Round 1

Reviewer A:

For author and editor

I enjoyed reading this article and think it offers a new data product, a machine learning based pick database derived from continuous waveforms from Earthscope, NCEDC, SCEDC cloud datasets, that seismologists will be able to use in their research. Also the exploration of the data processing using cloud technology will be helpful to other researchers who want to work with large datasets.

I do have some comments/questions that I think the authors should consider before publishing. The first number refers to the line number in the article the question is based on.

Sections 2.1 and 2.2 Would be good to include some reference to a description of AWS resources. Perhaps in supplementary material?

96 - 97. I think readers would appreciate a discussion of cost to run the entire pipeline.

145. Can you offer any commentary for the differences in number of picks between the networks? Is it due to the tectonic environment (volcanic, triple junction, transform fault), or difference in station availability?

176. re: We expect an average of approx. 20 P-wave picks per M1. Is there any consideration of network coverage - would sparse regions cause underreporting of events?

182. Would you be able to provide other benchmarks to show that 25% association rate is conservative?

185. "the state wide California catalog contains 325,000 events" -- this number seems low. I just queried the SCEDC catalog (uses sta-lta and analysts for picks) for 2001-2024 (local) and retrieved over 400 K events. Are you sure we are evaluating a catalog that is equivalent in time span and completeness?

Figures:

Figure 3: It appears the vertical axis of Figure B is the jobid, but that could be any arbitrary time segment in any part of the NCEDC or SCEDC archive, or do the jobids progress chronologically through the archives? It looks like there are a number of jobids launched at the same time for SCEDC at beginning. Is that just due to the beginning of the processing when all 1500 jobs are available at the same time?

Figure 5 - Should x axis be labeled "Days after mainshock"? I assume we are progressing in time from left to right.

Figure 5 "we count all events a certain dist from reference station" - since you are referring to the number of events after mainshock - wouldn't it be better to choose the event origin as opposed to a reference station?

Figure 5 - Could you provide any discussion of the flattened part of the curve immediately after mainshock?

Reviewer B: Piero Poli

For author and editor

Dear Authors,

I found your work on global-scale picking with machine learning extremely interesting. The application is highly compelling, and the results are impressive. I have no particular issues with the manuscript; it is, in my opinion, ready for publication in *Sismica* as it stands.
I would just like to suggest one minor point: in line 106, you include a brief discussion on station channels. Since stations are also characterized by a LOC code, it might be interesting to include a few more details on this aspect.

Best regards,
Piero Poli

Response to reviewers for Seismica submission "A Global-scale Database of Seismic Phases from Cloud-based Picking at Petabyte Scale"

Dear Editor and Reviewers,

We appreciate your time in carefully reviewing our manuscript and providing constructive feedback. We have carefully made revisions addressing your comments and concerns. Along with the detailed response below, we hope that you find the revised manuscript more suitable for publication at *Seismica*.

Best wishes,
Yiyu Ni and the coauthors

Reviewer 1

Sections 2.1 and 2.2 Would be good to include some reference to a description of AWS resources. Perhaps in supplementary material?

Response: Thank you for this comment. We added the reference to a review paper in the Introduction section (Line 56). The review provides comprehensive descriptions of various AWS resources commonly used for scientific computing and including the ones used in our experiment.

96 - 97. I think readers would appreciate a discussion of cost to run the entire pipeline.

Response: Given the configurations detailed in the Workflow section, this experiment results in a total cost of ~$15,000 for us as the data consumer. We added this rough estimation in Section 3.1 (Job Statistics, Line 133).

145. Can you offer any commentary for the differences in number of picks between the networks? Is it due to the tectonic environment (volcanic, triple junction, transform fault), or difference in station availability?

Response: Yes, the differences in the number of picks may be factorized into multiple aspects.

1) The amount of data available. For example, the NC network has more than 1200 stations (according to the FDSN metadata), while UW has fewer stations (~800). Some stations may have been in operation for decades, while some may be installed just recently. Therefore, the amount of data available may vary significantly from network to network.

2) Tectonic settings where stations are installed.

3) Non-tectonic noise signals may be incorrectly picked.

4) Noise level affects the performance of the phase picker.

176. re: We expect an average of approx. 20 P-wave picks per M1. Is there any consideration of network coverage - would sparse regions cause underreporting of events?

Response: Thank you for pointing this out. The quantitative estimates used to highlight the overwhelming contribution of arrivals from small events in the database were from California. Naturally, the exact numbers will depend on the specific regional seismicity pattern and network density. However, we show that similar patterns occur in very different scenarios, such as subduction zones with sparser networks, even though at different magnitudes. We adjusted the discussion to highlight this scale-invariance and now precise that most arrivals are expected to come from events near the regional detectability threshold.

182. Would you be able to provide other benchmarks to show that 25% association rate is conservative?

Response: To estimate this value, we compared typical association ratios from studies in different regions, including dense networks with shallow seismicity, sparse networks in subduction, and ocean-bottom surveys. We now added references to these different studies to the manuscript in the Discussion section.

185. "the state wide California catalog contains 325,000 events" -- this number seems low. I just queried the SCEDC catalog (uses sta-lta and analysts for picks) for 2001-2024 (local) and retrieved over 400 K events. Are you sure we are evaluating a catalog that is equivalent in time span and completeness?

Response: thank you for pointing this out. The “325,000 event” is actually the catalog in the CEED dataset by Zhu et al. (2025). It used a filtered version of the original SCSN catalog. Here we update the text as below.

For context, the Southern California Seismic Network catalog contains approximately 450,000 events.

Figure 3: It appears the vertical axis of Figure B is the jobid, but that could be any arbitrary time segment in any part of the NCEDC or SCEDC archive, or do the jobids progress chronologically through the archives? It looks like there are a number of jobids launched at the same time for SCEDC at the beginning. Is that just due to the beginning of the processing when all 1500 jobs are available at the same time?

Response: For Figure 3b: yes, we submit jobs for NCEDC and SCEDC all at once, therefore one may not be able to tell which year of the data each job was working on. We extract this information from the submission logs.

All jobs are submitted sequentially but we use a python script to automate it so that job submission is very fast. Another important fact here is that at the beginning of the processing, the job queue was empty so all submitted jobs can go into the running state immediately. When the queue is filled, new submitted jobs are required to wait until previous jobs are finished.

Figure 5 - Should x axis be labeled "Days after mainshock"? I assume we are progressing in time from left to right.

Response: We updated the x axis label as “Days after mainshock”.

Figure 5 "we count all events a certain dist from reference station" - since you are referring to the number of events after mainshock - wouldn't it be better to choose the event origin as opposed to a reference station?

Response: We agree with the reviewer that for typical aftershock studies we'd count the number of events near the mainshock hypocenter. However, in this case we want to obtain a measure that is comparable to the number of picks which depends on the station location and is independent of the event location. Therefore, we use this station base criterion. Nonetheless, as the aftershock sequence will be the dominant source of seismicity, both selection strategies will lead to the same Omori decay behaviour, just that the station-base criterion is expected to return to a background rate earlier.

Figure 5 - Could you provide any discussion of the flattened part of the curve immediately after mainshock?

Response: The flattened part of the curve is most likely the expression of aftershock incompleteness in the pick count. In the early times after the mainshock, there may be too many signals from the mainshock coda and we only pick one event per window, therefore reducing the completeness of the picking model.

Reviewer 2 - Piero Poli

I would just like to suggest one minor point: in line 106, you include a brief discussion on station channels. Since stations are also characterized by a LOC code, it might be interesting to include a few more details on this aspect.

Response: Thank you for your comment. When stations have multiple location codes or channel types, we pick and record them all, but separately in the workflow. One can also see from the pick collection in Table 1 that a trace can be uniquely identified with a trace_id (network code, station code, location code) and a channel code.

Round 2

Reviewer A:

Thank you for responding to my comments. I have 3 very minor suggestions for consideration that do not affect the findings of the paper:

1) line 33-34: Upon reading this statement the second time, "

However, these methods are highly sensitive to background seismic noise, limiting their effectiveness to small-magnitude events or recordings from particularly quiet stations."

I agree that these methods are sensitive to background seismic noise, but disagree that they haven't been useful for large magnitude events. Current earthquake monitoring, which mostly still relies on these methods, have performed well in alerting public about felt events.

2) Line 189: should "picks associated with earthquake near the regional" be "picks associated with earthquakes near the regional..."

3) Line 195: "For context, the Southern California Seismic Network..." add "For context, in the time period of interest, the Southern California...."