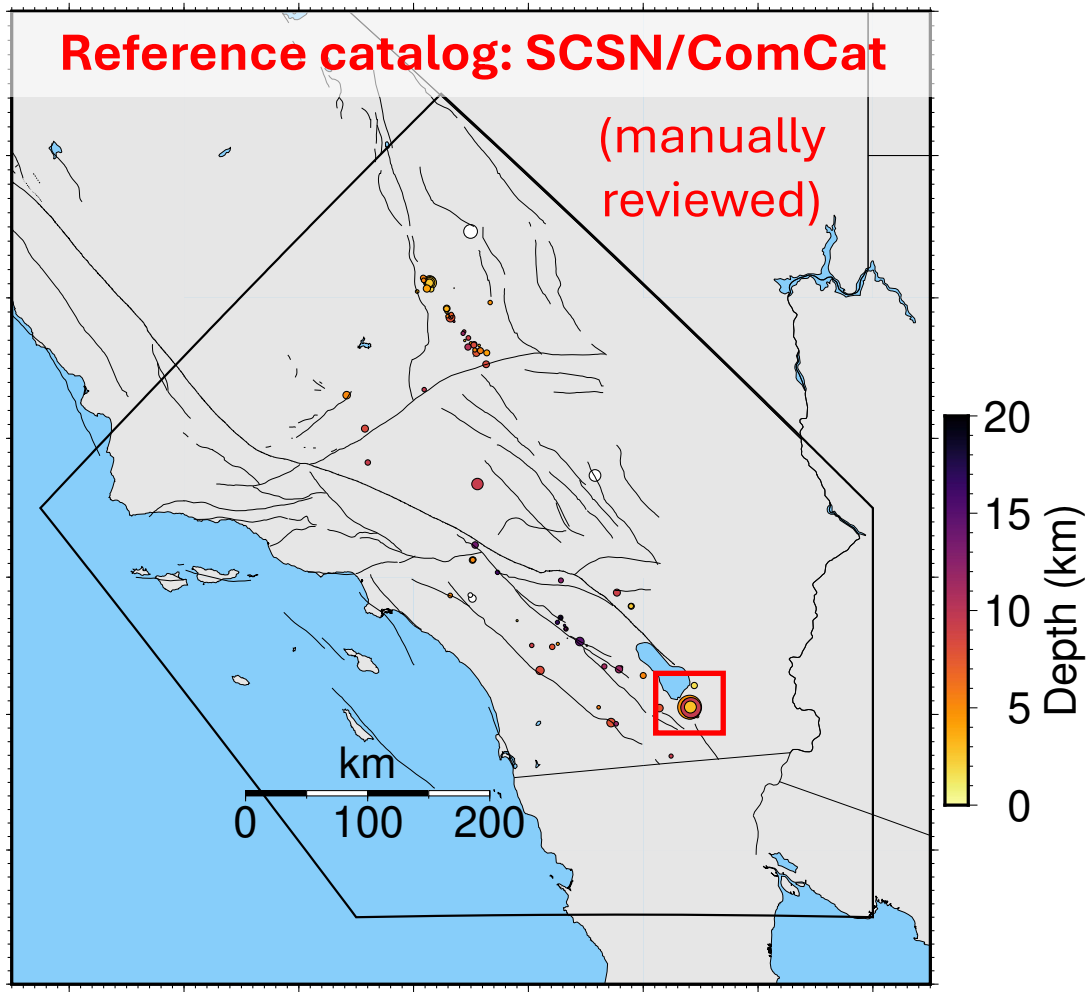
- Compare enhanced catalog with reference catalog on plots
  - Plot reference & enhanced catalogs side-by-side
  - Separate plots for MATCH, NEW, MISSED category events

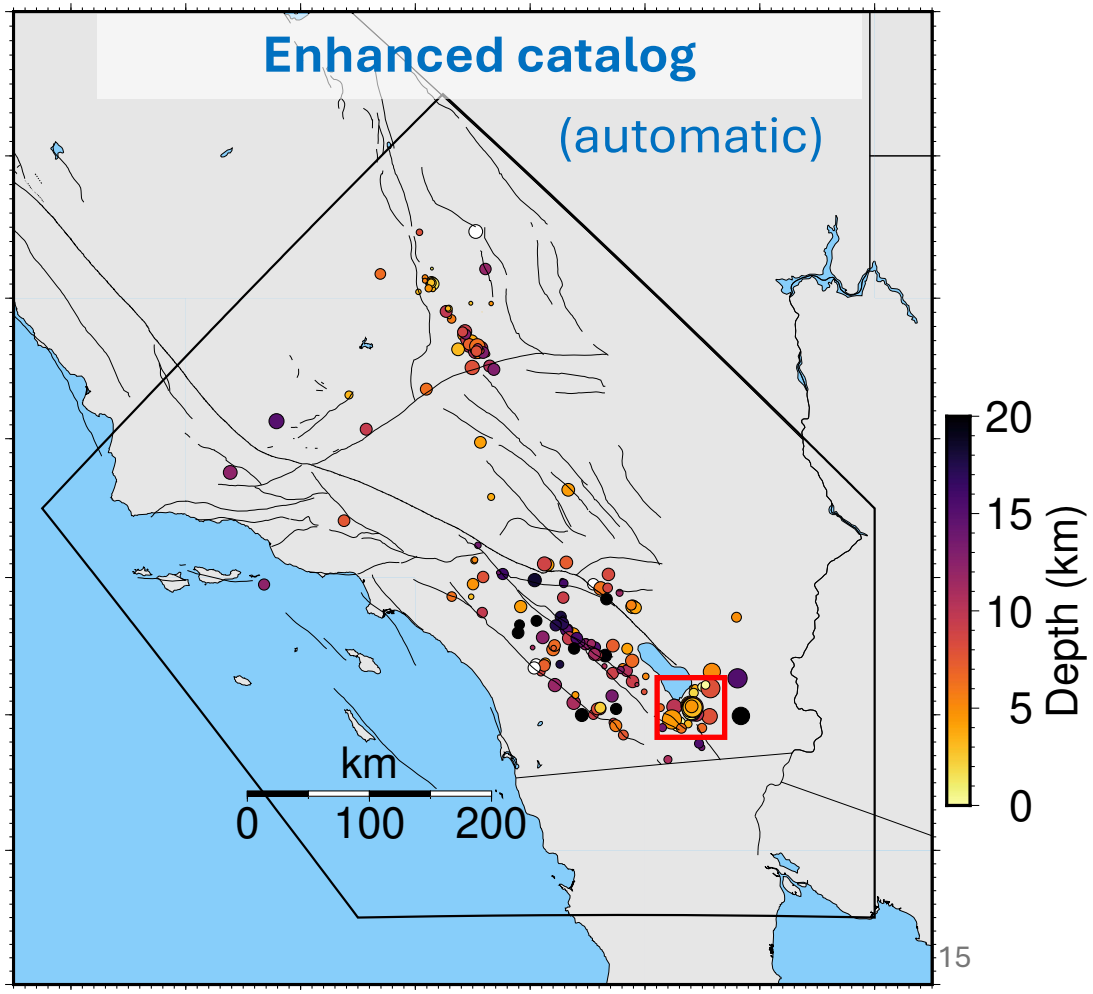
- Visualize enhanced catalog → are results reasonable for earthquakes?
  - Locations (epicenters) in map view
    - Are events near known seismogenic areas, past seismicity, faults? Known quarries?
  - Depth cross-sections
    - Are events at expected seismogenic depths (0-30 km, unless in subduction 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/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

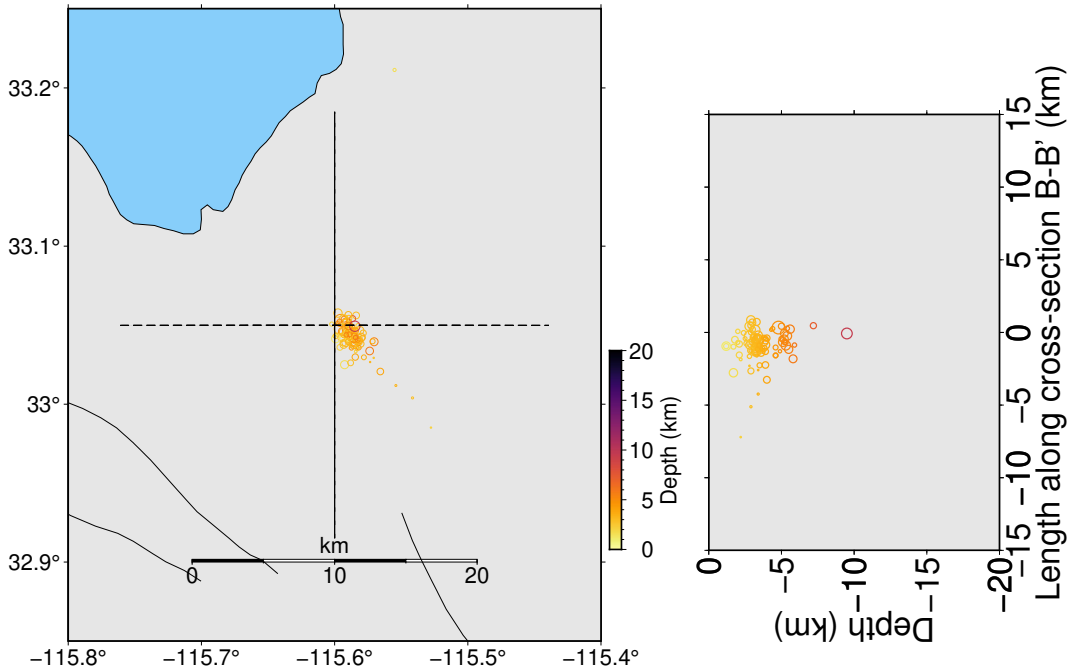


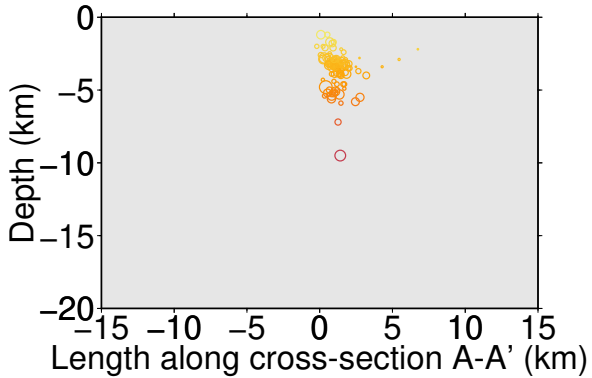


Catalog QC, zoomed locations in map view & depth cross-section: SCSN

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

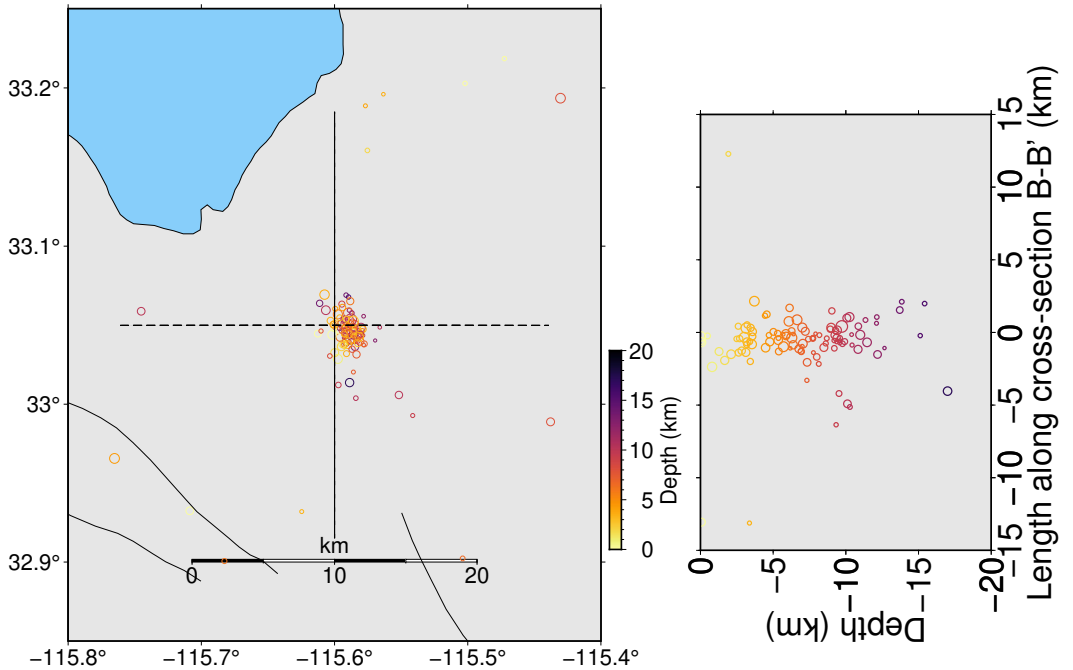
Enhanced catalog: Similar epicenters. Larger range of depths, but still reasonable

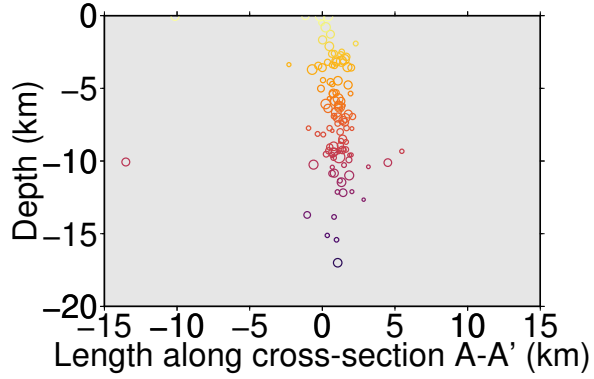




Reference catalog: SCSN/ComCat

(manually reviewed)





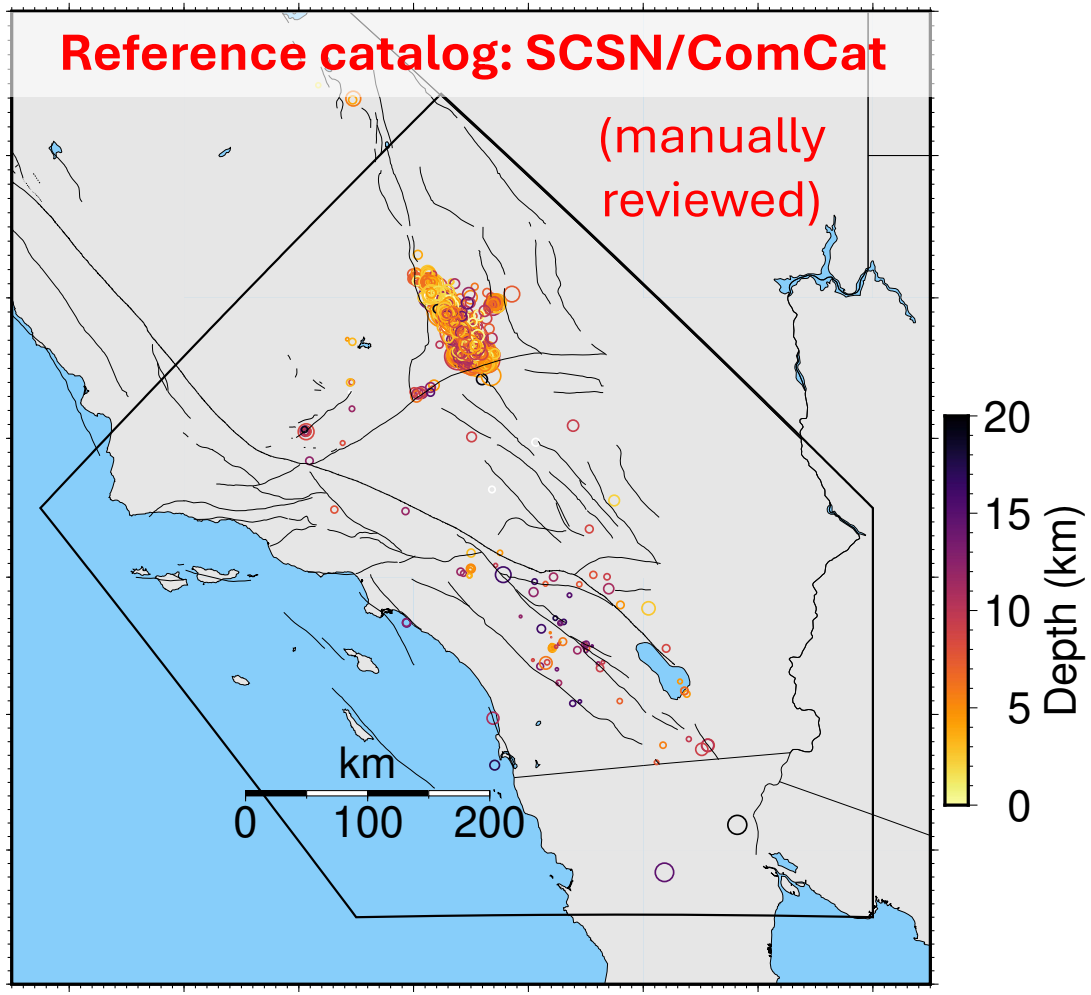
Enhanced catalog

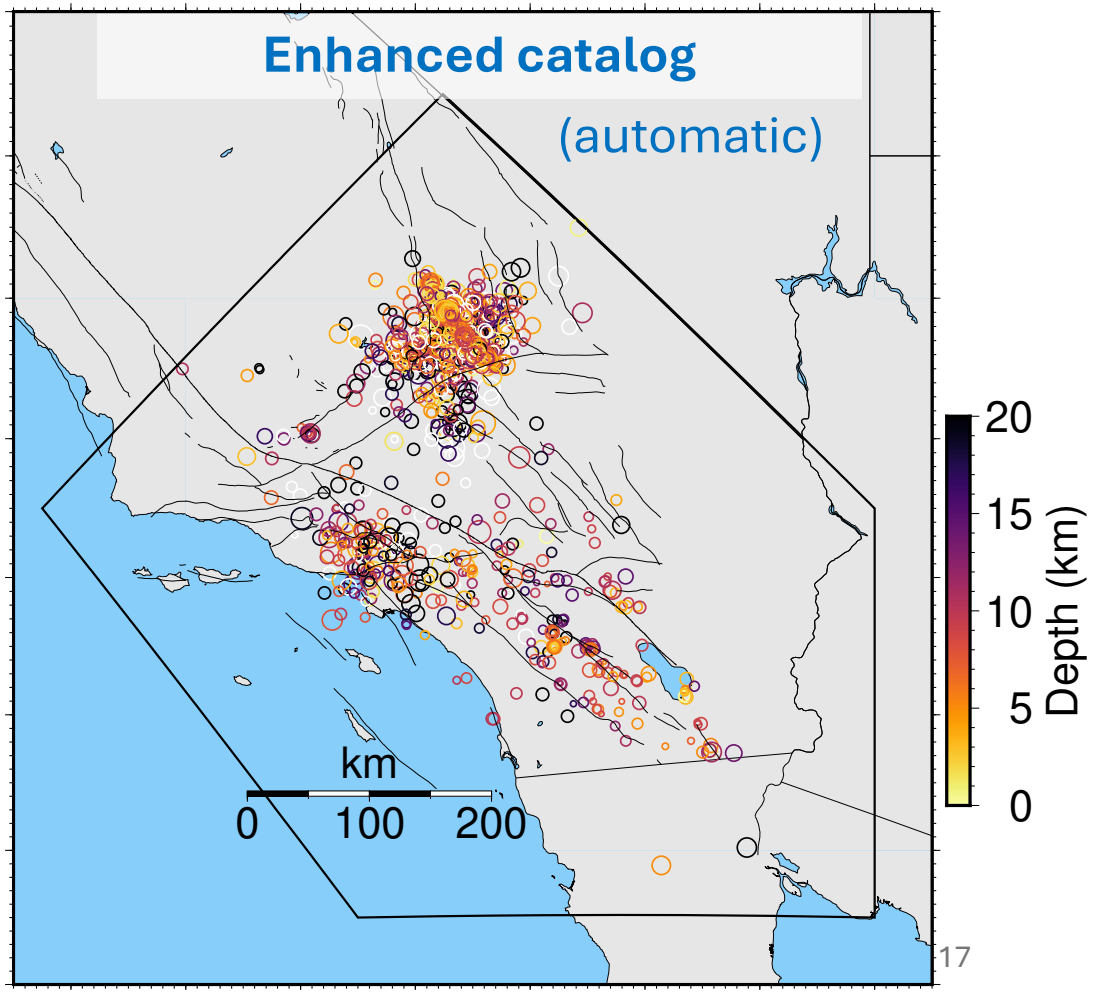
(automatic)

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





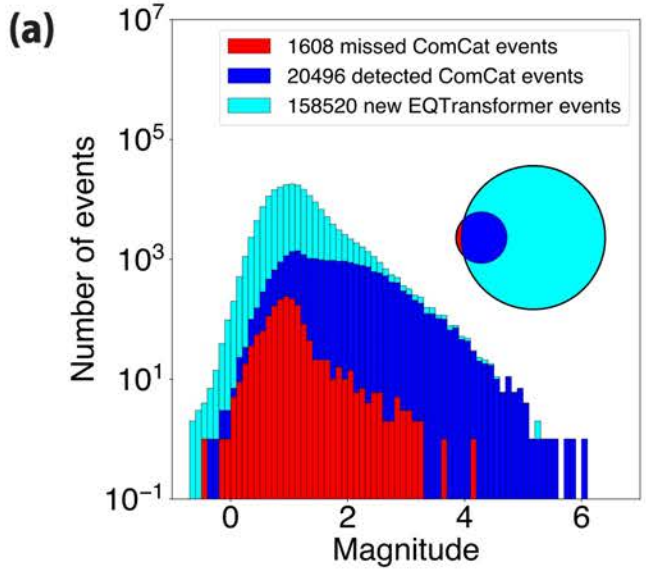
Earthquake Catalog QC Tips: Check Parameters

- 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/time after mainshock (Omori decay)?

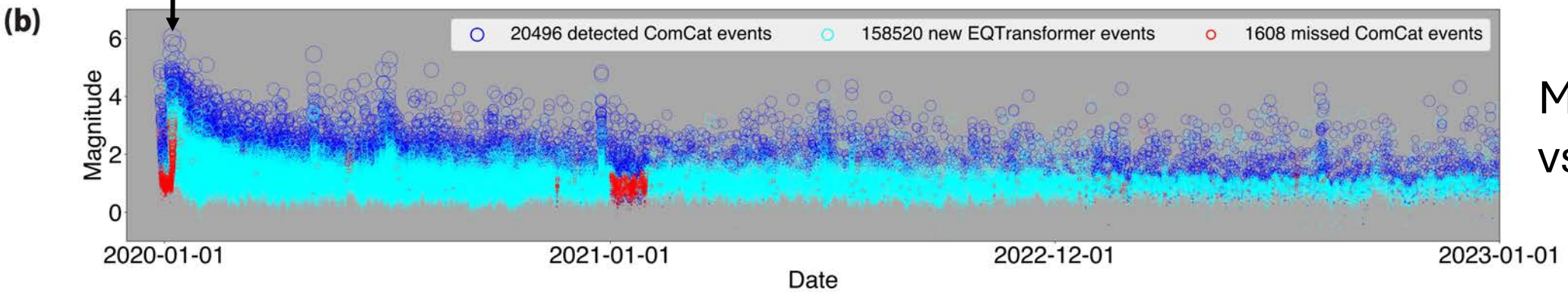
Catalog QC: Puerto Rico

MATCH,
MISSED,
NEW

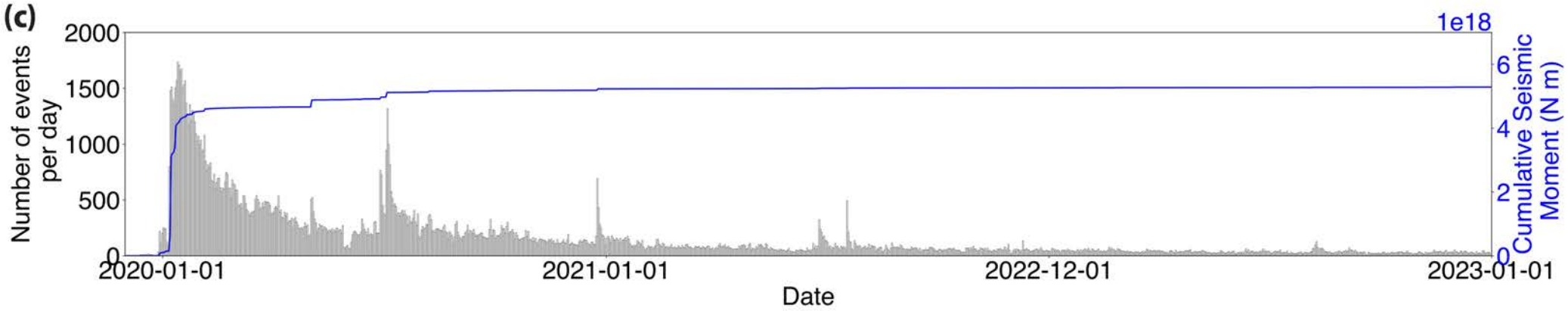


Magnitude-frequency distribution (MFD)

M_w 6.4 mainshock



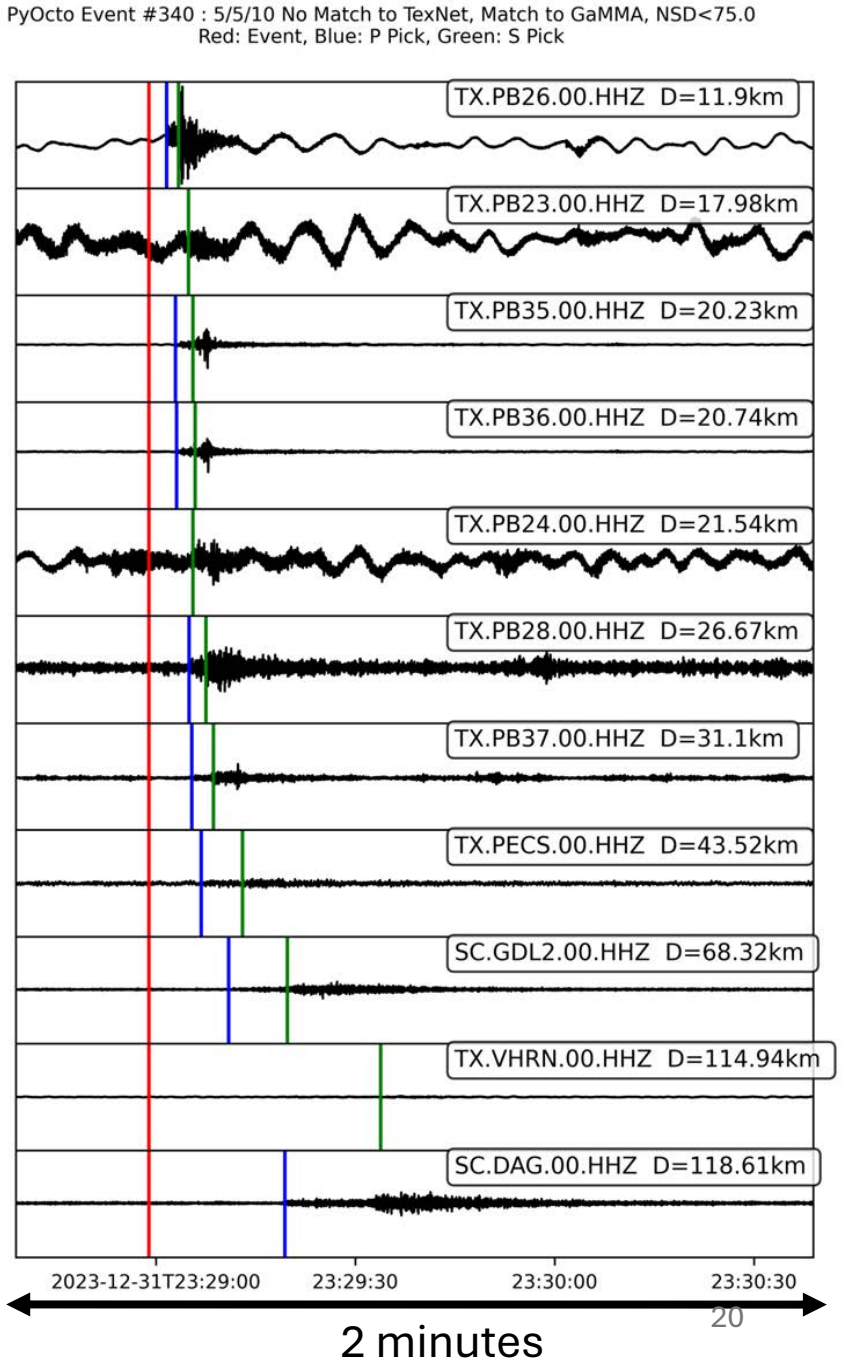
Magnitude vs. time



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)?



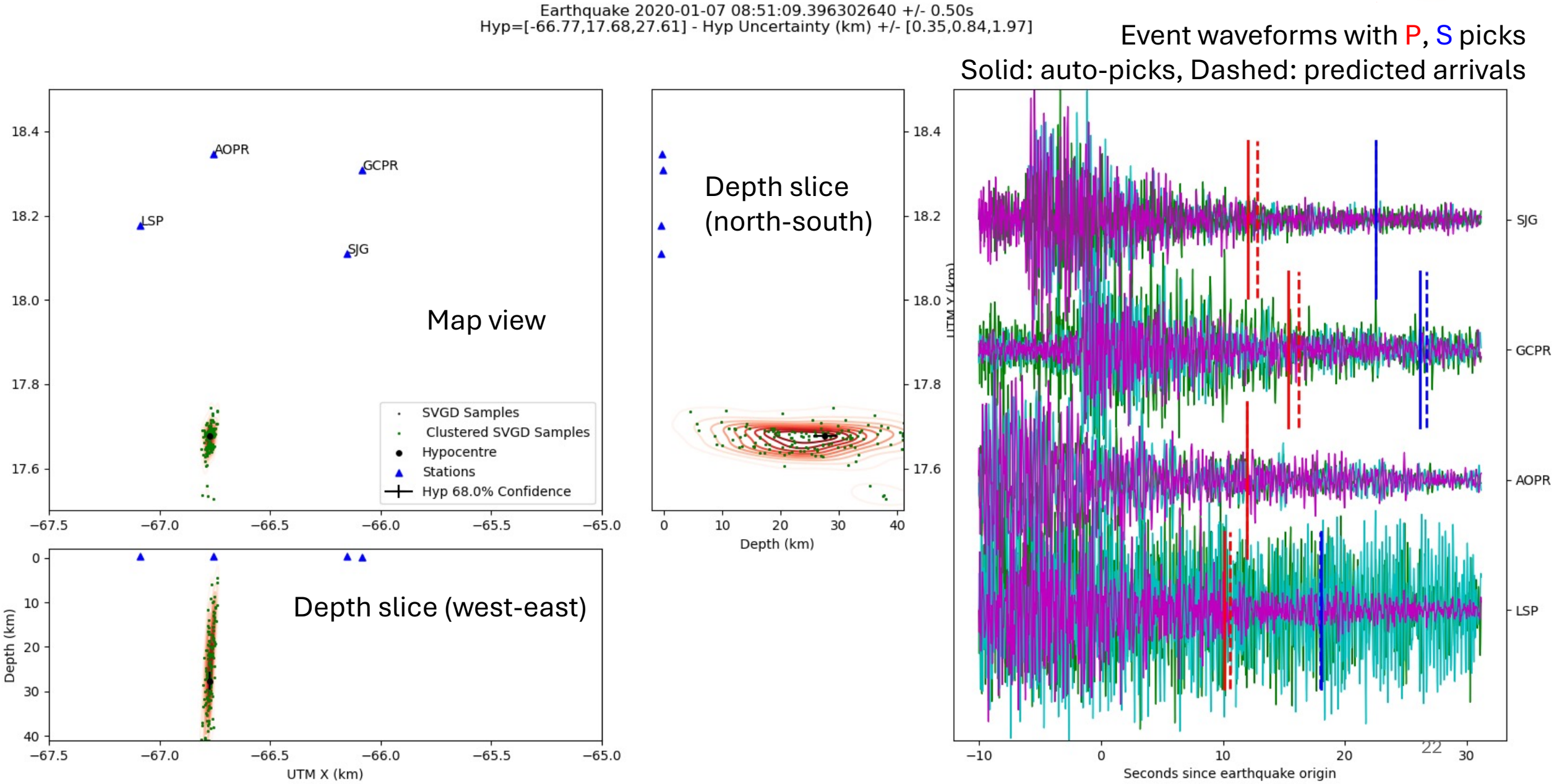
Check "NEW" events in enhanced catalog, not in reference catalog

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

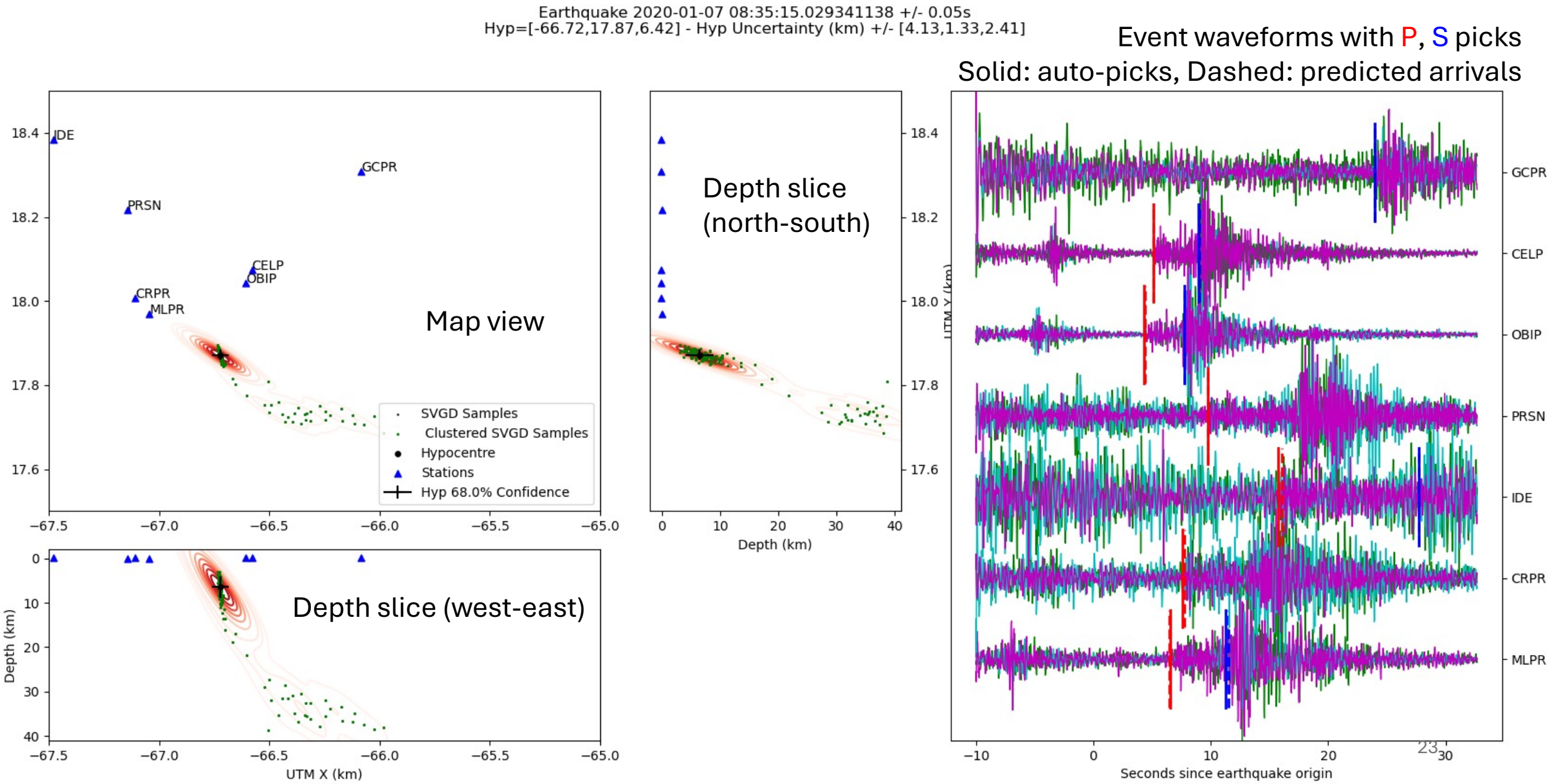
Catalog QC: False detection in coda of larger earthquake (also, too deep?)

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



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



Check "NEW" events in enhanced catalog, not in reference catalog

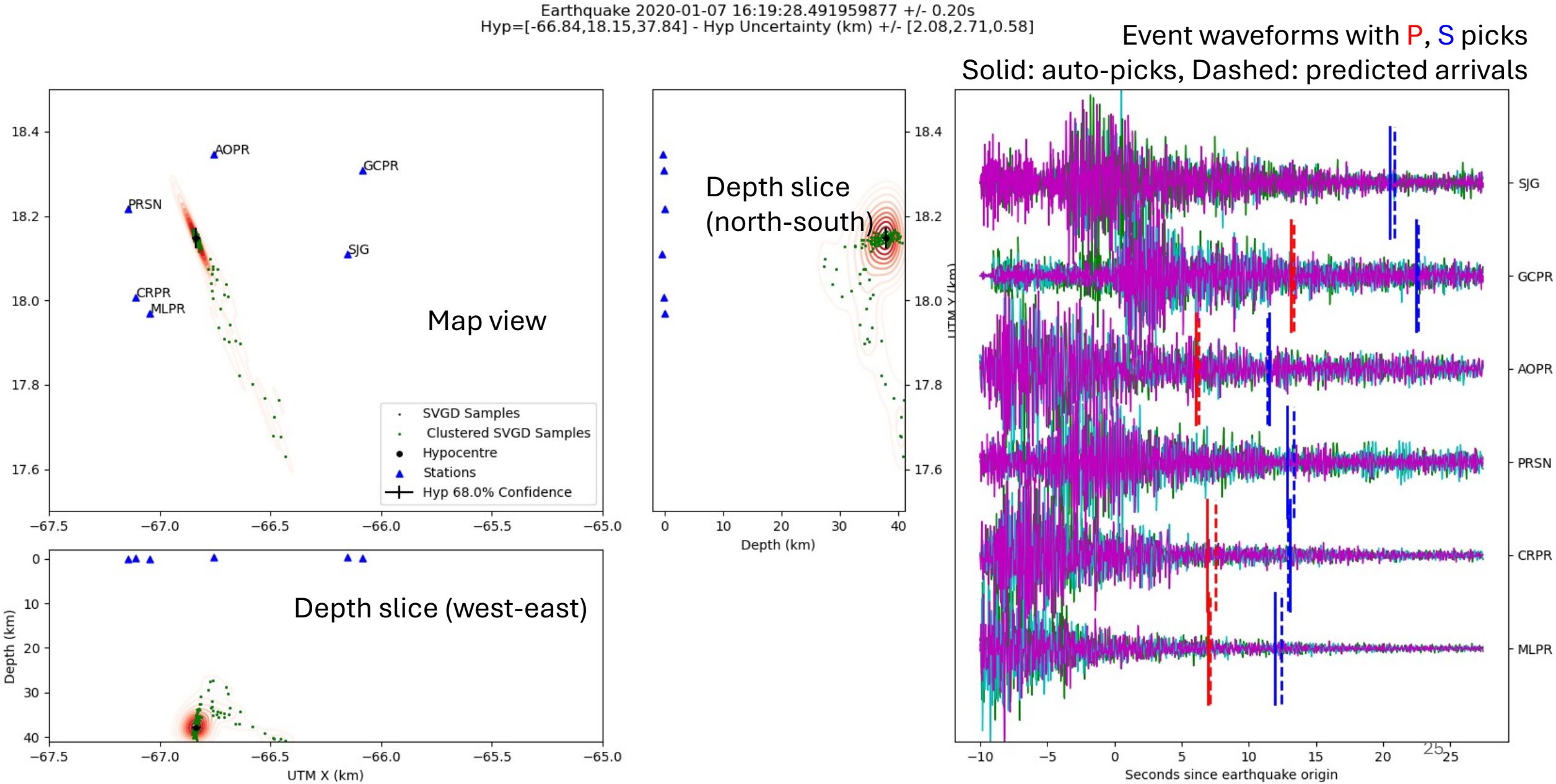
Sorted by **depth**
(Southwest Puerto Rico)

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?)

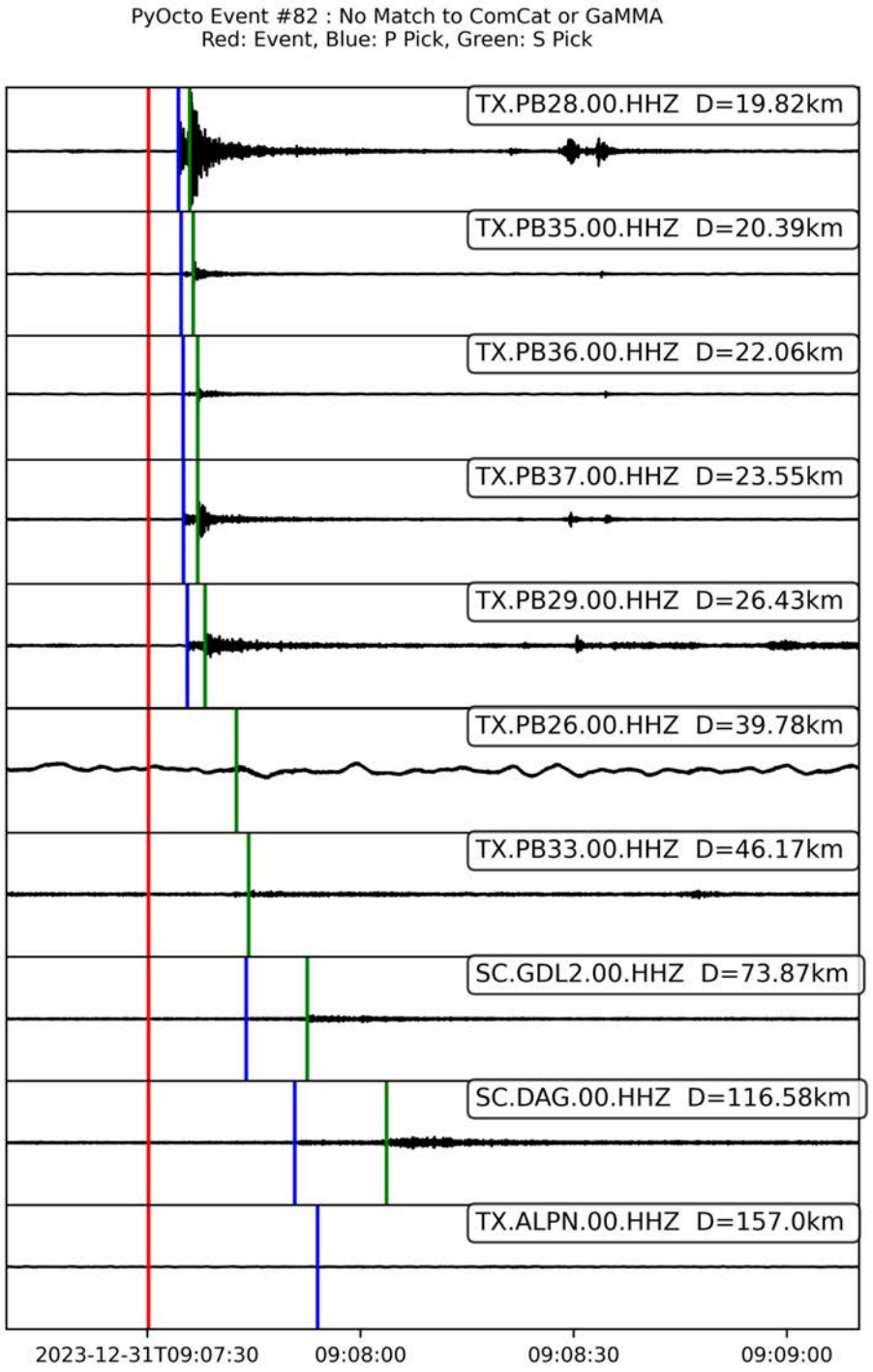
2020-01-07T16:19:28.490000 58768.490000 18.146667 -66.835500 **37.840000 2.980000** 1001012

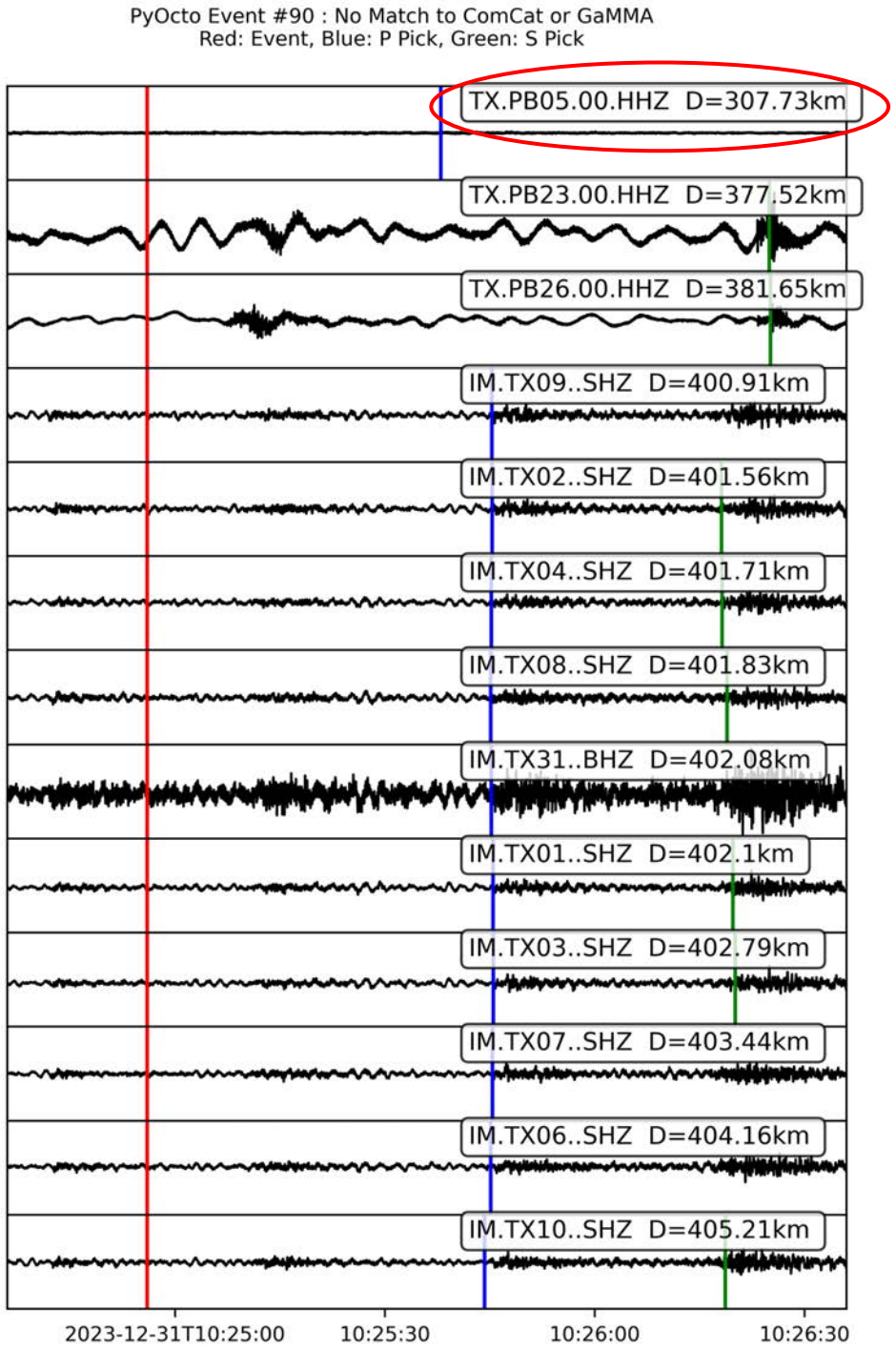


Post-processing: automatic removal of false detections in coda of larger earthquakes

- 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

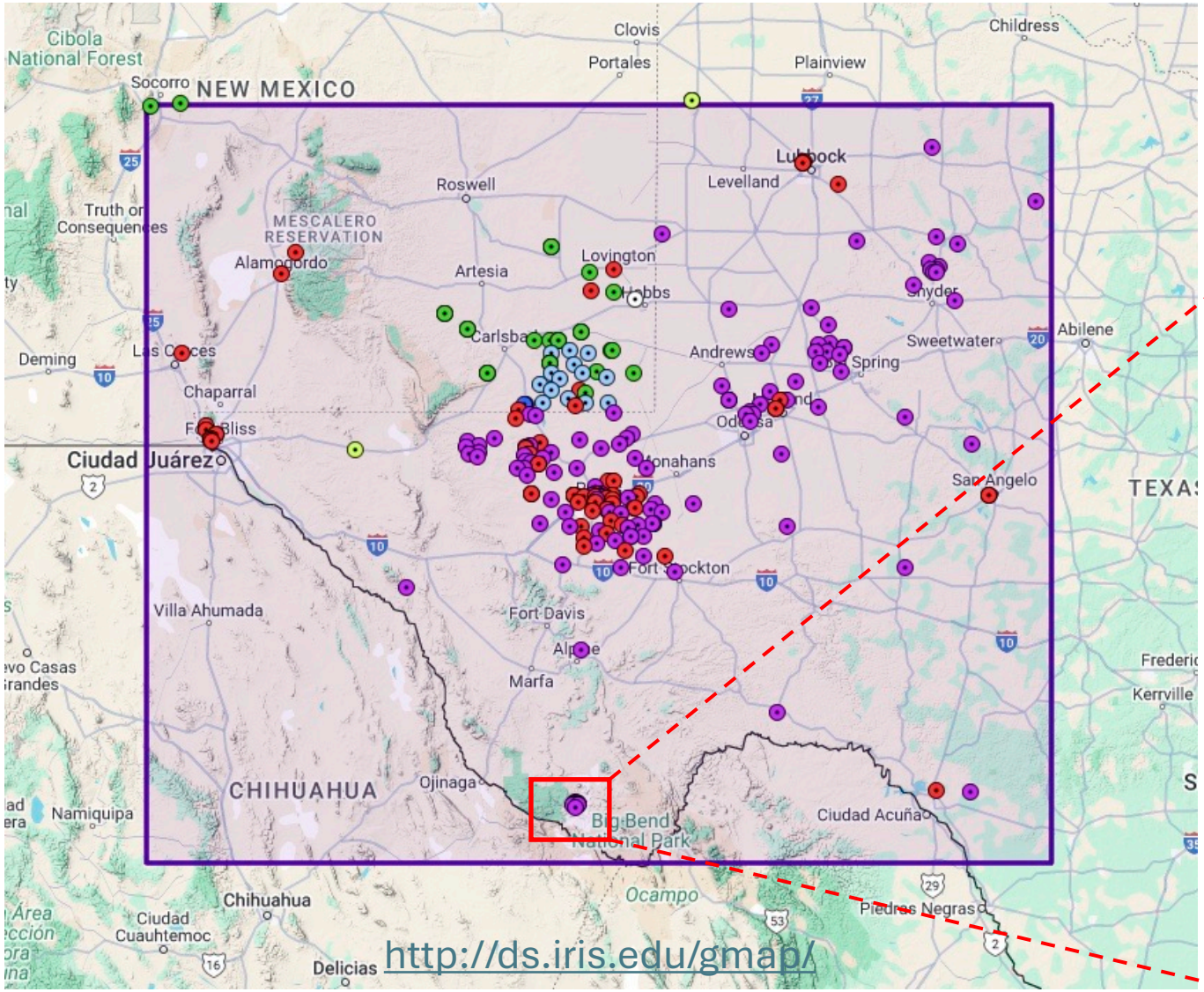




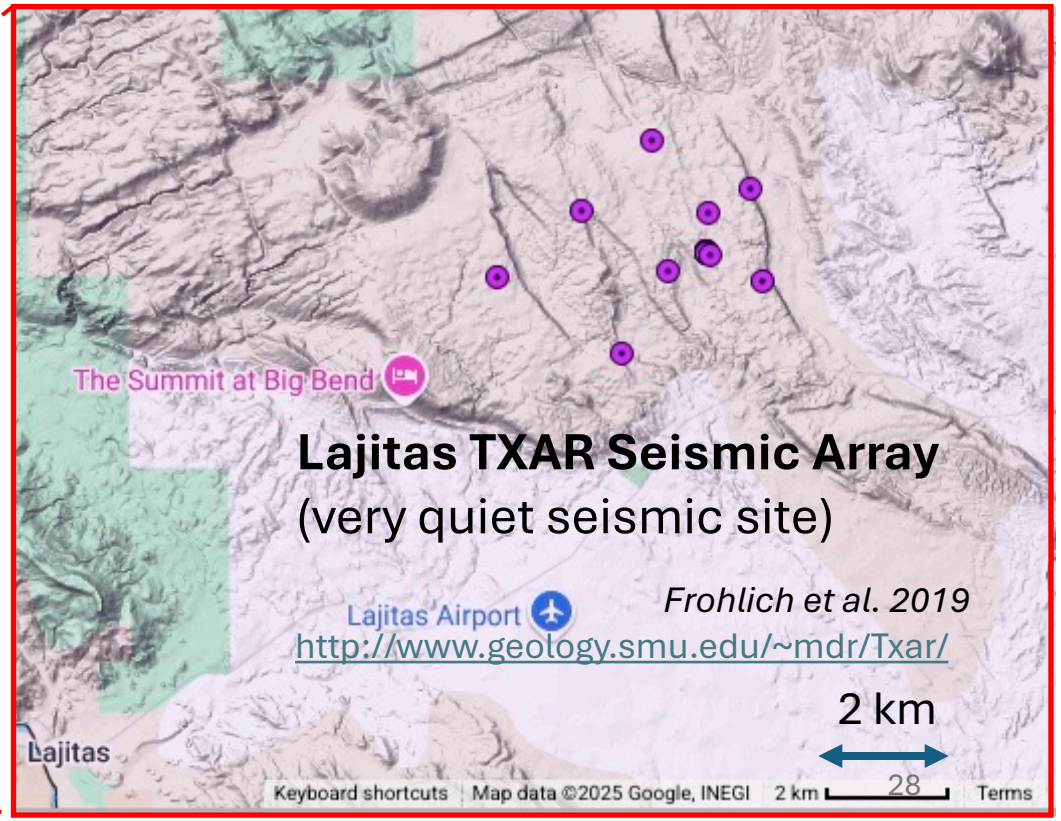
- Catalog QC: West Texas
- After event association
- (Left) new local earthquake 😀
- (Right) false detection ☹️
  - *large distance (307 km) to nearest station; IM network?*

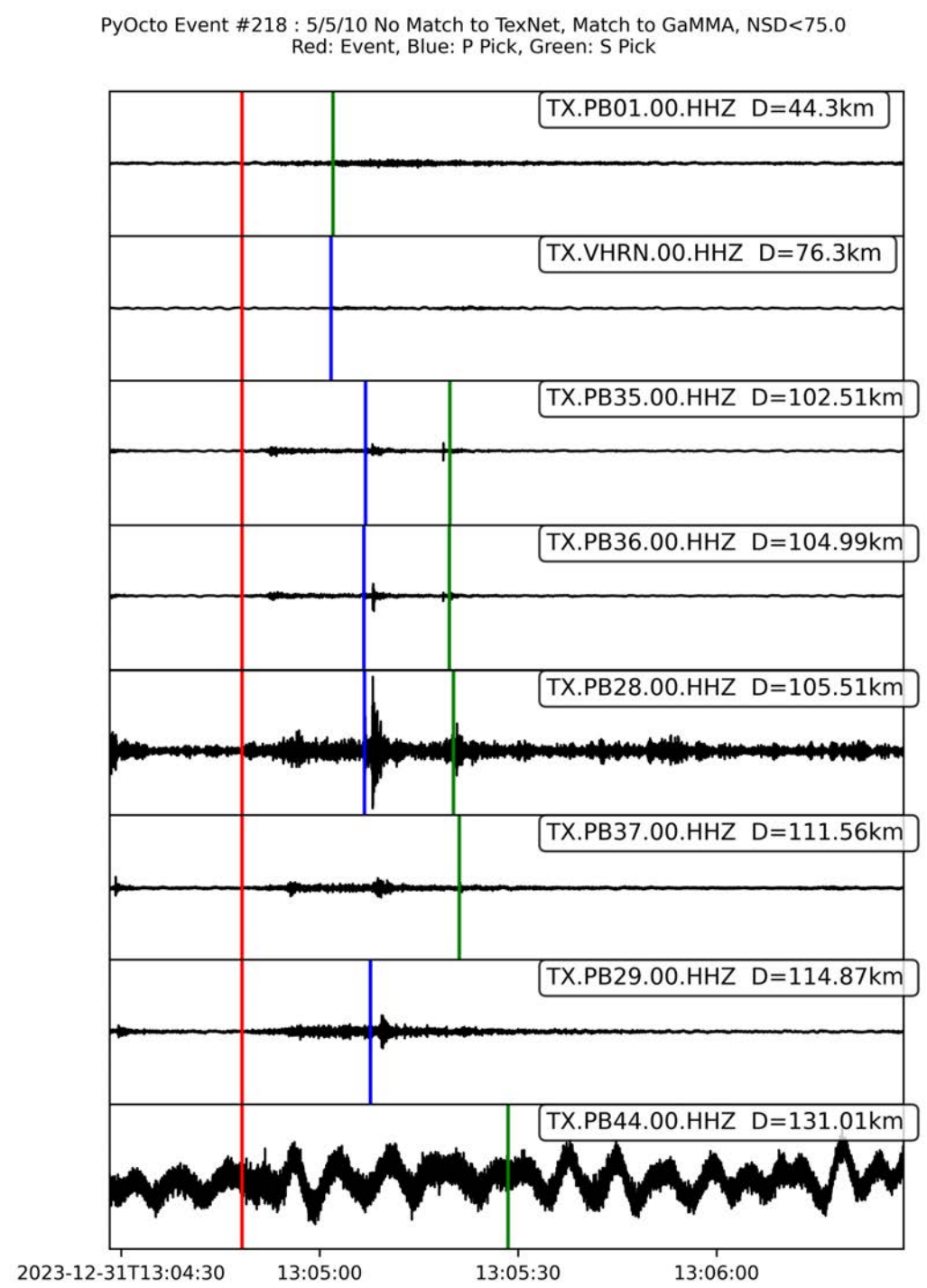
2-minute waveforms
Vertical lines:
Origin time, P, S

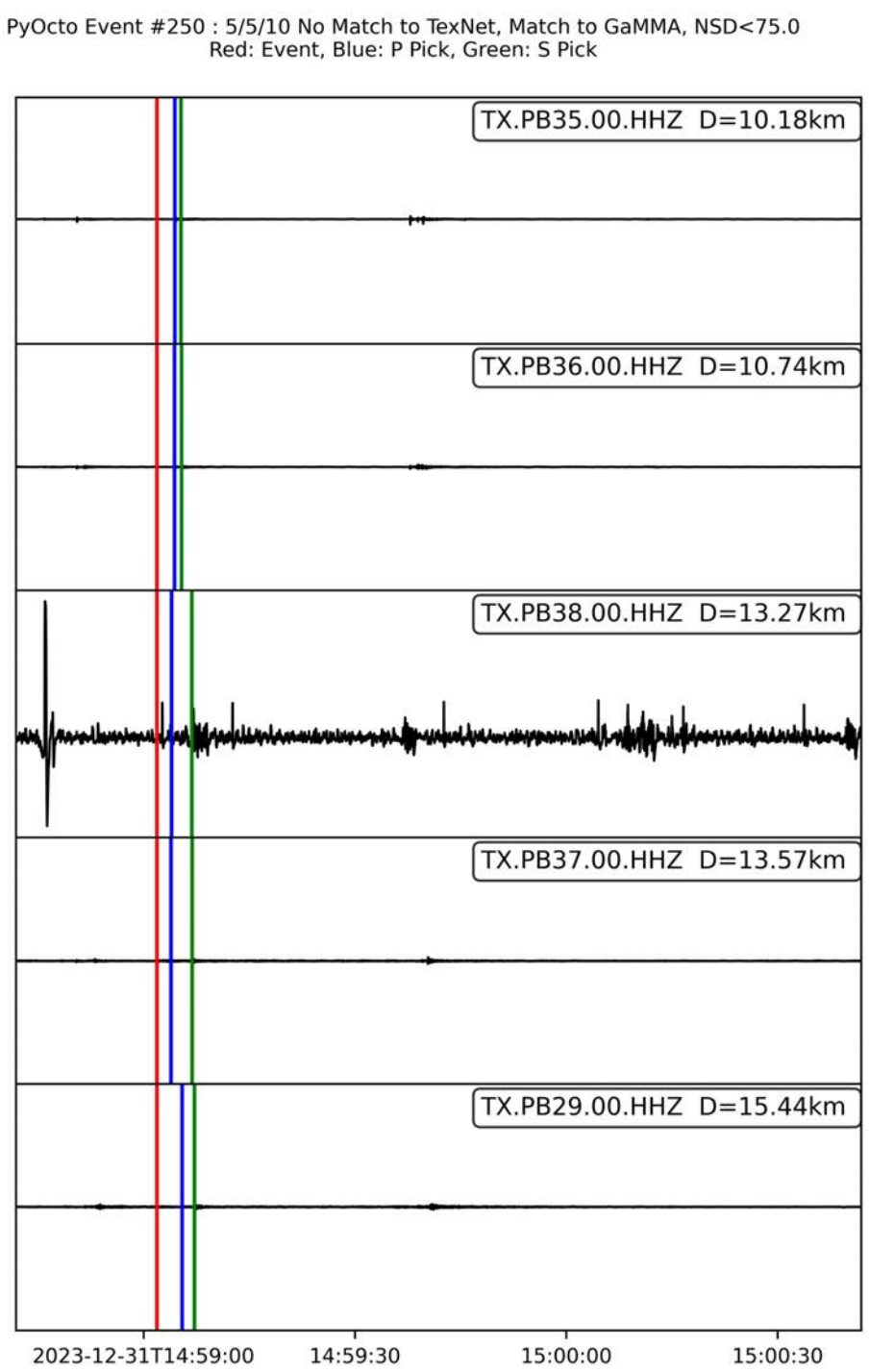
Problem: Incorrect association/location due to **uneven station distribution** (array far from rest of network & seismic sources)



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

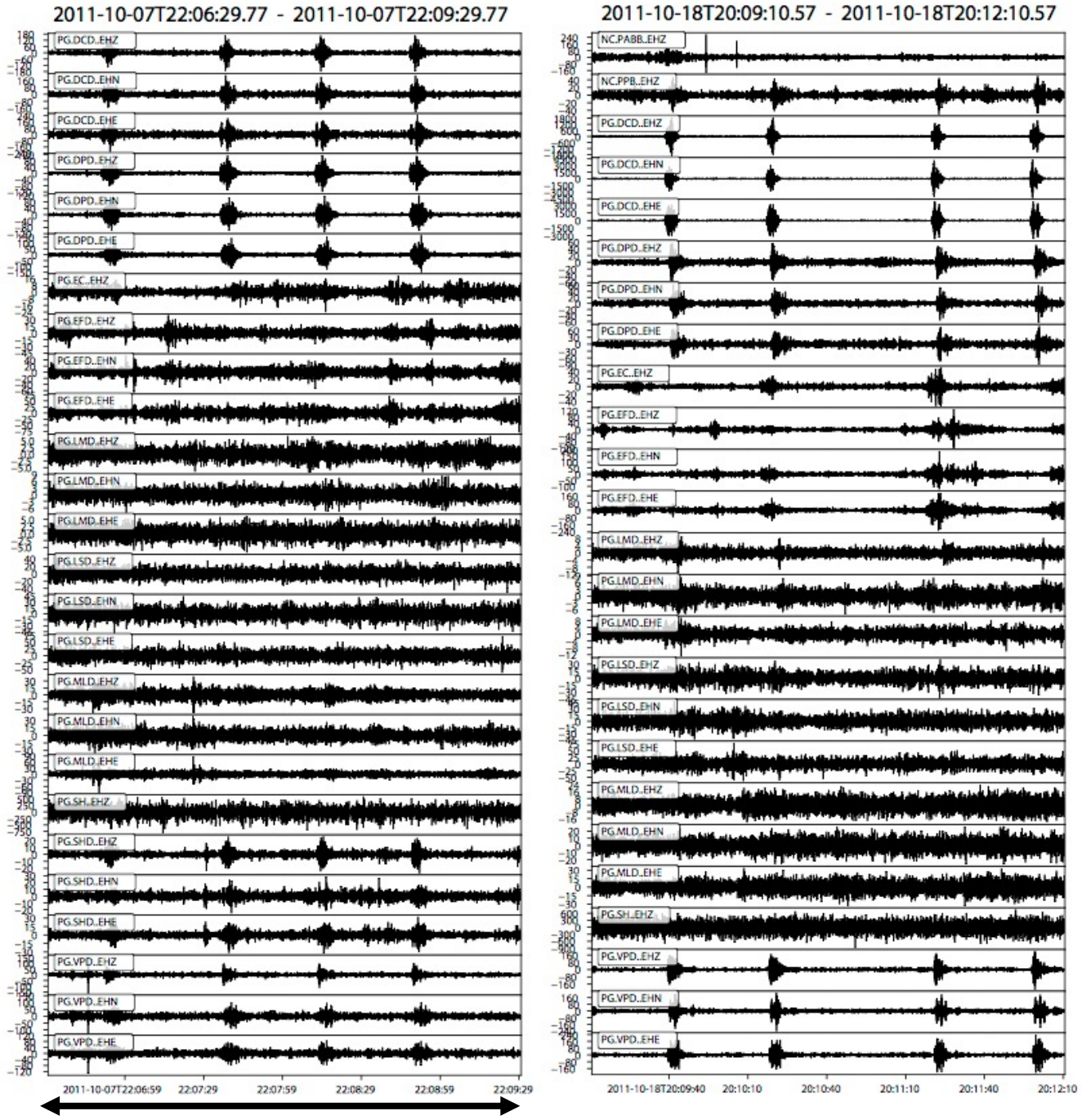






- Catalog QC: West Texas
- **Bad** event association ☹
- (Left) P, S picks from 2 different earthquakes associated as one event
- (Right) Noise picks associated as event → wrong moveout

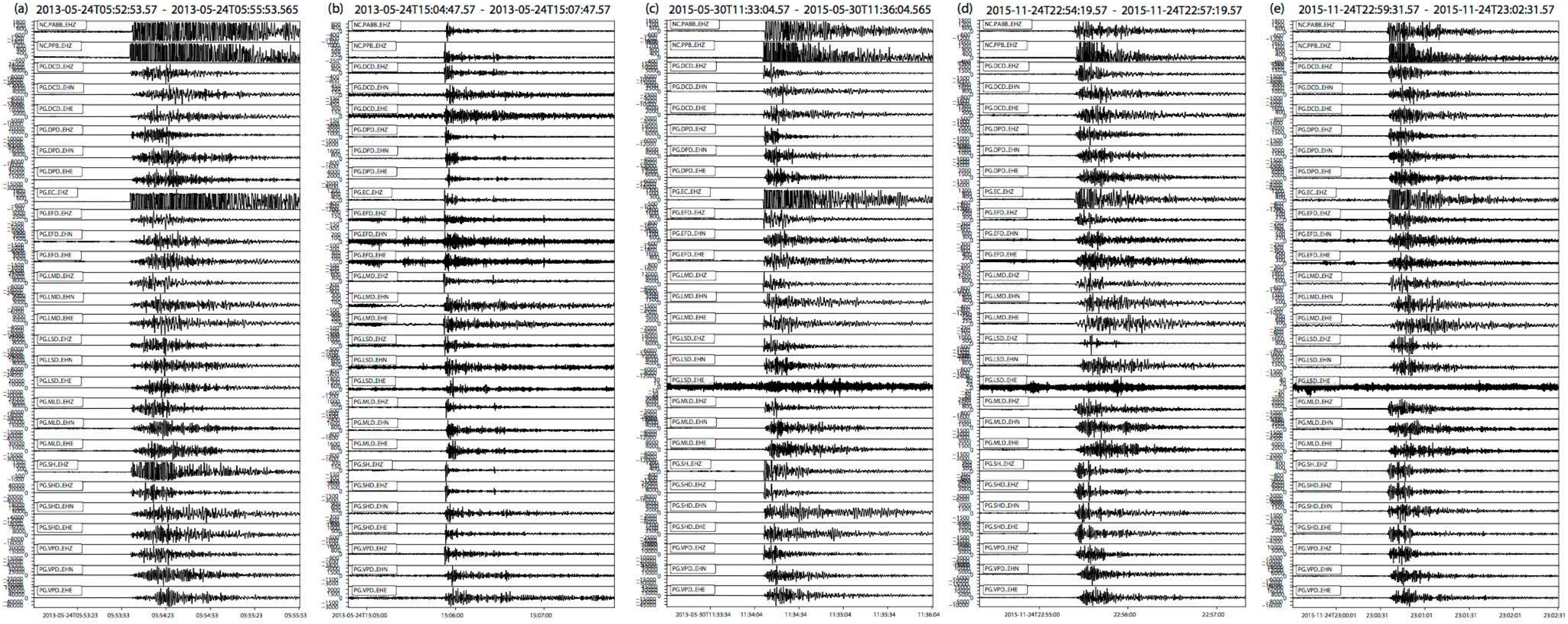
2-minute waveforms
Vertical lines:
Origin time, P, S



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

Yoon et al. (2019), BSSA

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

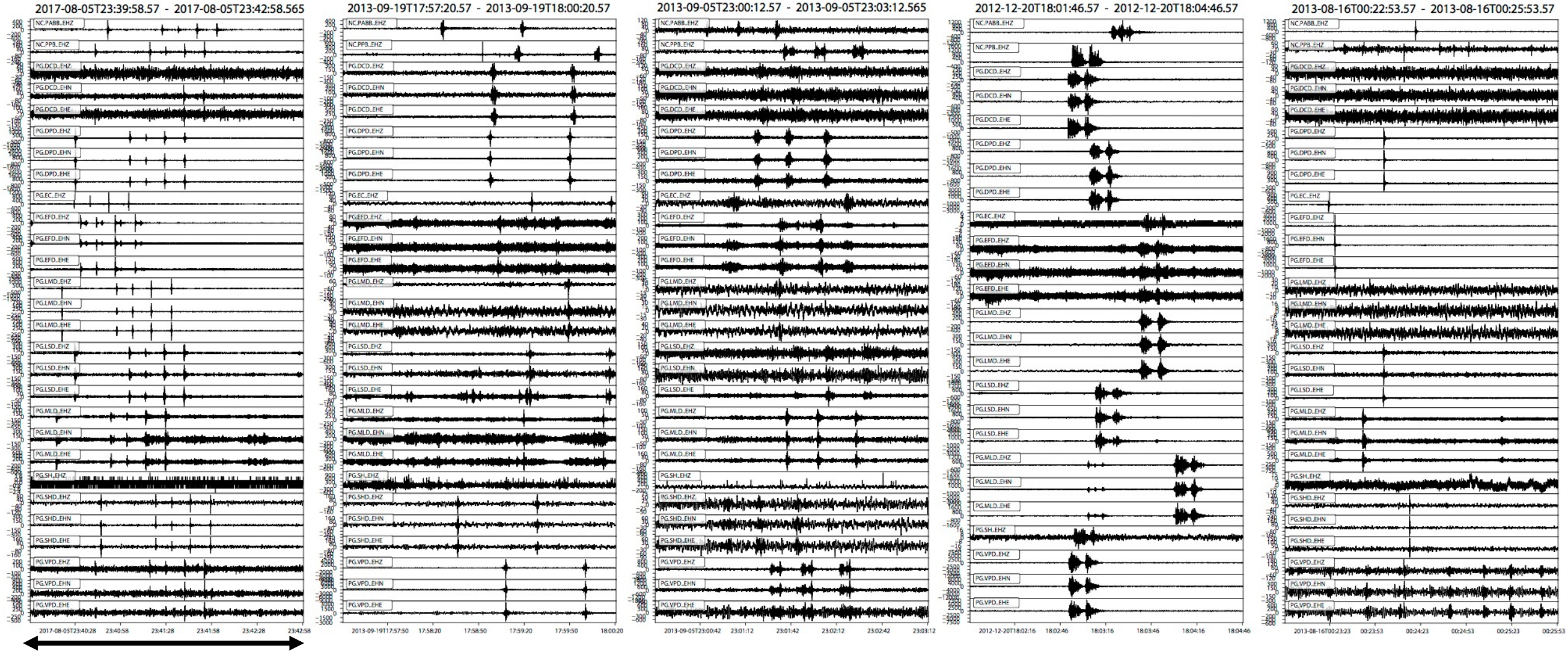


3 minutes

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

Yoon et al. (2019), BSSA

Infrasound signals recorded in Central CA



3 minutes

Slower moveout from sound waves

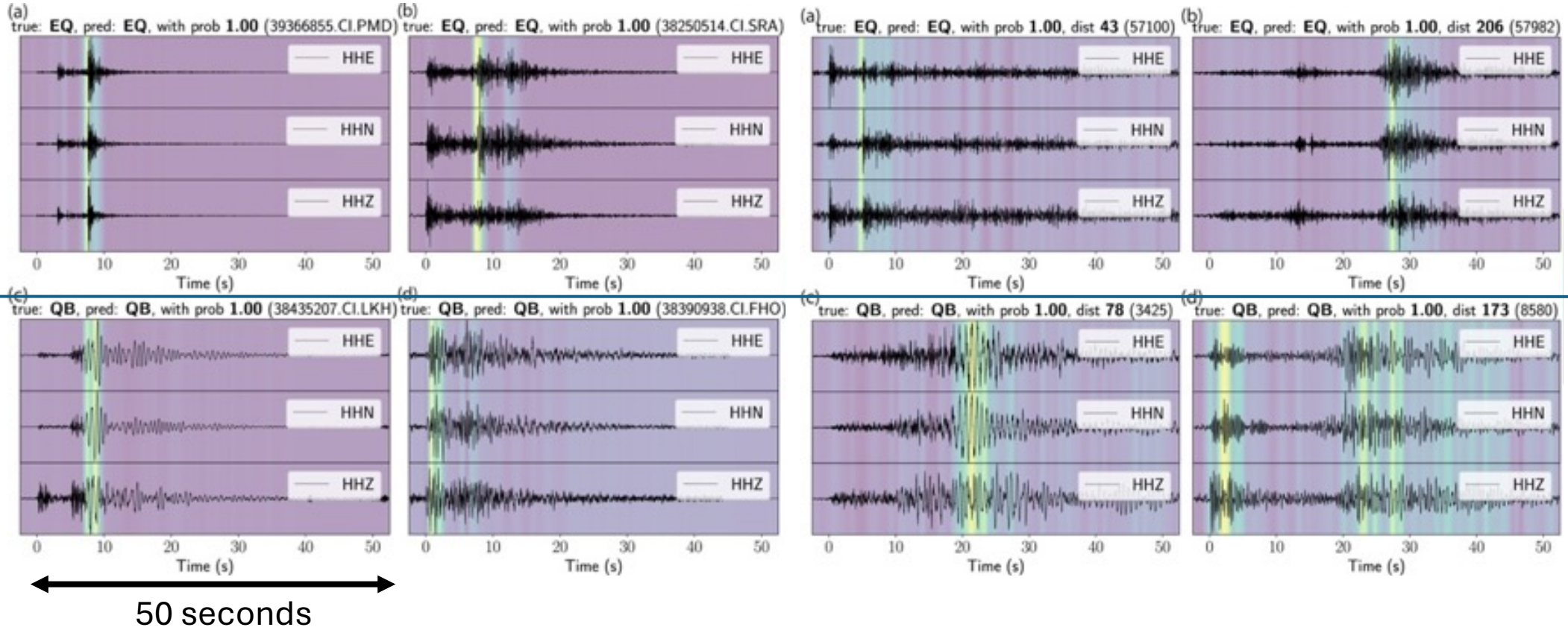
Yoon et al. (2019), BSSA

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

Check ComCat for blast events

earthquake

quarry blast



Check waveform plots: lower frequency content?

Events located near known quarry?

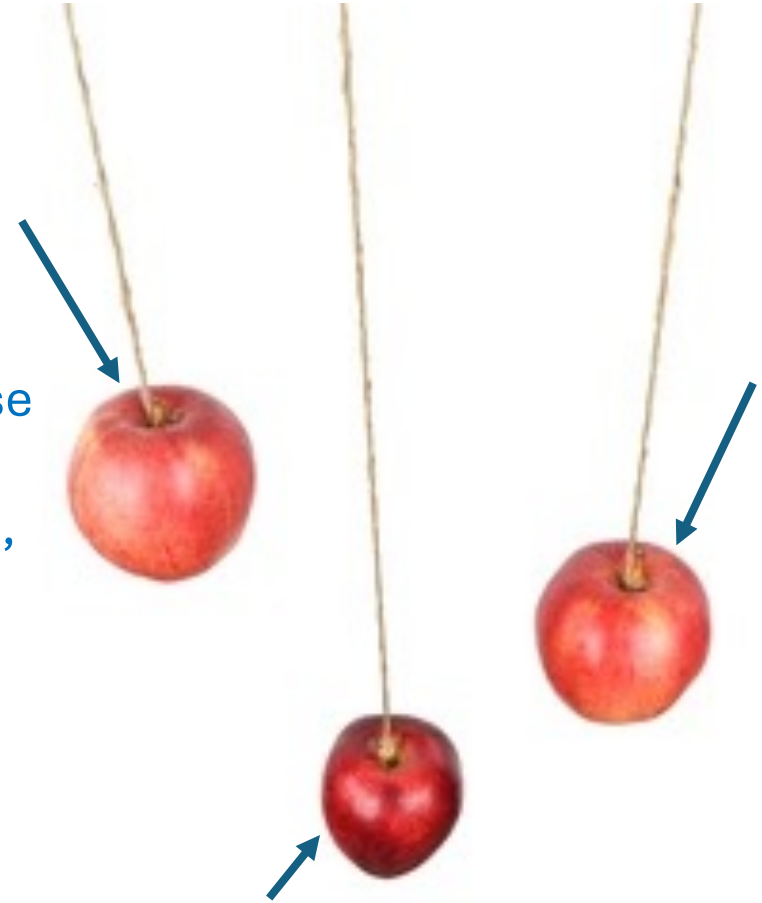
Events occur only in local daytime?

Figures: Zhu et al. (2024), GJI

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

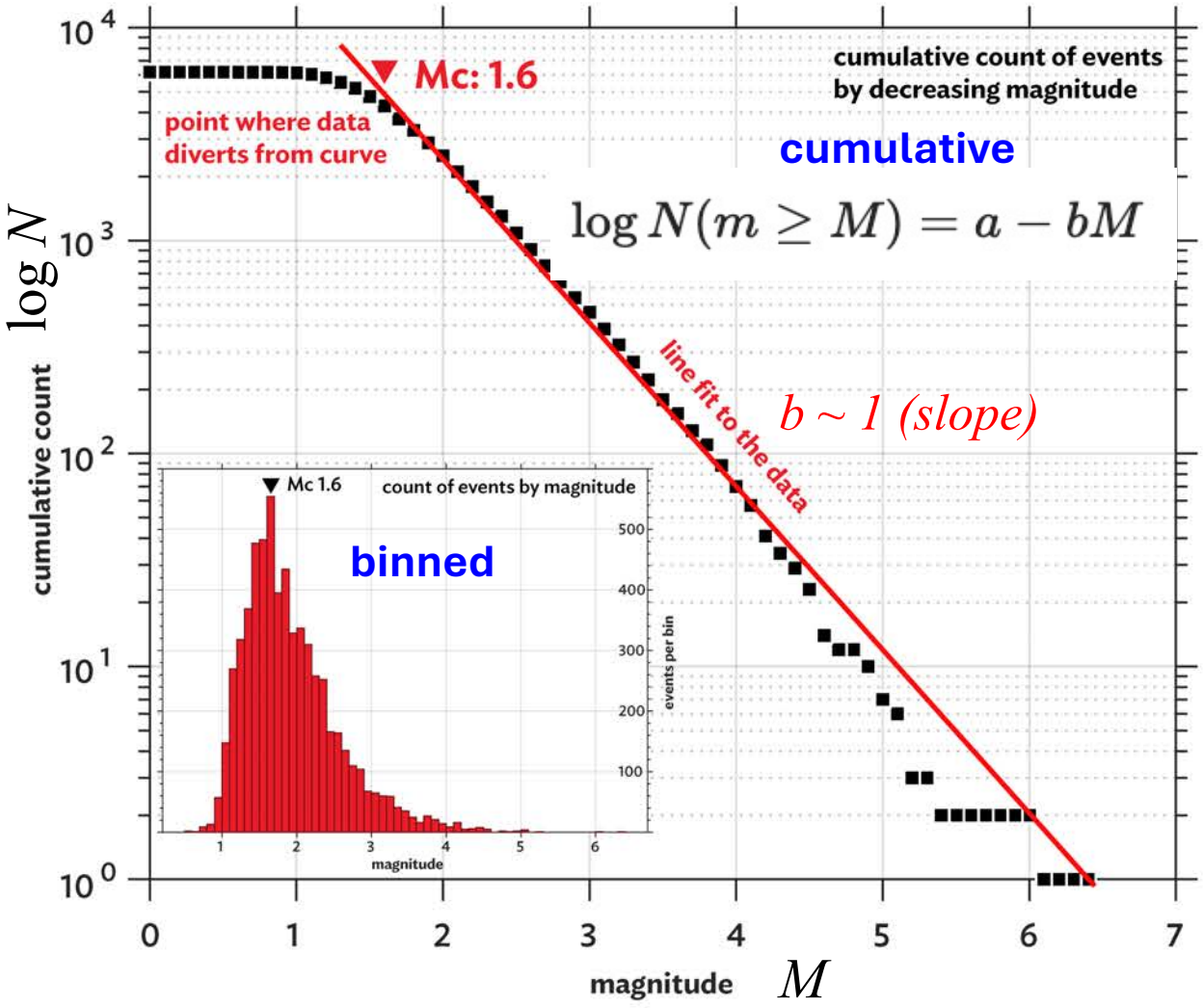
- Restrict locations to a certain region and/or depth range
- Set minimum threshold on:
  - Number of (P, S) picks
  - Number of stations per event
  - Output probabilities for P, S picks
- Set maximum bound on:
  - Distance to nearest station
  - Distance to farthest contributing station with picks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

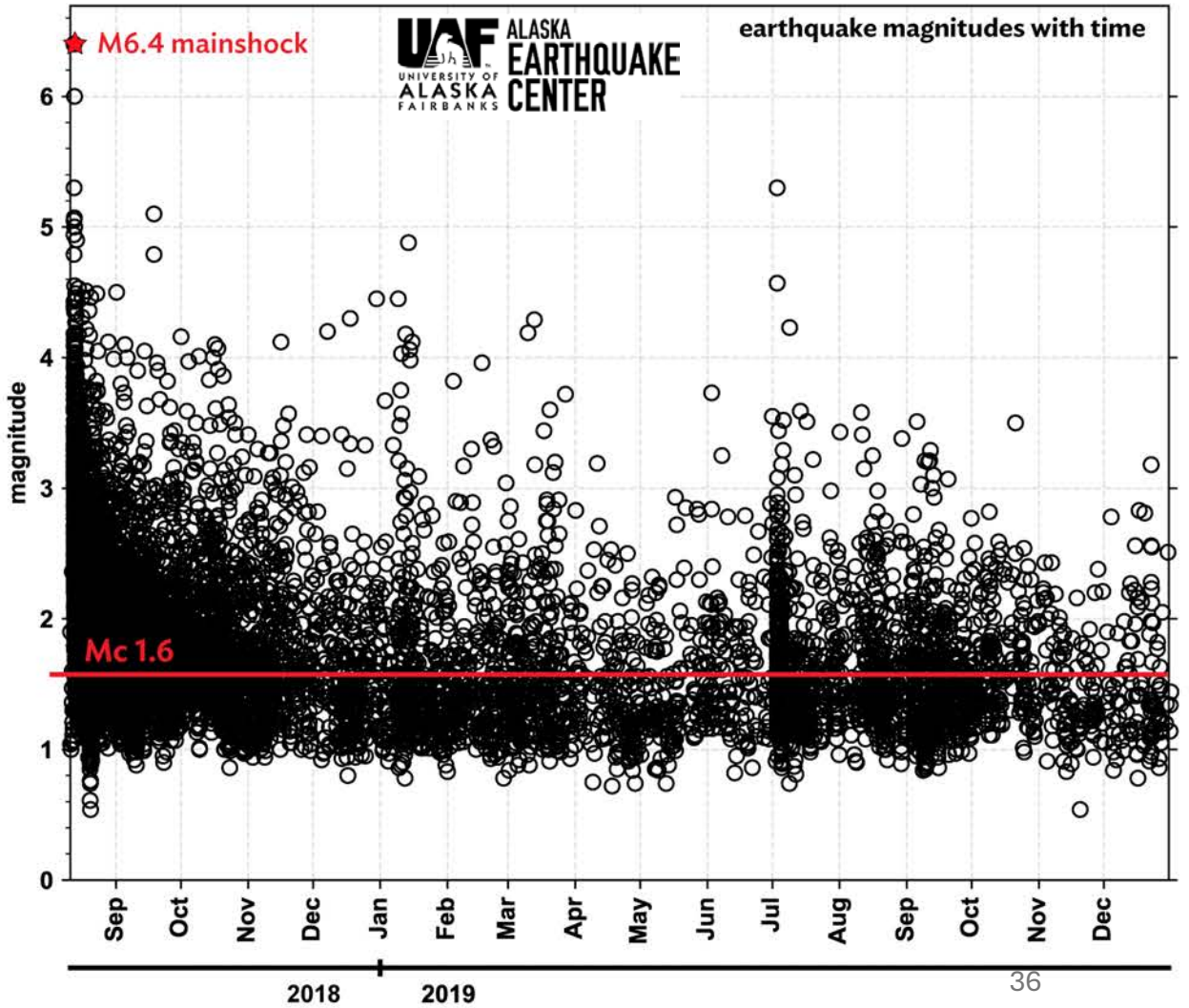
Magnitude of Completeness (Mc)

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

Lower Mc is better (more complete)

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





https://earthquake.alaska.edu/what-magnitude-completeness

Many different methods to get catalog Mc

- **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

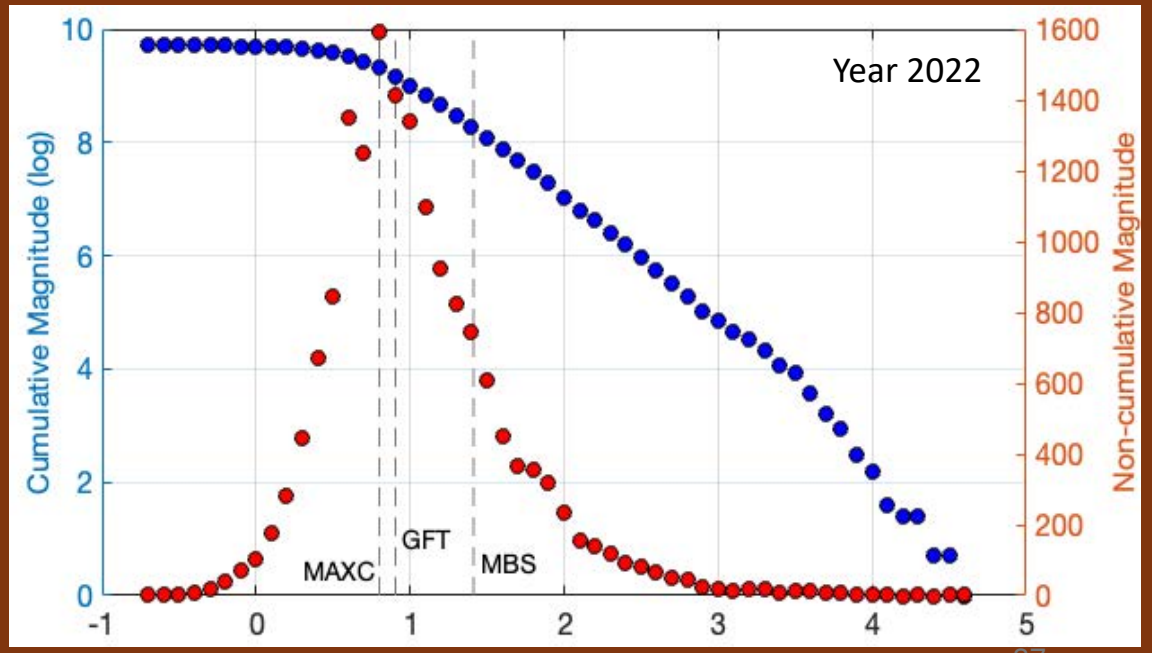
- **Goodness-of-fit Test (GFT)**: lowest magnitude cut-off where Gutenberg-Richter relation holds

- **Mc by b-value Stability (MBS):** first magnitude increment where $|b_{avg} - b| <$ uncertainty of b
  * tends to produce the highest (i.e., most conservative) Mc

* 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

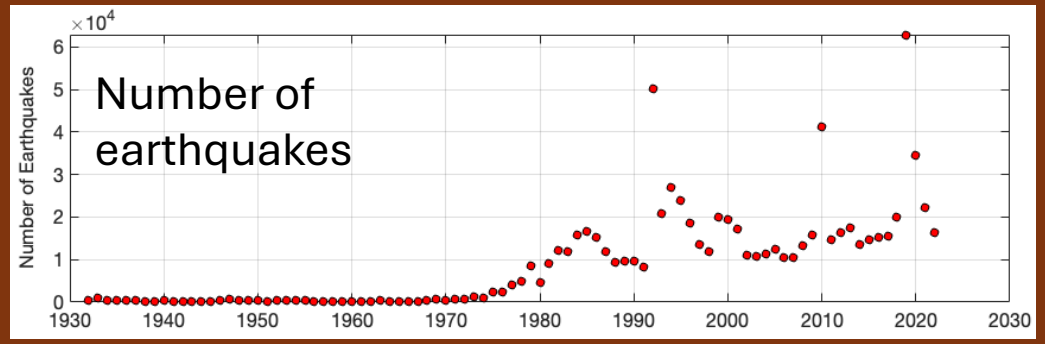
* Other methods to get Mc:
ZMAP (Wiemer, 2001)
b+ positive (van der Elst, 2021)

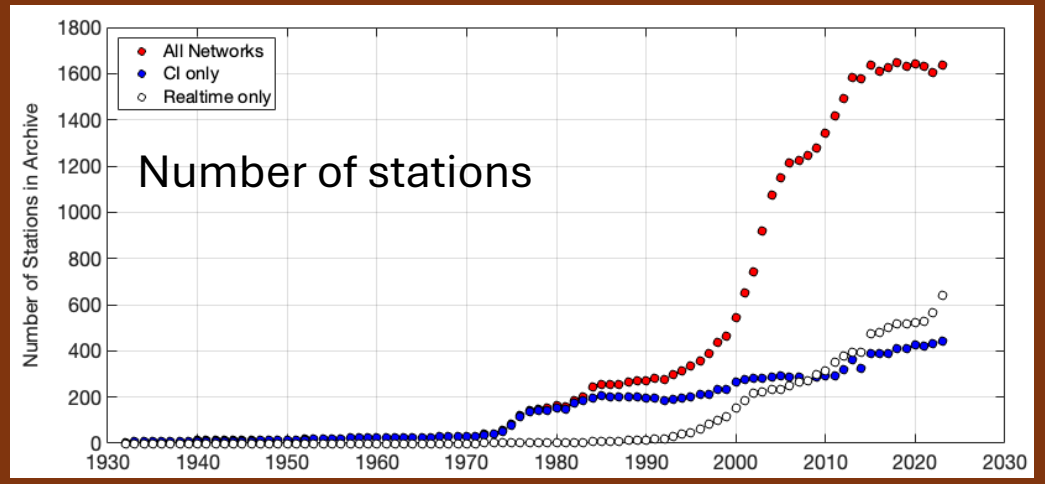


Temporal Mc changes

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

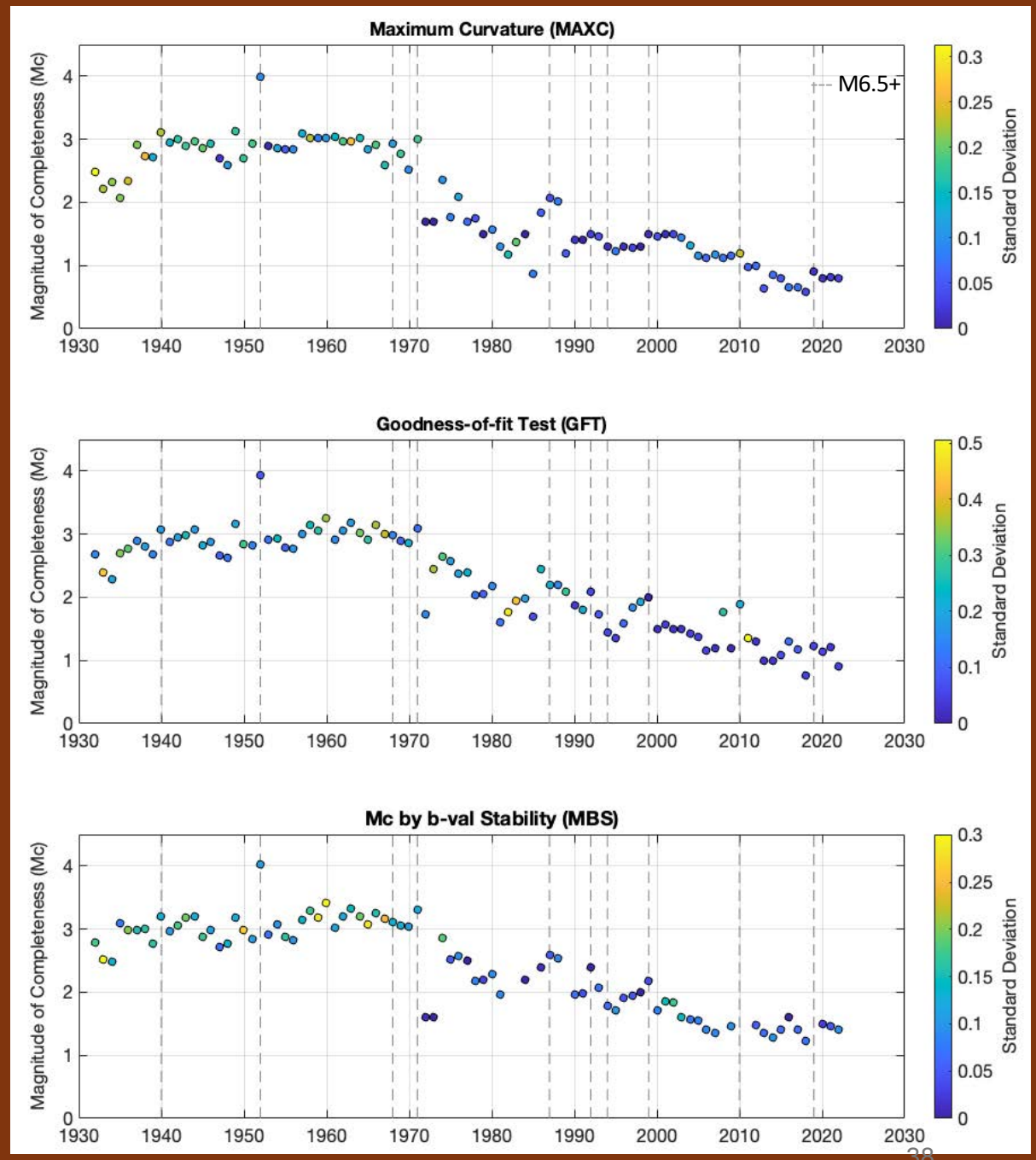
Full SCSN Catalog





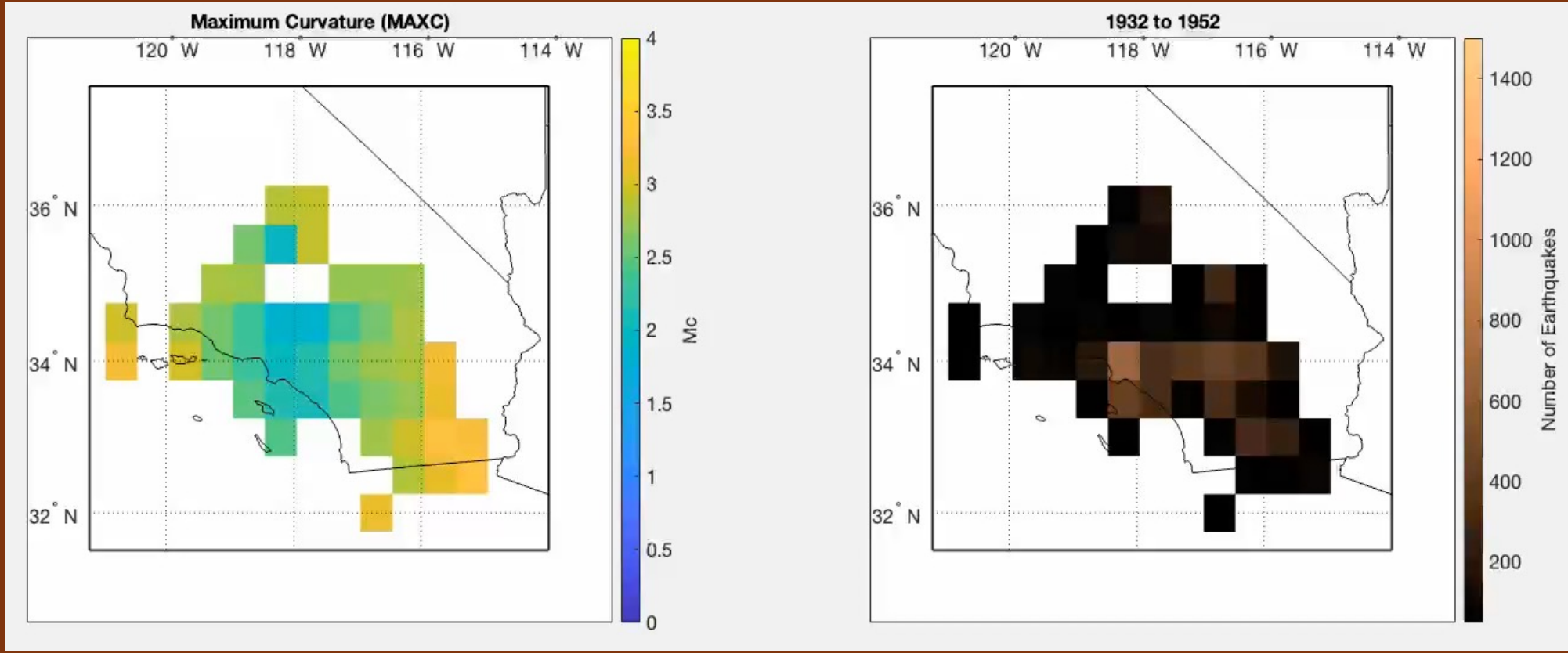
year

Mc



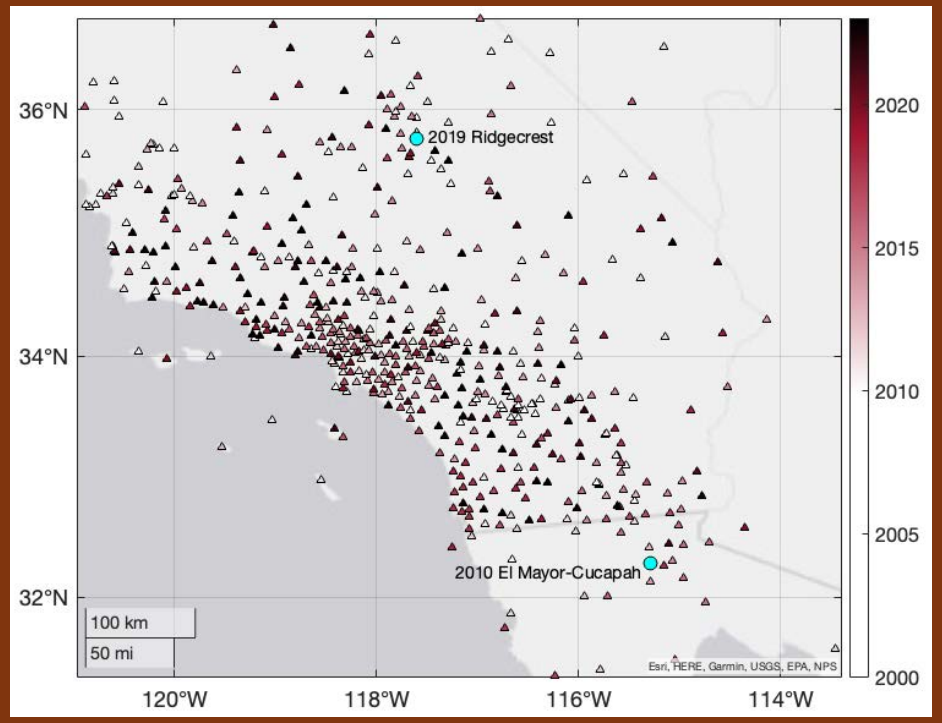
year

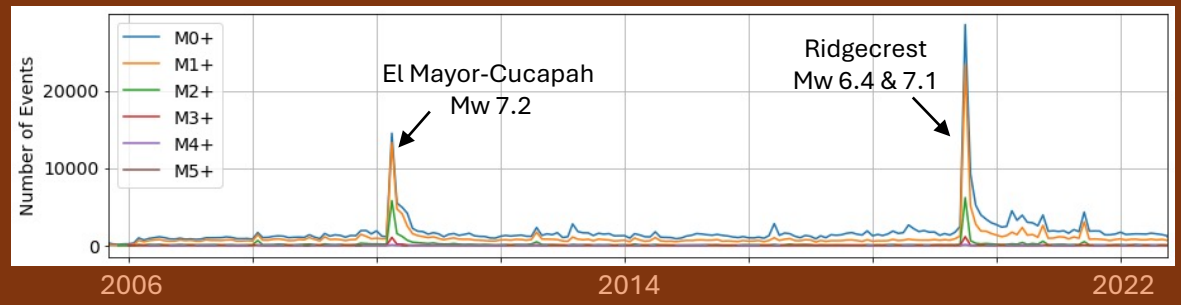
Spatial Mc Changes



- 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

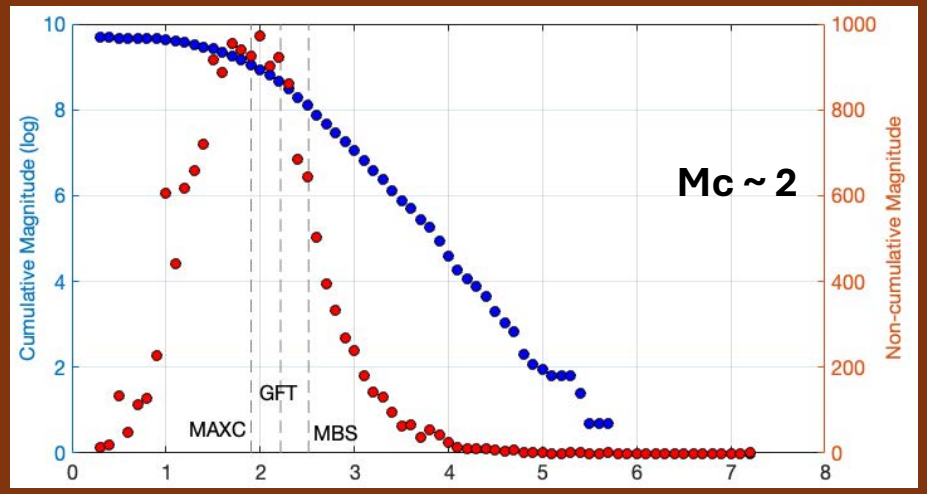




Expect higher Mc earlier in aftershock sequence

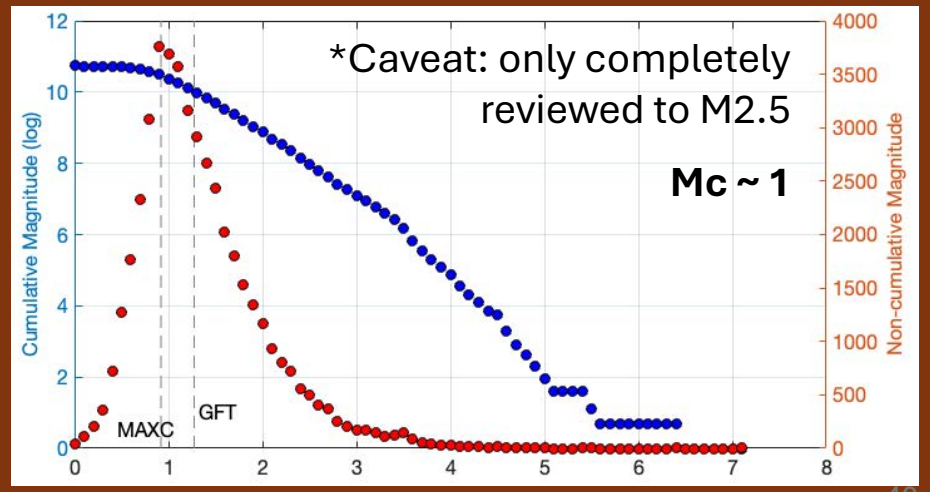
Ridgecrest Mc is ~1 mag. unit lower than El Mayor-Cucapah Mc

2010 M_w 7.2 El Mayor-Cucapah sequence



magnitude

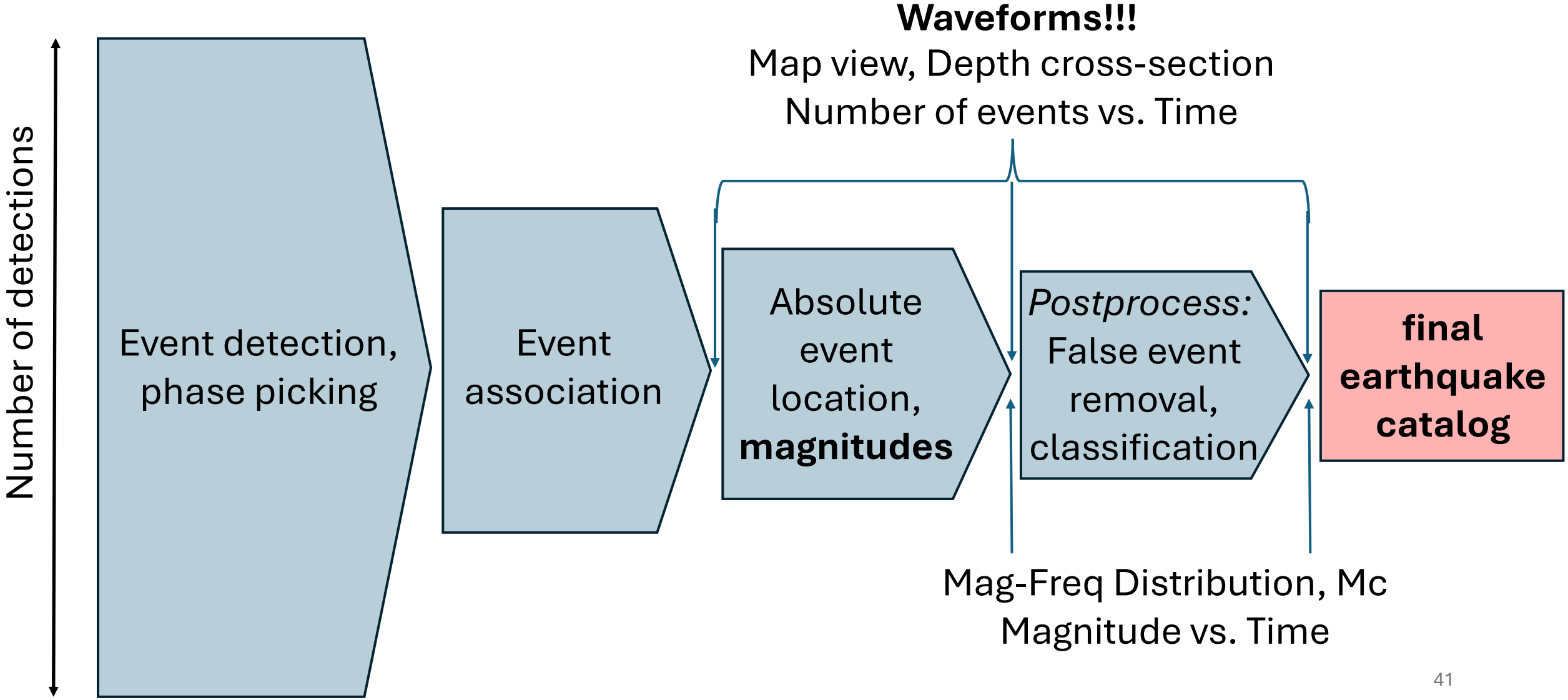
2019 M_w 6.4 & 7.1 Ridgecrest sequence



magnitude

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/teleseismic 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/

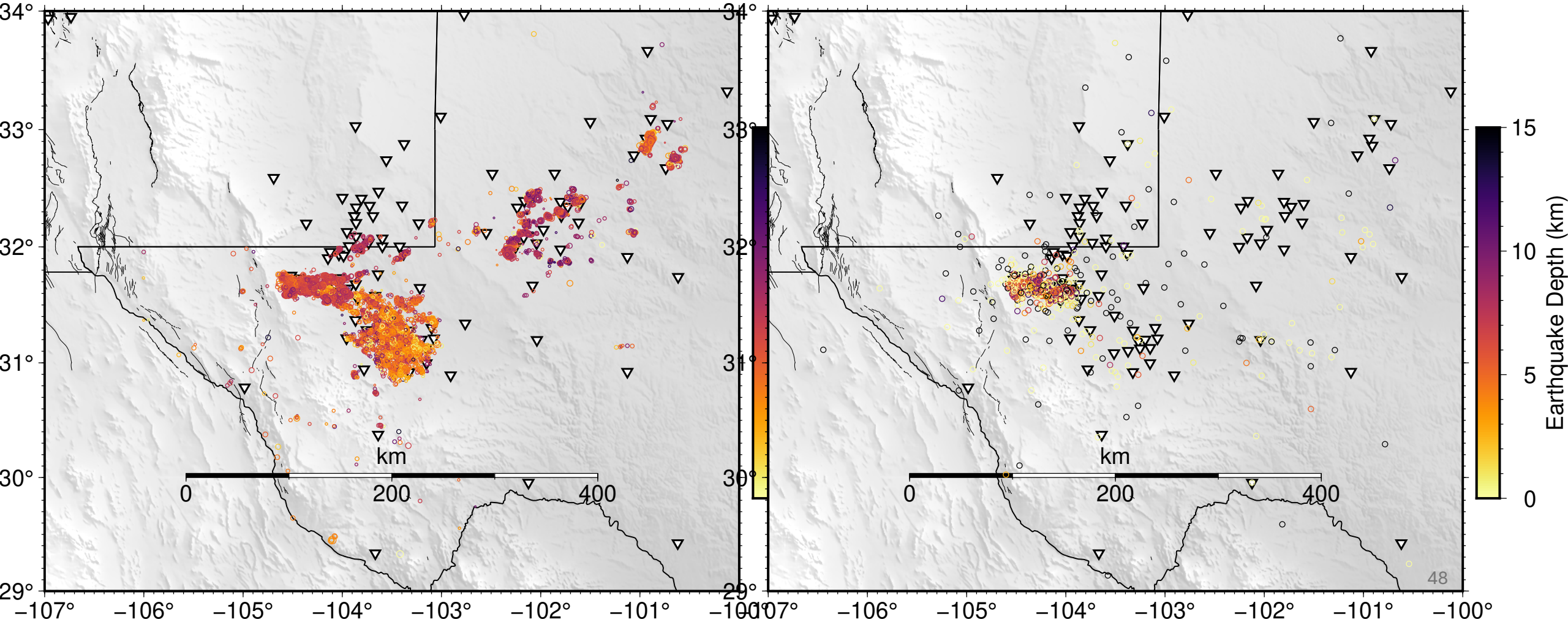
- **NEW** events in your enhanced catalog might be from reference catalog events outside your selected bounding region
  - Check waveforms: could be a regional distance earthquake if higher duration and amplitude

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)

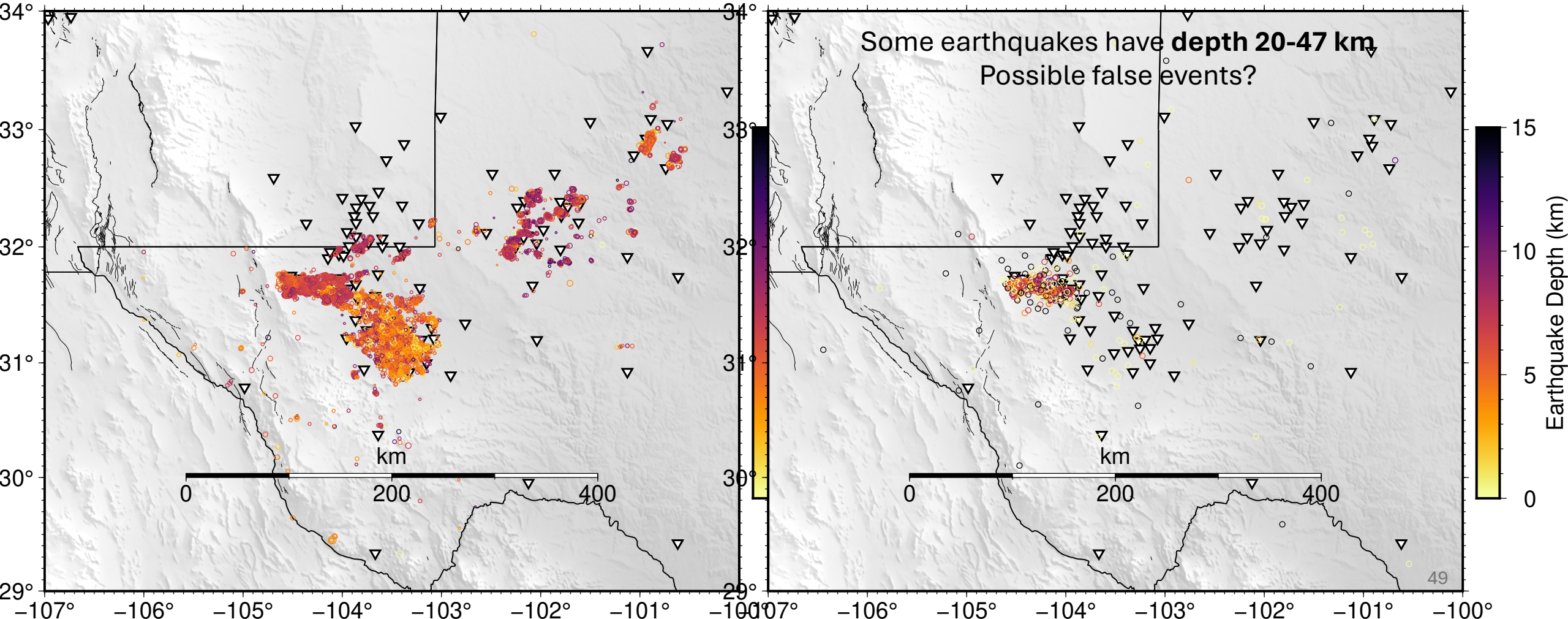


Catalog QC, locations in map view: TexNet

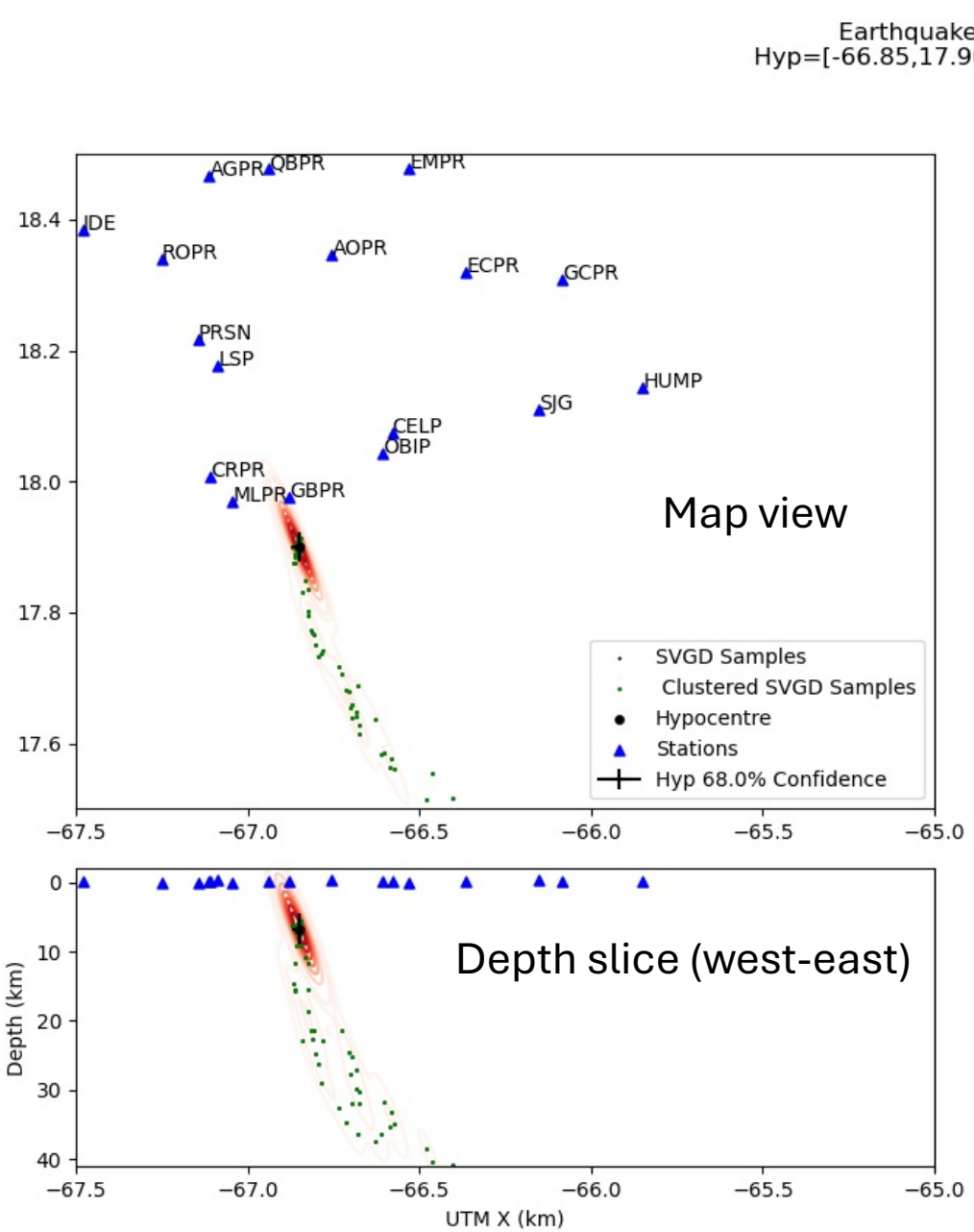
Removed false detections (fewer scattered events)
with stricter association criteria

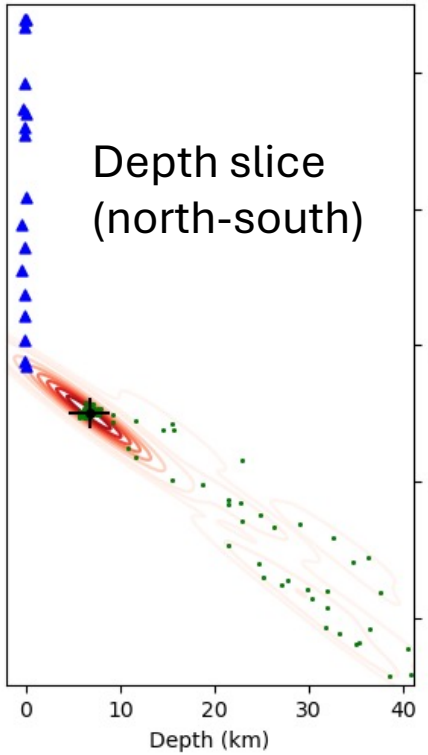
Reference catalog:
TexNet (2015-2024: 9 years)

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



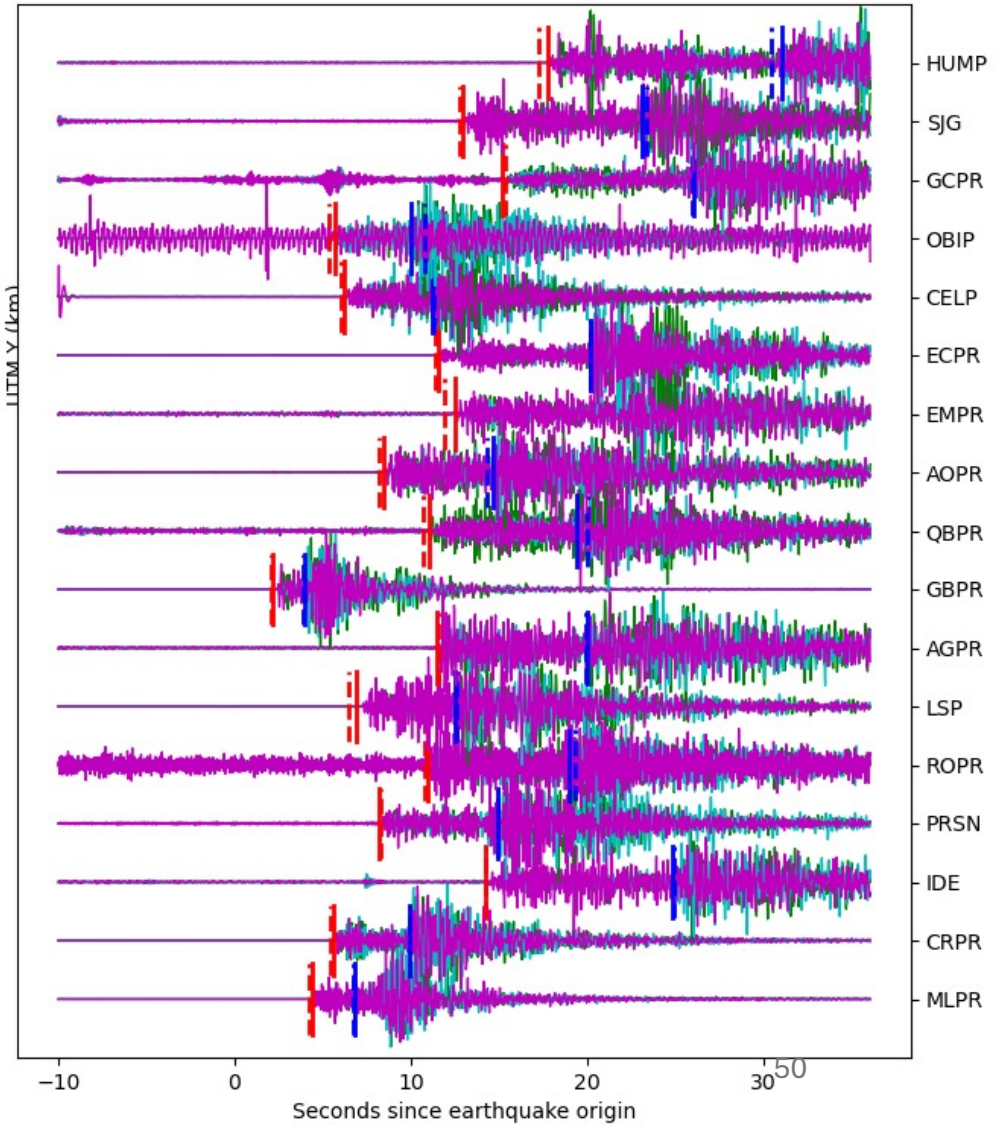
Catalog QC: 2020 M2.8 earthquake, southwest Puerto Rico





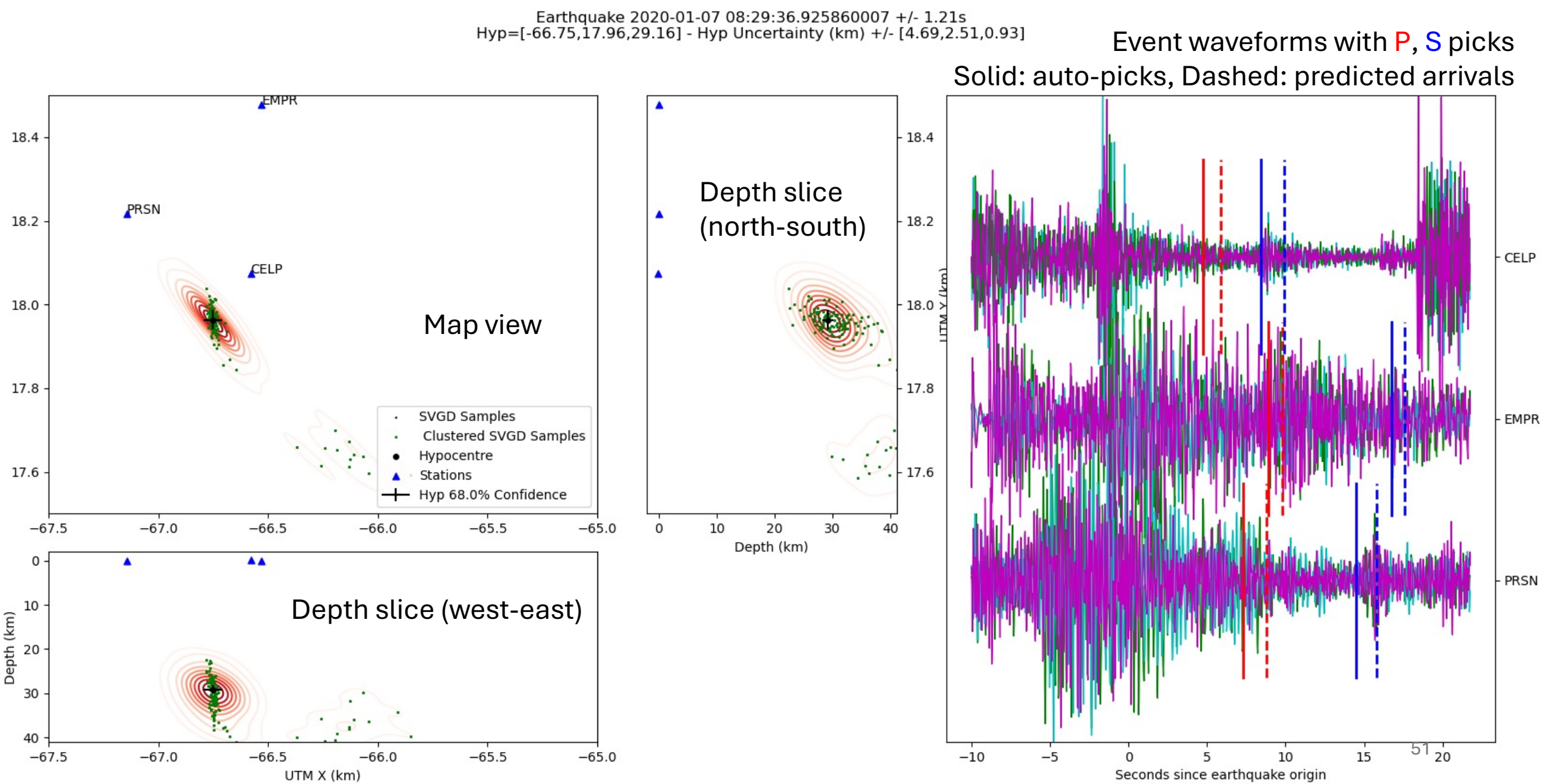
Event waveforms with P, S picks

Solid: auto-picks, Dashed: predicted arrivals



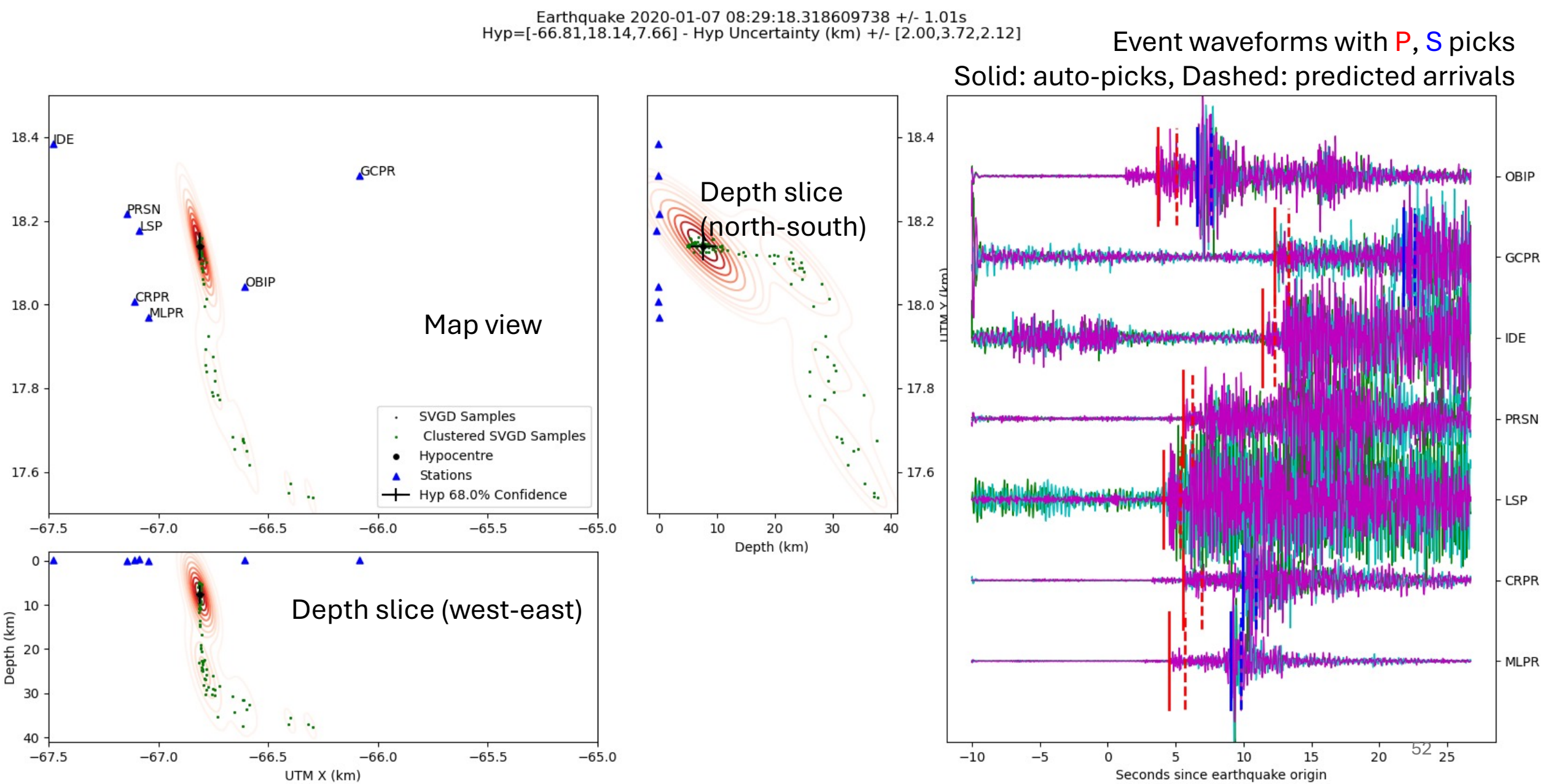
Catalog QC: False detection in coda of larger earthquake (also, too deep?)

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



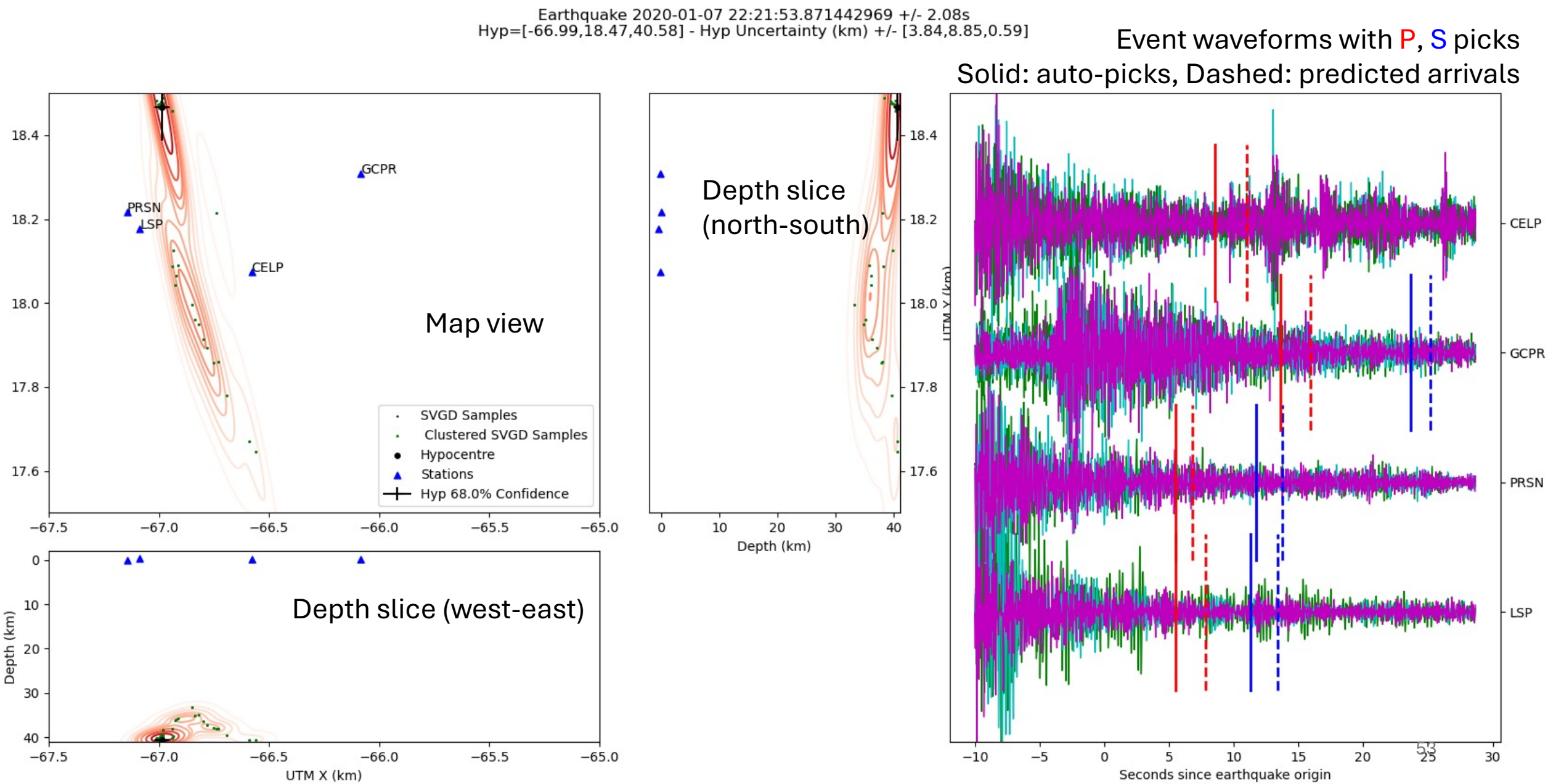
Catalog QC: Real (newly detected) earthquake

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



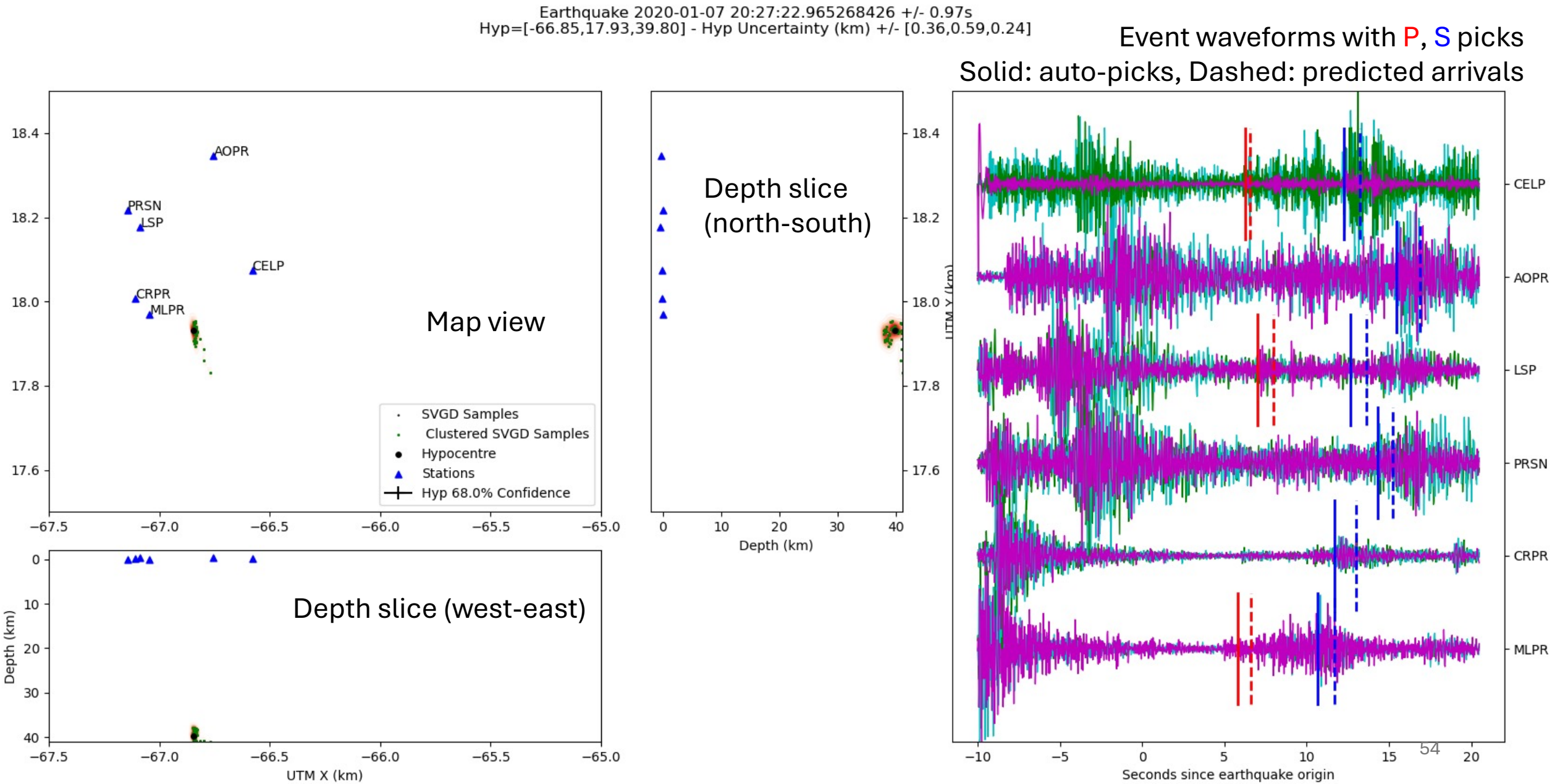
Catalog QC: False detection in coda of larger earthquake (also, too deep?)

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



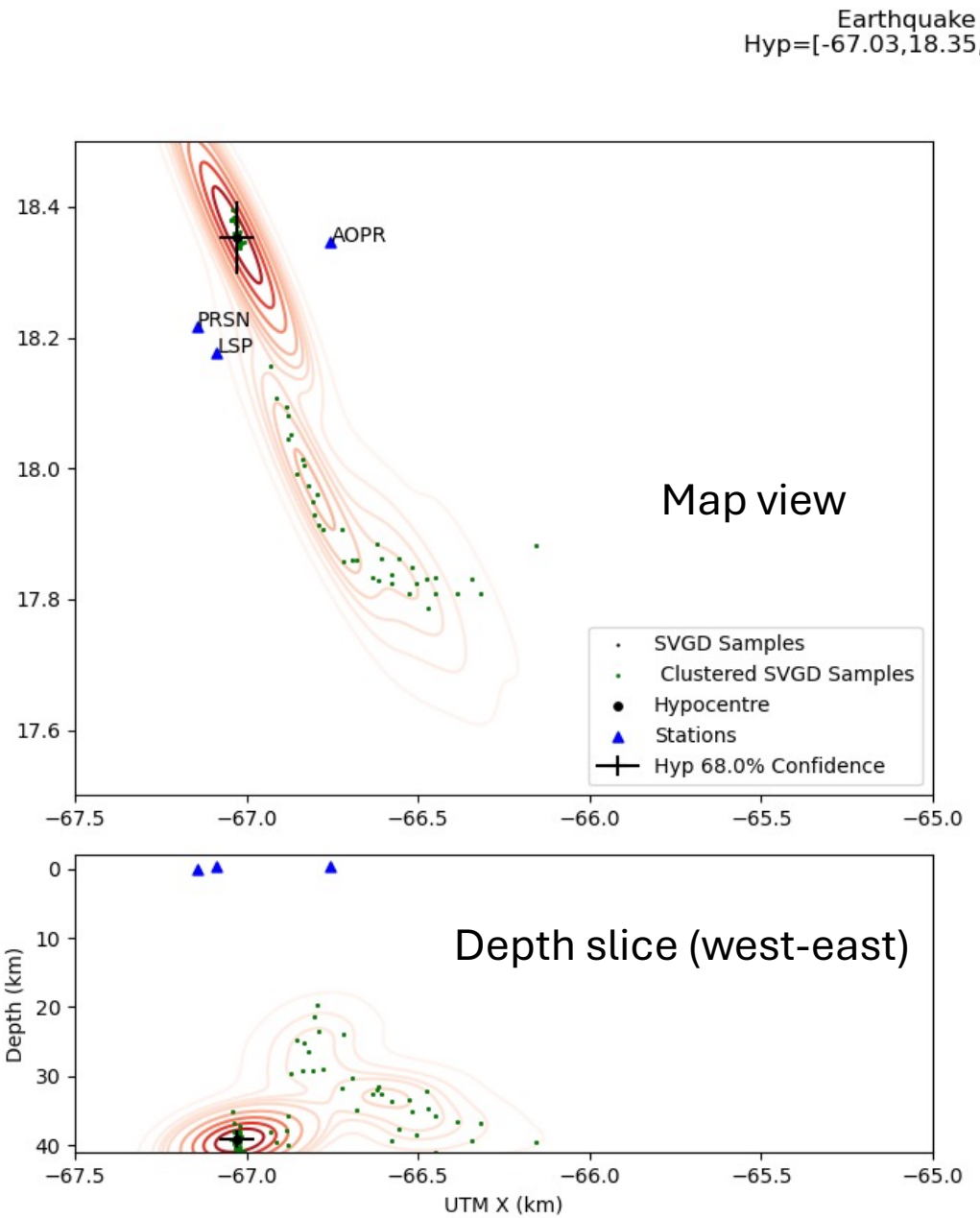
Catalog QC: False detection in coda of larger earthquake (also, too deep?)

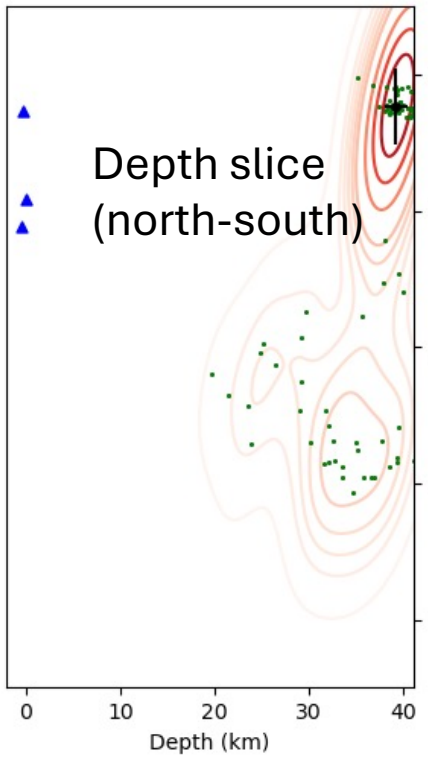
`2020-01-07T20:27:22.970000 73642.970000 17.931000 -66.846500 **39.800000 1.560000**  1001320`



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

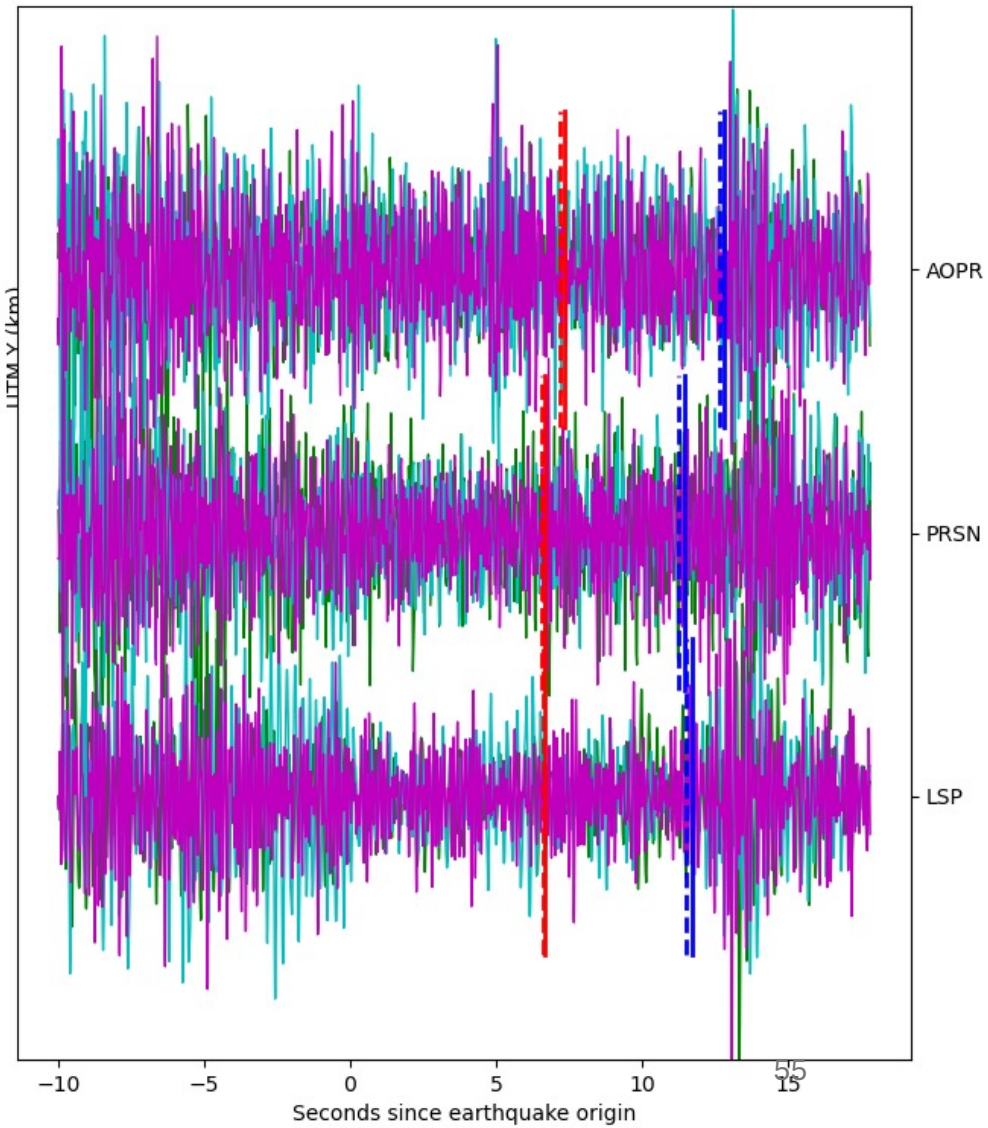
`2020-01-07T15:58:52.370000 57532.370000 18.353667 -67.028000` **`39.160000 1.310000`** `1000980`





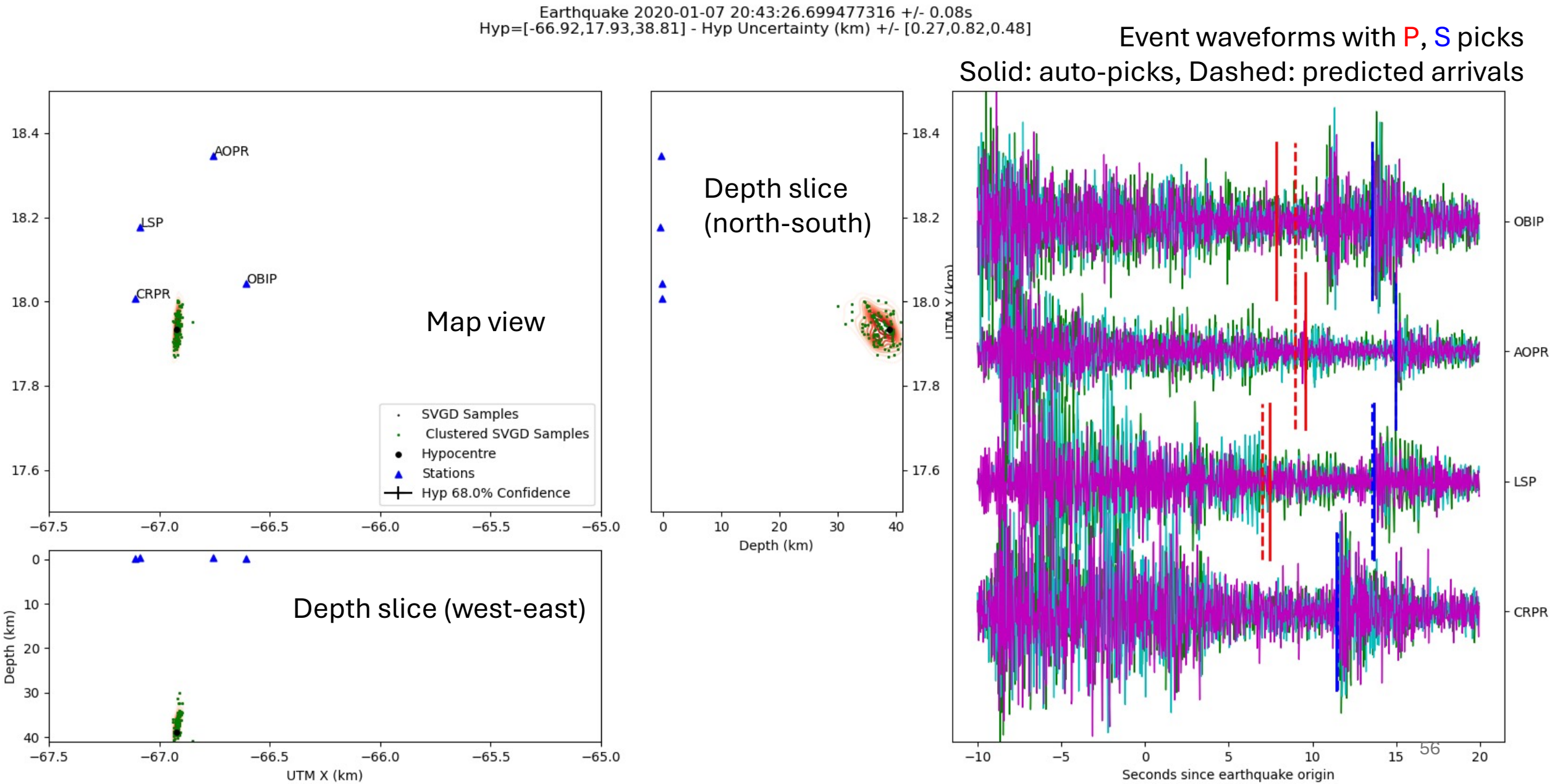
Event waveforms with P, S picks

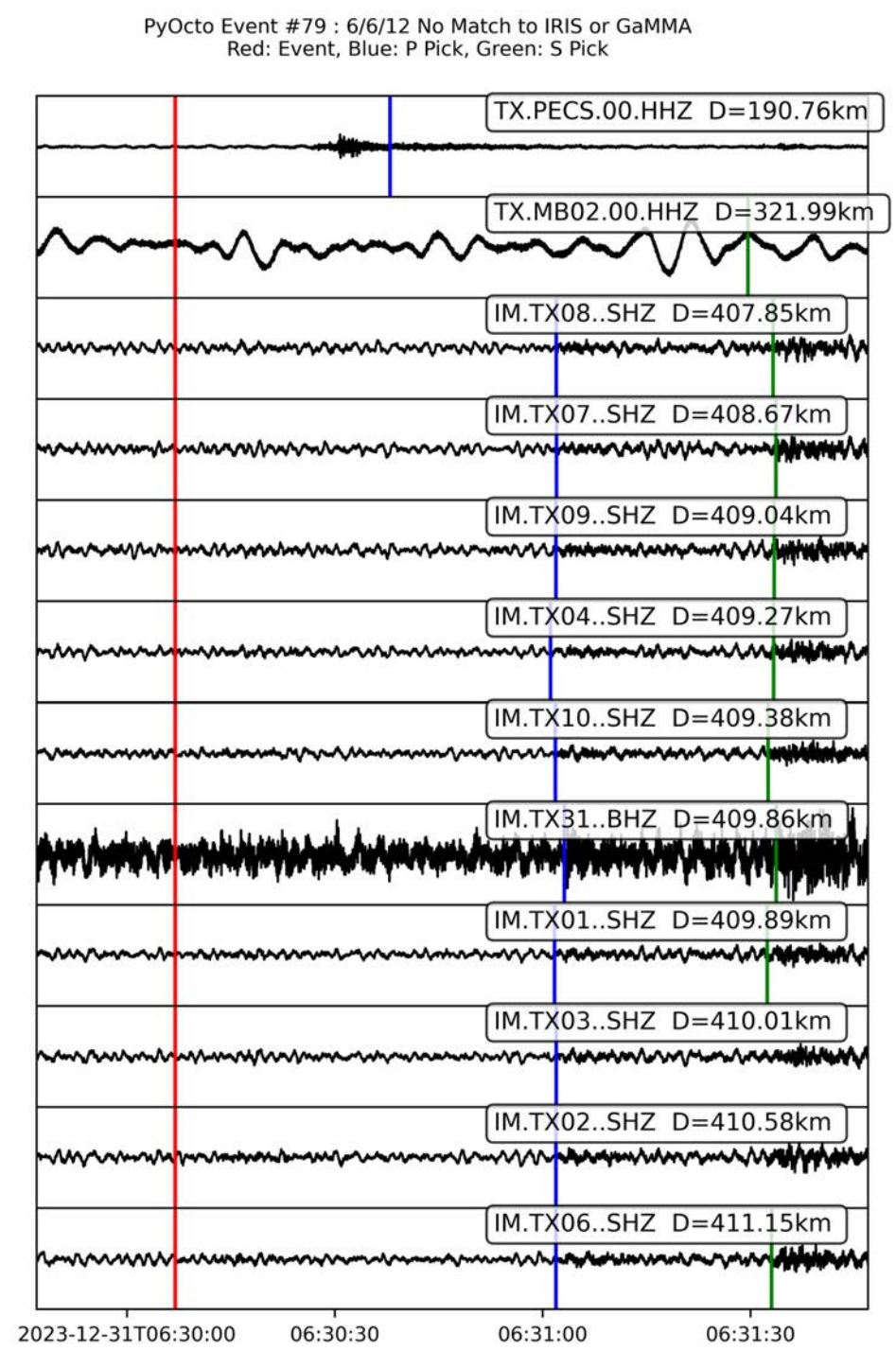
Solid: auto-picks, Dashed: predicted arrivals

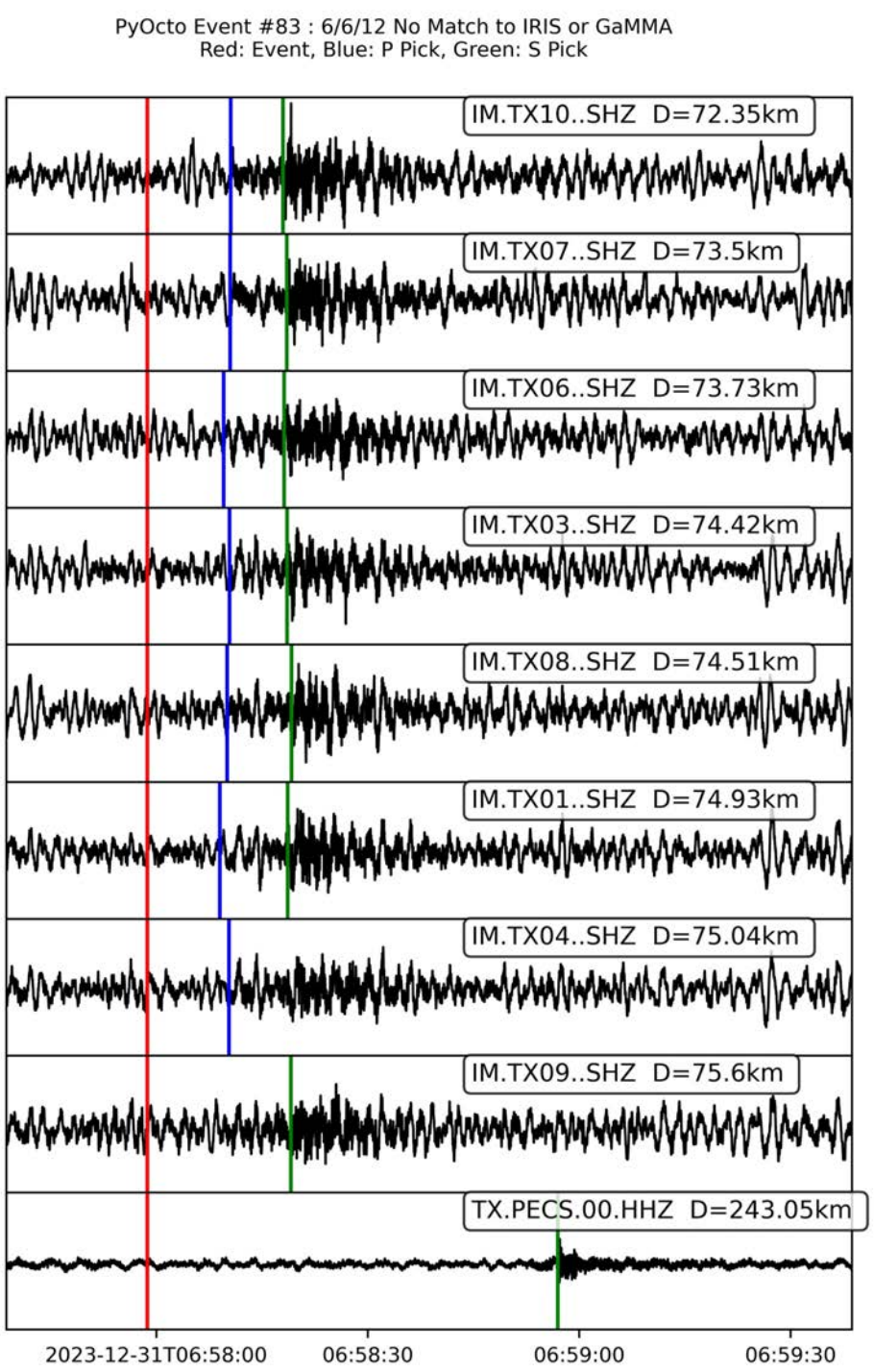


Catalog QC: False detection in coda of larger earthquake (also, too deep?)

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



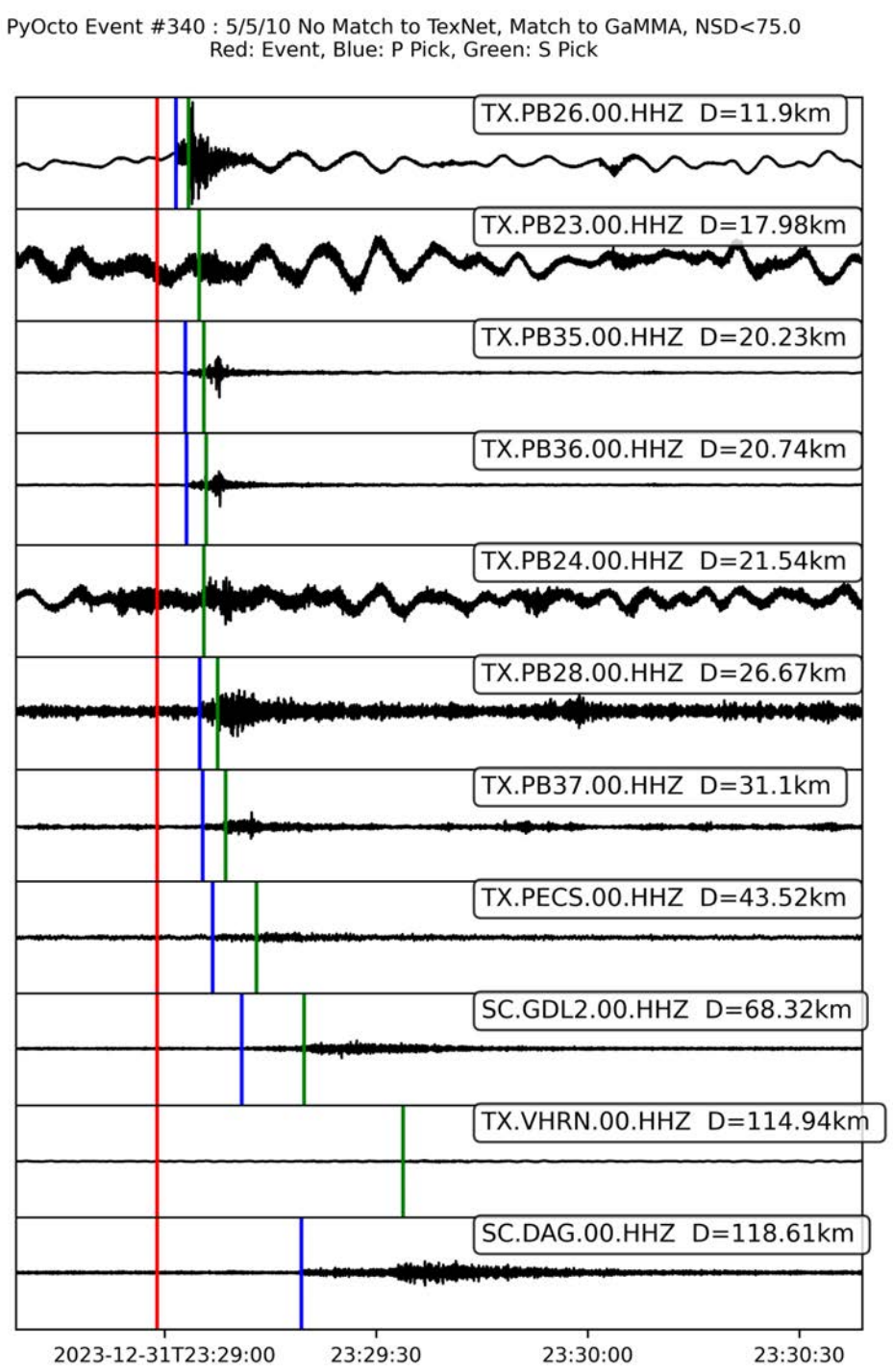


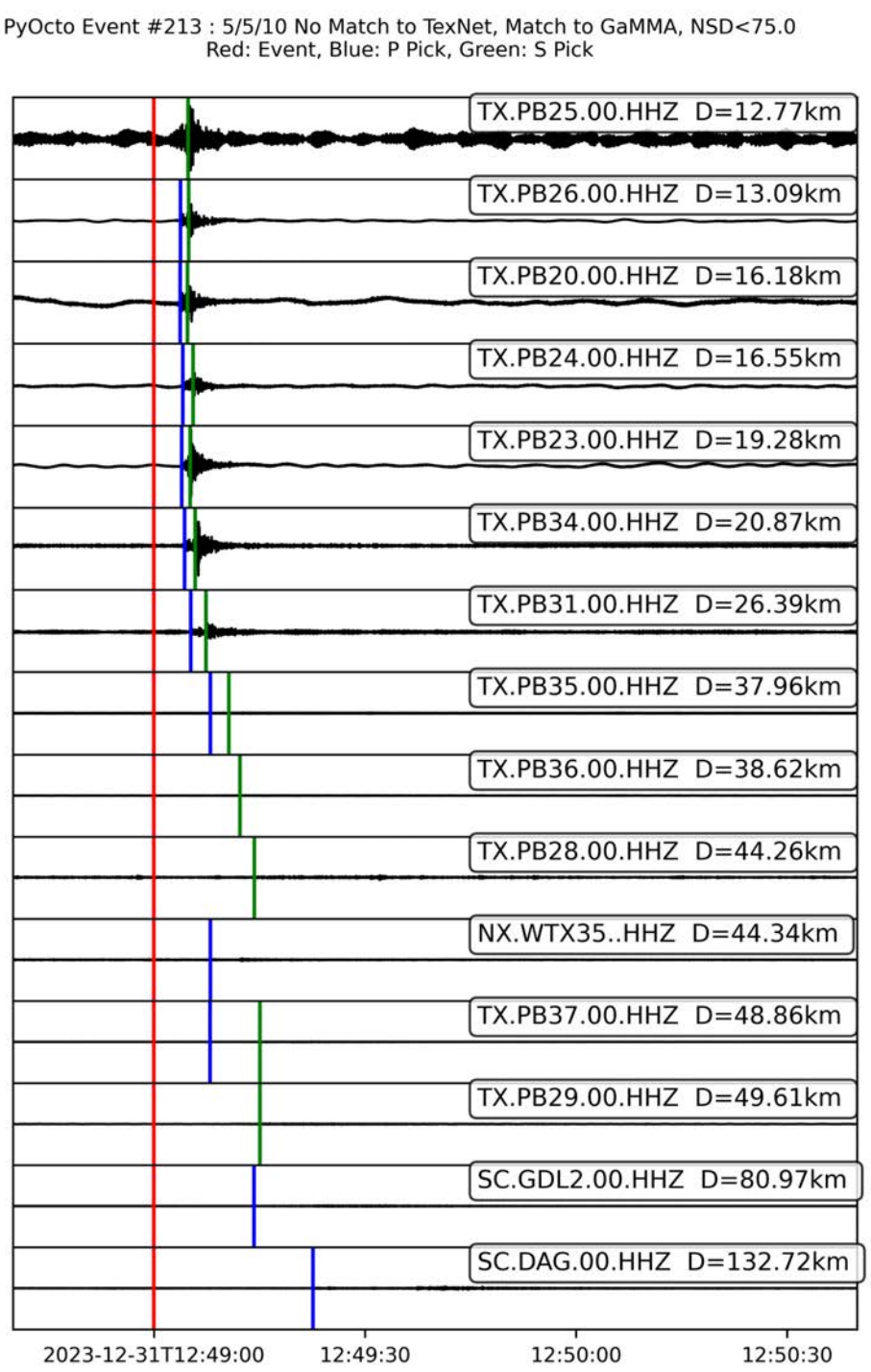


- Catalog QC: West Texas
- Bad event associations – dominated by array
- Notice large distance to nearest station

2-minute waveforms

Vertical lines:
Origin time, P, S





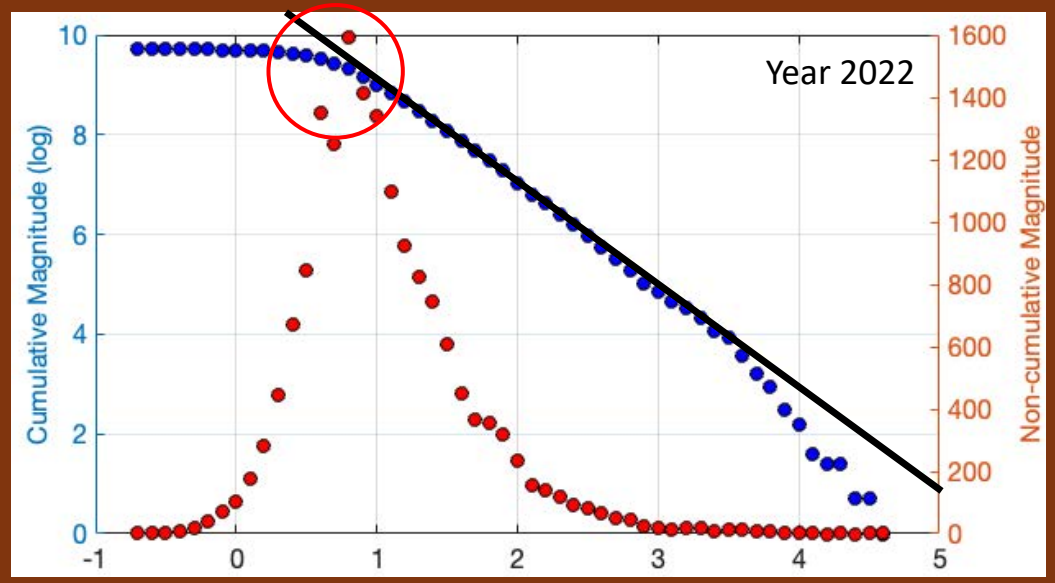
- Catalog QC: West Texas
- **Good** event association examples
- P, S picks on earthquake signals with expected moveout

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 Mc
- spatial variations of Mc across region

* More info and original R codes from CORSSA: doi: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

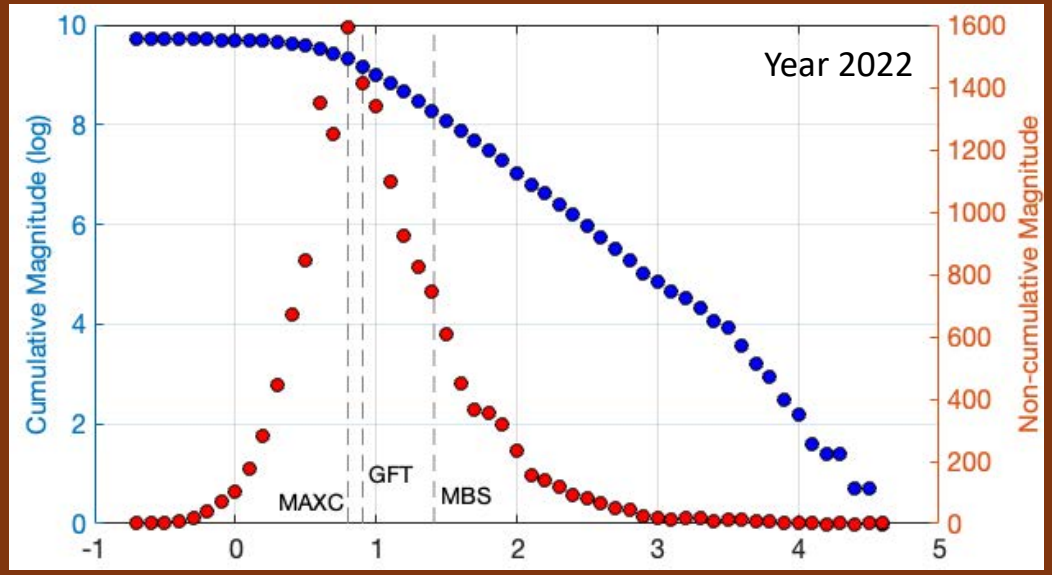
- Mc by b-value Stability: first magnitude increment where $|b_{avg} - b| <$ uncertainty of b
  * tends to produce the highest (i.e., most conservative) Mc

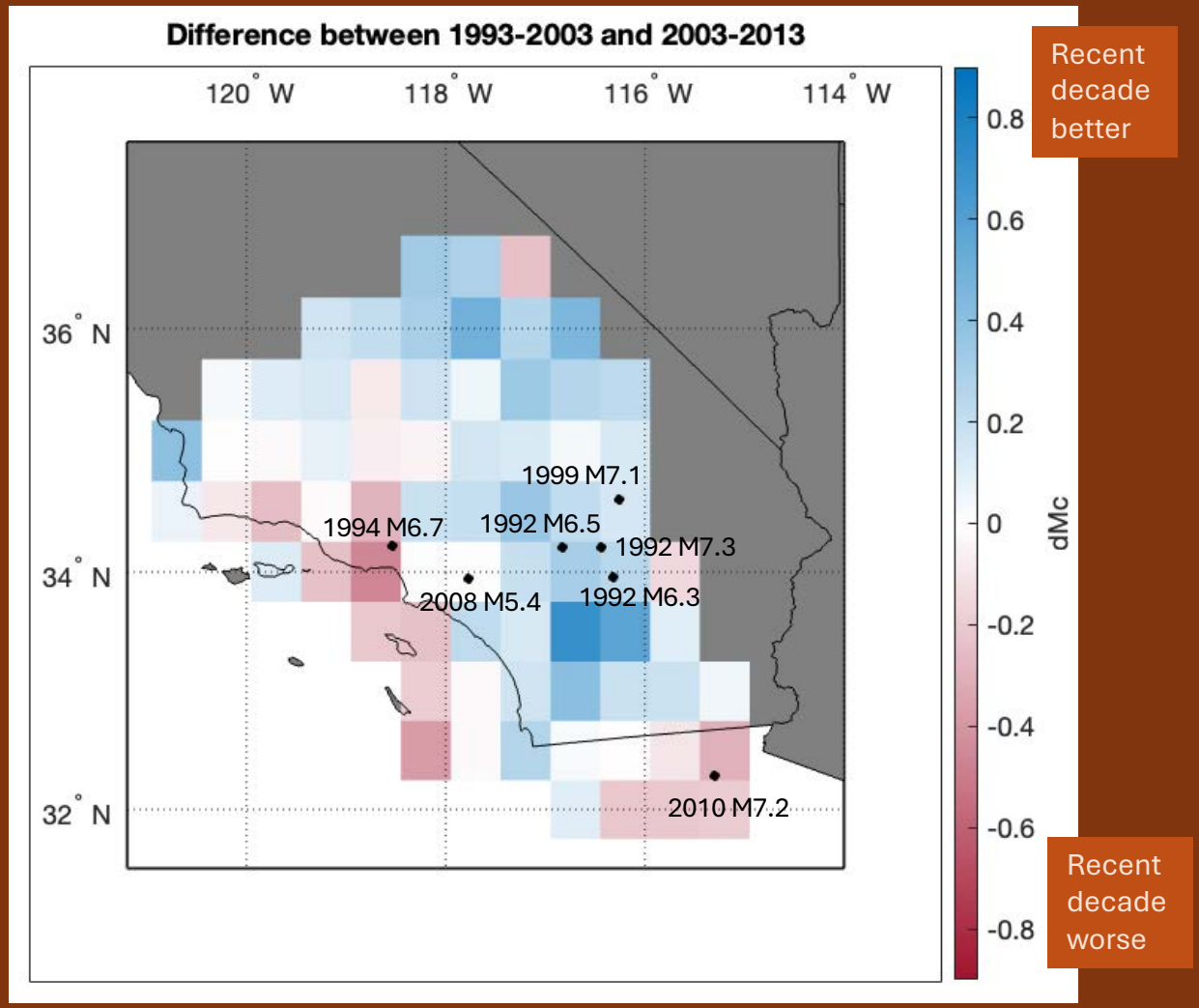
Bootstrapping

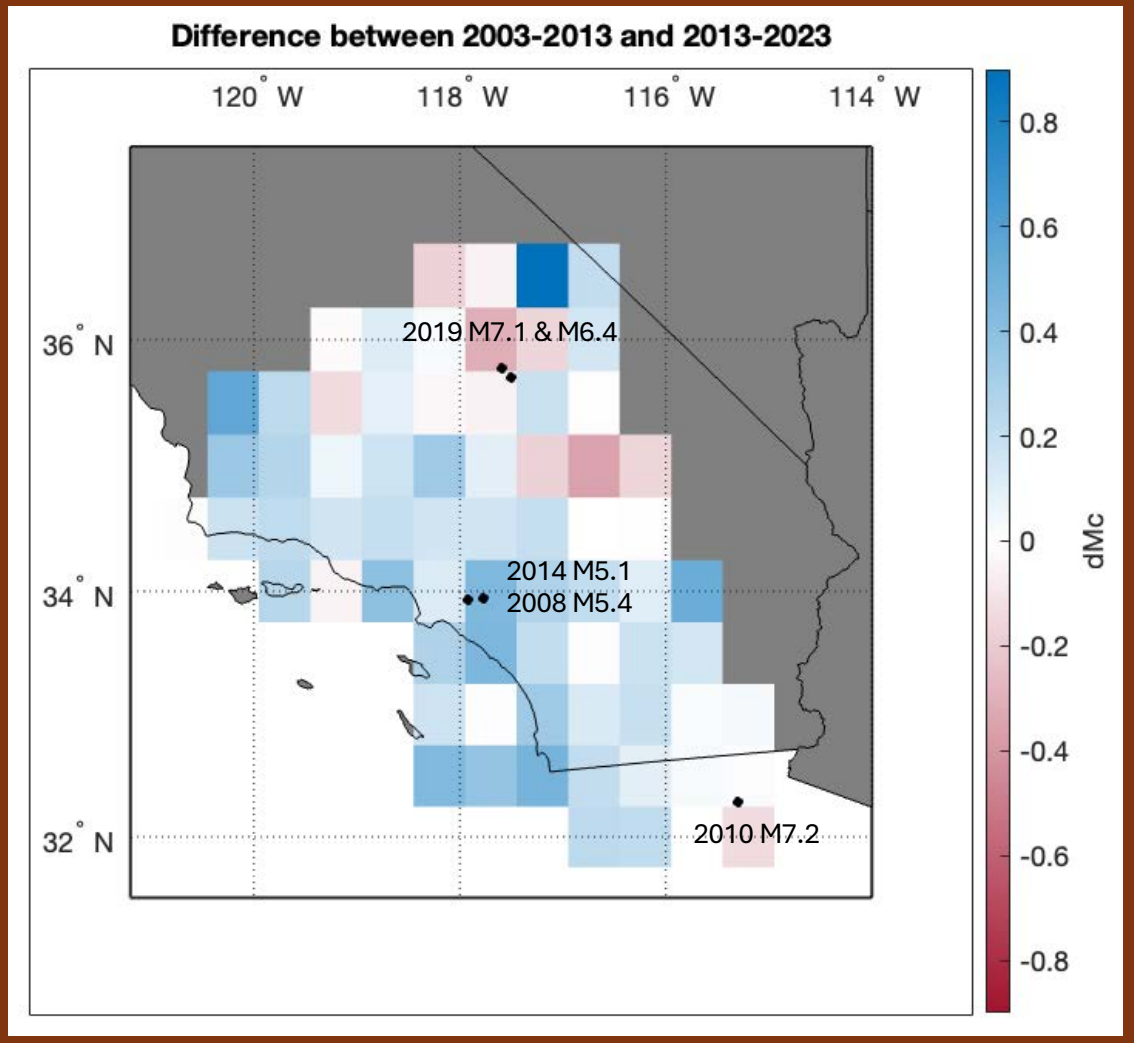
- gives sense of variation from sampling
- 200 sets with resampling
  - mean and st. dev.

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



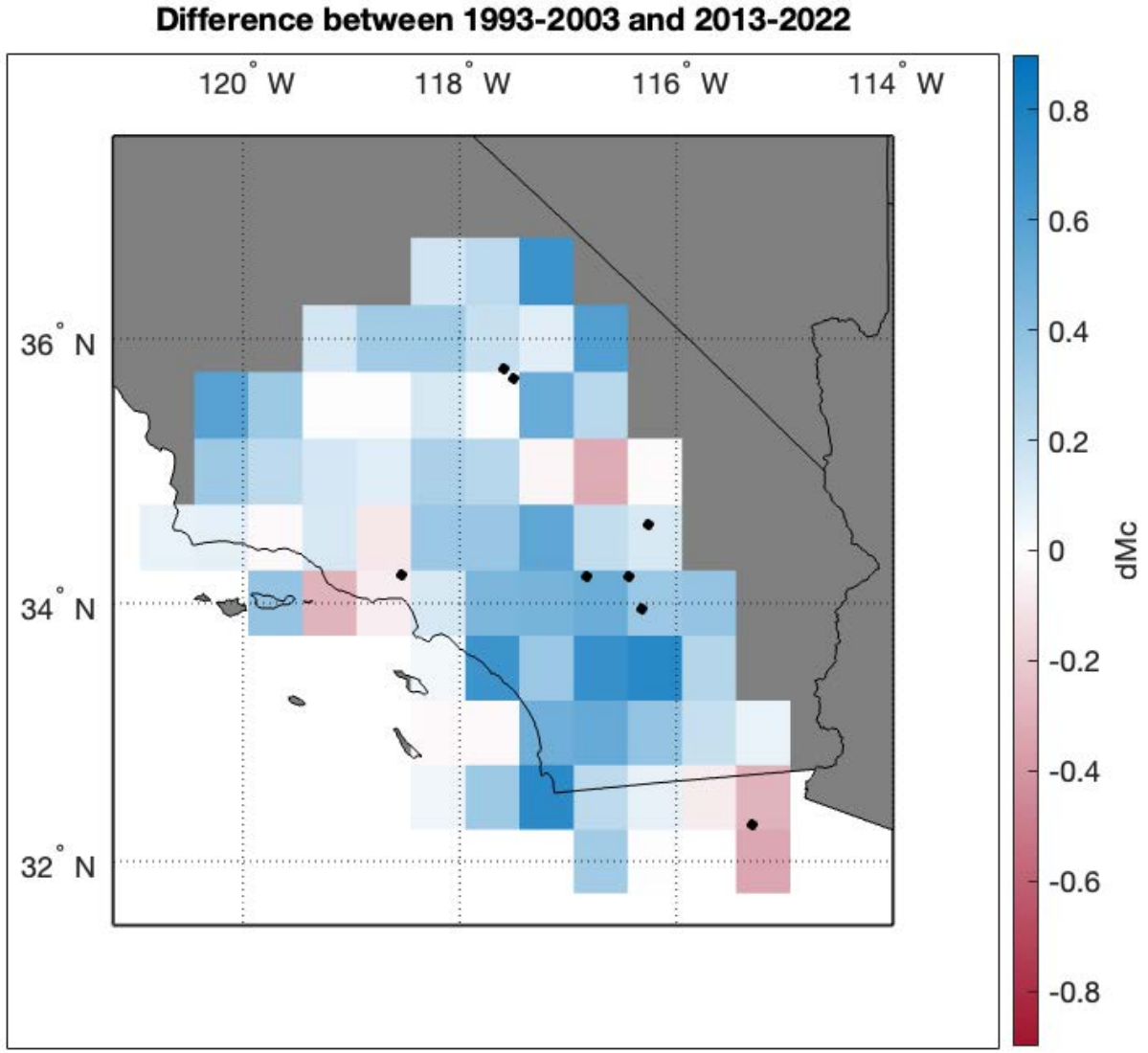


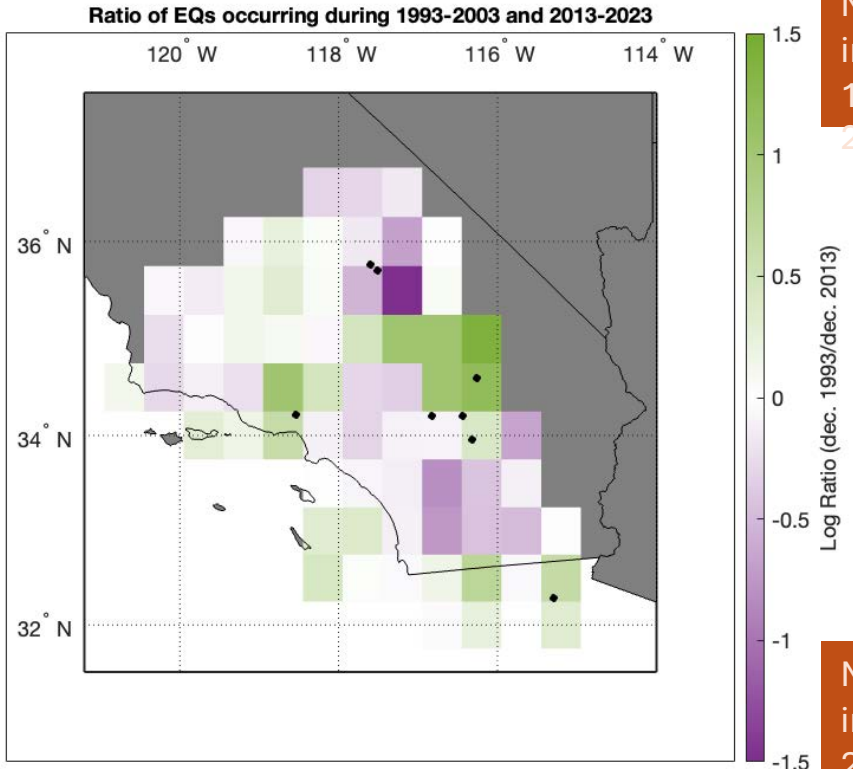


Better: 1994 Northridge

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

*incomplete





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