

Picking Regional Seismic Phase Arrival Times with Deep Learning

A.L. Aguilar Suarez *¹, G.C. Beroza ¹

¹Department of Geophysics, Stanford University, Stanford, USA

Author contributions: *Conceptualization:* A.L. Aguilar Suarez & G.C. Beroza. *Formal Analysis:* A.L. Aguilar Suarez & G.C. Beroza. *Writing - Original draft:* A.L. Aguilar Suarez & G.C. Beroza.

Abstract Sparse instrumental coverage for much of the Earth requires working with regional seismic phases for effective seismic monitoring. Machine learning phase pickers to date have focused on local earthquake recordings. Here we present deep learning models designed and trained to be effective at picking the arrival times of earthquake phases at distances up to 20 degrees. We trained our models on the CREW dataset, which includes 1.6 million earthquake waveforms with over 3.2 million labeled arrivals on 5 minute long three component seismograms. We present models that accurately pick the first arriving P and S waves and models that pick and classify Pn, Pg, Sn, and Sg phase arrivals. We apply these models in a variety of settings and compare their performance to established machine learning models that were trained on local earthquake recordings. We demonstrate the abilities of our models by finding new earthquakes in the Gorda plate offshore northern California. Finally, we use our multiple phase picker to find new examples with secondary arrivals from our massive training dataset. The goal of this method is to improve automatic earthquake monitoring in regions of sparse instrumental coverage and seismicity in remote regions far from instrumentation.

Production Editor:
Andrea Llenos
Handling Editor:
Hongyu Sun
Copy & Layout Editor:
Oliver Lamb

Received:
August 24, 2024
Accepted:
March 17, 2025
Published:
April 30, 2025

1 Introduction

Many observational seismology tasks have been enhanced using machine learning. Deep learning models have proven effective in phase picking (Zhu and Beroza, 2019; Mousavi et al., 2020), phase association (McBrearty and Beroza, 2023), earthquake location (Smith et al., 2022), magnitude estimation (Mousavi and Beroza, 2020), polarity assessment (Ross et al., 2018; Uchide, 2020), and signal denoising (Zhu et al., 2019; Novoselov et al., 2022) and there are easy to use platforms to support their adoption like Seisbench and Qseek (Woollam et al., 2022; Isken et al., 2025). Improved earthquake catalogs are enabling insights into the dynamics of seismicity in a variety of settings, including crustal seismicity (Tan et al., 2021), deep earthquakes (Xi et al., 2024); induced seismicity (Park et al., 2020; Schultz et al., 2023) and volcanic seismicity (Wilding et al., 2023). A common factor of these earthquake catalogs is that they were built in regions with dense seismic monitoring and close station spacing.

The data and model landscape of deep learning based phase pickers is dominated by local recordings, i.e. seismograms from small earthquakes at short distances. Two of the most popular pickers were trained on such data. PhaseNet (Zhu and Beroza, 2019) was trained on data at distances less than 140 km. EQTransformer (Mousavi et al., 2020) was trained on the Stanford Earthquake Dataset (Mousavi et al., 2019) that contains data up to distances of 350 km, but is dominated by data at shorter distances, with only 8 % of the data covering the

110-350 km range.

Here we introduce models that are developed for machine learning powered earthquake catalog creation in regions with sparse instrumentation or regions where seismicity occurs far from seismic instruments, such as subduction trenches that typically lie more than 100 km away from the closest land instrument. The datasets that include regional distances and are global in coverage are the MLAAPDE (Cole et al., 2023), NEIC (Yeck et al., 2020), GEOFON (Woollam et al., 2022), Curated Regional Earthquake Waveforms (CREW) (Aguilar Suarez and Beroza, 2024), and the Seis-PnSn (Kong et al., 2024) datasets. Among these, the CREW dataset is dedicated to, and emphasizes, regional distance data. It is the only one for which every waveform contains both P and S labeled arrivals. In addition, several examples contain labeled secondary arrivals, e.g. including both Pn and Pg picks. We trained our models on the CREW dataset (Aguilar Suarez and Beroza, 2024).

With multiple effective machine learning models for first arrival picking available, attention has started to shift to secondary phases. Secondary phases can play an essential role in constraining earthquake depths (Münchmeyer et al., 2023; Wang and Klemperer, 2021), provide insights about crustal and upper mantle structure (Al-Damegh et al., 2004; Pasyanos et al., 2009), and are essential for nuclear test-ban-treaty verification (Rodgers and Walter, 2002). In this paper we also present models with the ability to simultaneously pick and classify crustal (Pg, Sg) and mantle arrivals (Pn, Sn).

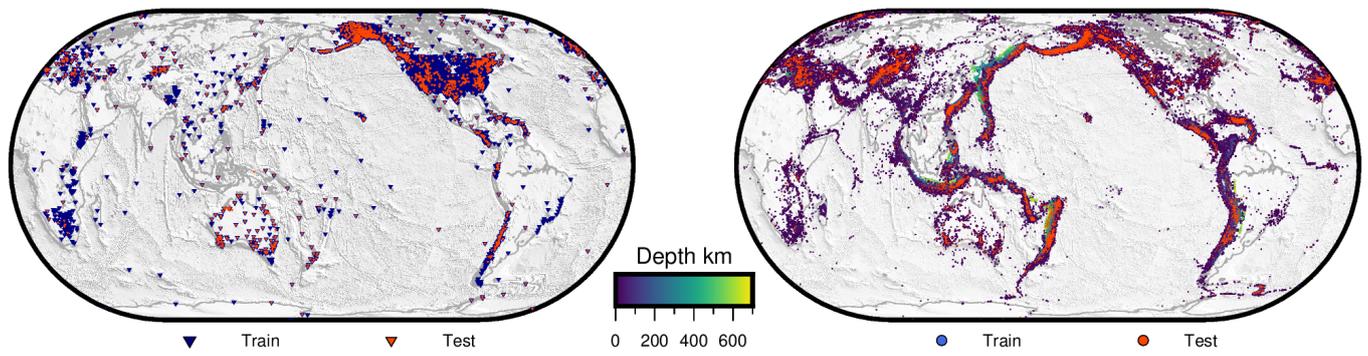


Figure 1 Left. Distribution of stations in the CREW dataset, where the training stations are shown as navy triangles and the testing stations as orange triangles. Right. Earthquakes in the CREW dataset, events in the training split are color coded by depth with the scale shown, and events in the test split are shown as orange circles.

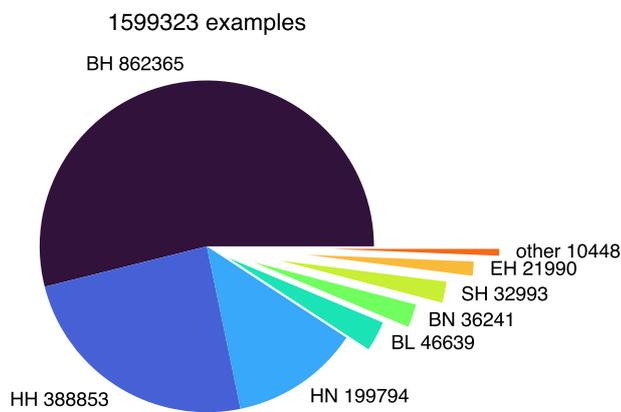


Figure 2 Instrument types and number of examples in the CREW dataset. (BH) broadband high gain seismometer, (HH) high broadband high gain seismometer, (HN) high broadband accelerometer, (BL) broad band low gain seismometer, (BN) broadband accelerometer, (SH) short period high gain seismometer, (EH) extremely short period high gain seismometer.

2 Data

The CREW dataset consists of 5-minute three-component waveforms from earthquakes recorded at distances between 1 and 20 degrees of source-receiver separation. In total there are 1,599,323 examples in the CREW dataset, from 523,294 earthquakes recorded at 4,071 locations around the globe (Figure 1). These waveforms come from a variety of instruments, including broadband seismometers, short period seismometers and accelerometers (Figure 2). All of the waveforms are resampled to 100 Hz. Furthermore, each example contains labeled arrivals for P and S waves, adding up to over 3.2 million labeled arrivals. The waveforms have been normalized by removing the mean and dividing by the maximum amplitude among the three components but are otherwise in their raw form, with no filtering or correction of the instrument response.

We split the CREW dataset into training and testing sets. The training set includes 1,536,731 examples, each of which is a three component seismogram with both

P and S labeled arrivals. These examples come from 504,552 earthquakes that span all latitudes, all longitudes, and all depths recorded at stations around the world. Figure 1 shows the locations of the stations in the training set, as well as the location of the earthquakes. The test set includes 62,592 examples, from 18,742 earthquakes, which is the most recently recorded part of the dataset, spanning part of 2021 and 2022. Figure 1 indicates which stations are in the train and test splits. There are 49 out of the total 4,071 stations in the CREW dataset that are exclusively represented in the test split.

3 Picking First Arrivals

We present **SKYNET**, **S**eismological **K**nowledge **Y**ardstick **N**ETworks. We trained an architecture similar to PhaseNet to pick the first arriving P and S waves, but with five additional layers. The original PhaseNet architecture is made of 19 convolution layers, whereas our architecture comprises 24 layers of convolution. Both models employ a U Net architecture with skip connections, the original with four downsampling and four upsampling stages, whereas ours has five downsampling and five upsampling stages. The original PhaseNet was trained on 30 second waveforms, but our training data is 300 second waveforms, which is ten times the duration used in the original PhaseNet. By adding these layers, we ensure a similar level of feature representation in the encoding. In PhaseNet, the number of features at the bottom of the U is 22x12, in ours it is 32x30, which is on the same order of magnitude, even though the waveforms are a factor of 10 longer. With these modifications we produced a longer receptive field, to ensure proper processing of the much longer seismograms. The size of the convolution kernel and the stride step were set to seven and four sample points, as in the original PhaseNet. The number of trainable parameters in our architecture is 79,442, compared to 38,734 in the original PhaseNet. This model takes as input three component waveforms and outputs three time series, for the probability of samples being the arrivals of P waves, S waves, or something else, likely noise.

We trained our phase picker on an augmented ver-

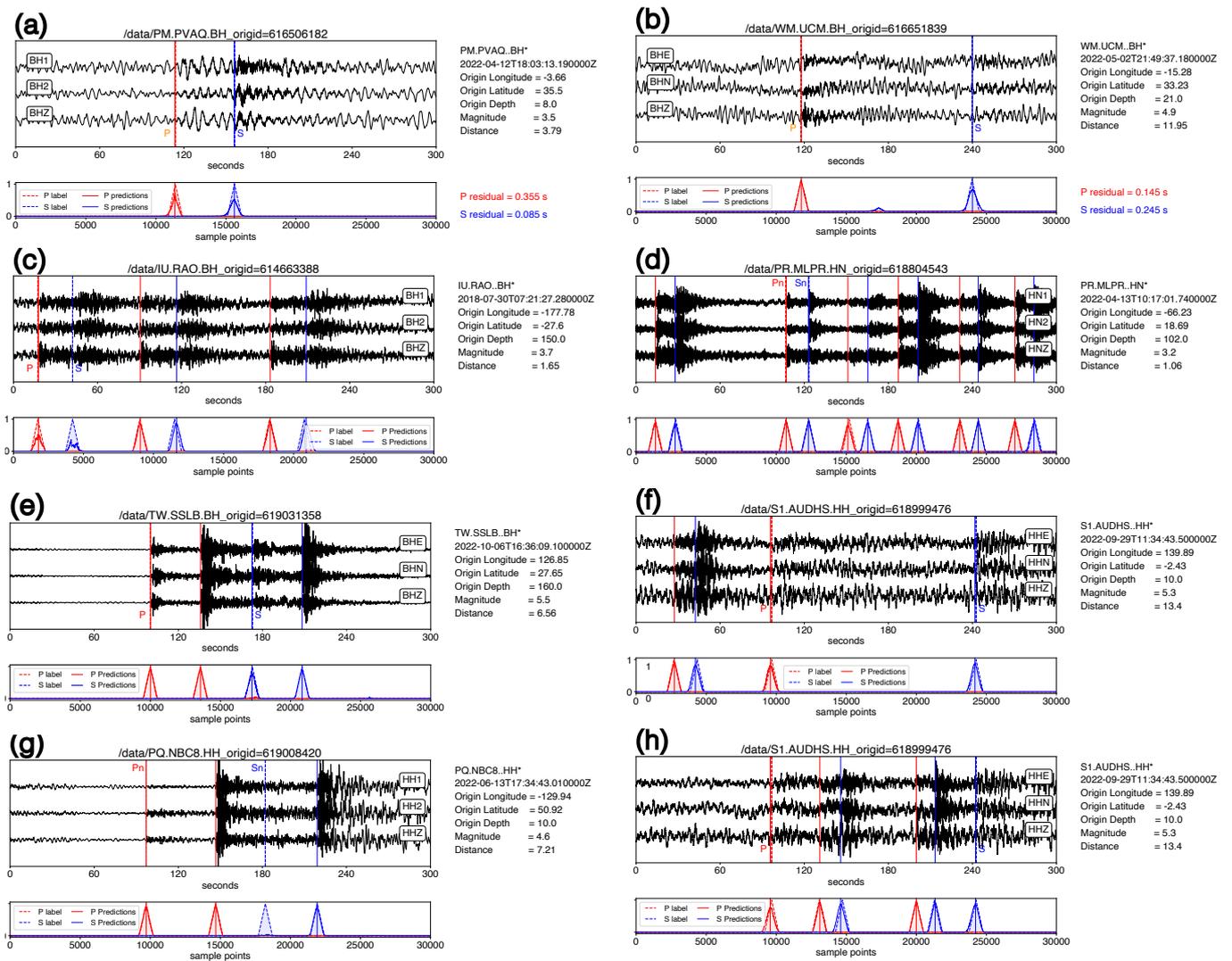


Figure 3 Picker performance in test data and augmented test data. Dataset labeled arrival times and triangle labels are shown with dotted lines, whereas model predictions and inferred pick times are shown with solid lines. (a-b) Raw examples from the test set. Residual times between labeled and predicted times of P and S waves are indicated on the bottom right. (c) augmented example by concatenating the same waveforms three times with a random polarity flip. (d) Six different examples merged together. (e) entangled waveforms of two earthquakes for which the P arrival of the second one arrives between the P and S arrival of the first one. (g) Same modality as (e), but the model misses the first S arrival. (f) Mixture of a local earthquake signal with one from 13.4 degrees source-receiver distance. (h) A regional earthquake with two closer earthquakes in between its P and S arrivals. All but two picks in these examples were recovered.

sion of the CREW dataset. Figure 3 illustrates various recipes used for data augmentation, which are mostly superpositions of two or more examples with random polarity flips, amplitude modulation and random shifting. First, If the S-P time, a proxy of the signal duration, was more than a third of the total duration (100 s or 10,000 samples), the data and labels were kept in their original form, as in (Figure 3(a,b)). If the duration of the signal was shorter than 10,000 samples, we imposed a random choice between 2 or 3 copies of the same example. In the case of 2 or 3 copies, we also introduced a random chance of flipping the polarity of the waveforms (Figure 3(c)). In every case, the training labels were adjusted accordingly as displayed in the bottom of each panel (Figure 3(c-h)), where multiple triangular shapes indicate the multiple arrival times.

To increase the complexity of the training data, we

combined different examples following a variety of formulas. First, for the shortest duration signals we added several earthquake signals consecutively, with a random time shift, a random polarity shift, and a random amplitude amplification (Figure 3(d)). For some examples, the waveforms were superimposed such that the P arrival of the second earthquake arrives between the P and S arrival of the first earthquake (3(e,g)). We also combined long duration signals with short duration signals (Figure 3(f,h)). The long and short duration signal mixtures include cases where the signals do not overlap (Figure 3(f)), and where the signals do overlap, such that the short duration signals are placed in between the P and the S arrival of the long duration signal (Figure 3(h)).

We used a synthetic Gaussian noise generator to create three-component noise samples of 300 s duration.

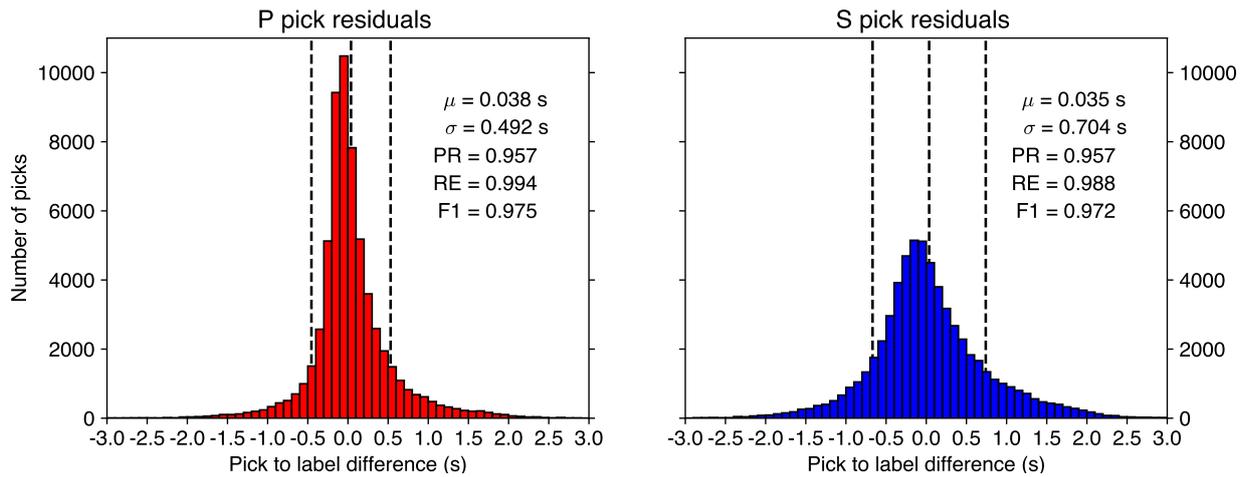


Figure 4 Evaluation metrics for picking P and S waves in the test set. PR: precision, RE: recall and F1 is the F1 score.

These noise waveforms contain a random number of spikes and shifts (Figure S1). In each batch of 32 samples, three quarters are a mix of raw and augmented data according to the recipes in Figure 3. The remaining quarter is synthetic noise. We used the Adam optimizer with a learning rate of 0.001 and trained for 225 epochs.

4 Model Selection

We trained three different models with different triangular label widths. We tried triangle base lengths of 1000 samples, 400 samples and 200 samples, which correspond to 10 s, 4 s and 2 s respectively. Of these three models, the one that performed the best in terms of consistently producing the higher peak classification probabilities was the model with the wider labels. That model has a label width of 1000 samples, which translates to a duration of 10 seconds. Figure S2 shows the distribution of detection values for the test set for the models trained on labels with different widths. We selected the model with the wider labels as it produces the best predictions with shapes that resemble the labels more closely, with the highest heights of the triangular shapes. This makes the model results less sensitive to a threshold choice. Figure 4 shows the distribution of residuals for P and S picks in the test set for the chosen model. The mean residual for both cases is below 40 milliseconds, which translates to less than 4 sample points. However, the S pick distribution is wider with a higher standard deviation. Both distributions resemble a double exponential (Laplace) distribution. Also, both distributions show some skew, with heavier tails on the positive side (picking later). The precision, recall and F1 scores are all greater than 0.95. For these metrics, we considered a true positive (TP) an arrival that is recovered within 3 seconds of a labeled arrival (1% of the duration of the waveforms), and false positives (FP) arrivals picked outside of this window. The accuracy of the picks seems independent of source to receiver distance and signal to noise ratio, for both P and S waves, as shown in Figure S3. A more relevant test for our model is to deploy it to find new earthquakes in data outside of the train/test splits.

5 Application to new and continuous data

One of the best applications for our models is for regions with sparse seismic networks and/or remote seismicity. The models introduced here are not intended to replace or update established models that are trained on local data, but to complement them at greater distances. Figure 5 shows the predictions from SKYNET alongside the predictions from PhaseNet for stations in Northern California near the Mendocino Triple Junction and the Gorda Ridge. The earthquakes indicated in Figure 5 are nearly 300 km away from the closest seismometer, which is well beyond the distance cutoff for the PhaseNet training data. If we counted the number of picks that PhaseNet produces here over the 0.5 threshold, we find 3 P picks and zero S picks within the expected window. We note that our model takes as input 300 seconds, processing all the waveforms shown in a single step, whereas PhaseNet takes inputs of 30 s duration, requiring sliding windows to process the waveforms. SKYNET picked 10 P, and 8 S wave arrival times compared to 3 P and 0 S from PhaseNet. We also note that the height of the predictions from our models seem very consistent through the different stations. The waveforms displayed in Figure 5 are bandpass filtered between 1 and 10 Hz to facilitate the visualization of the earthquake. Figure S4 shows the raw waveforms, which were the ones used as input to the model.

We applied the method to continuous data from northern California in the vicinity of the Gorda ridge. We run model predictions for the day of February 29, 2024, when the swarm of earthquakes shown in Figure 5 took place. We used GENIE (McBrearty and Beroza, 2023) for phase association of the picks. Our association model was trained for earthquakes outside the footprint of the seismic networks in northern California. During the day of the swarm, we detected 34 earthquakes, recovering all 10 of those reported by the USGS, plus 24 newly detected ones. We estimated local magnitudes for all earthquakes and merged USGS events with ours if the difference in origin time was less than 10 seconds, keeping the USGS magnitude. Figure 6 shows the time

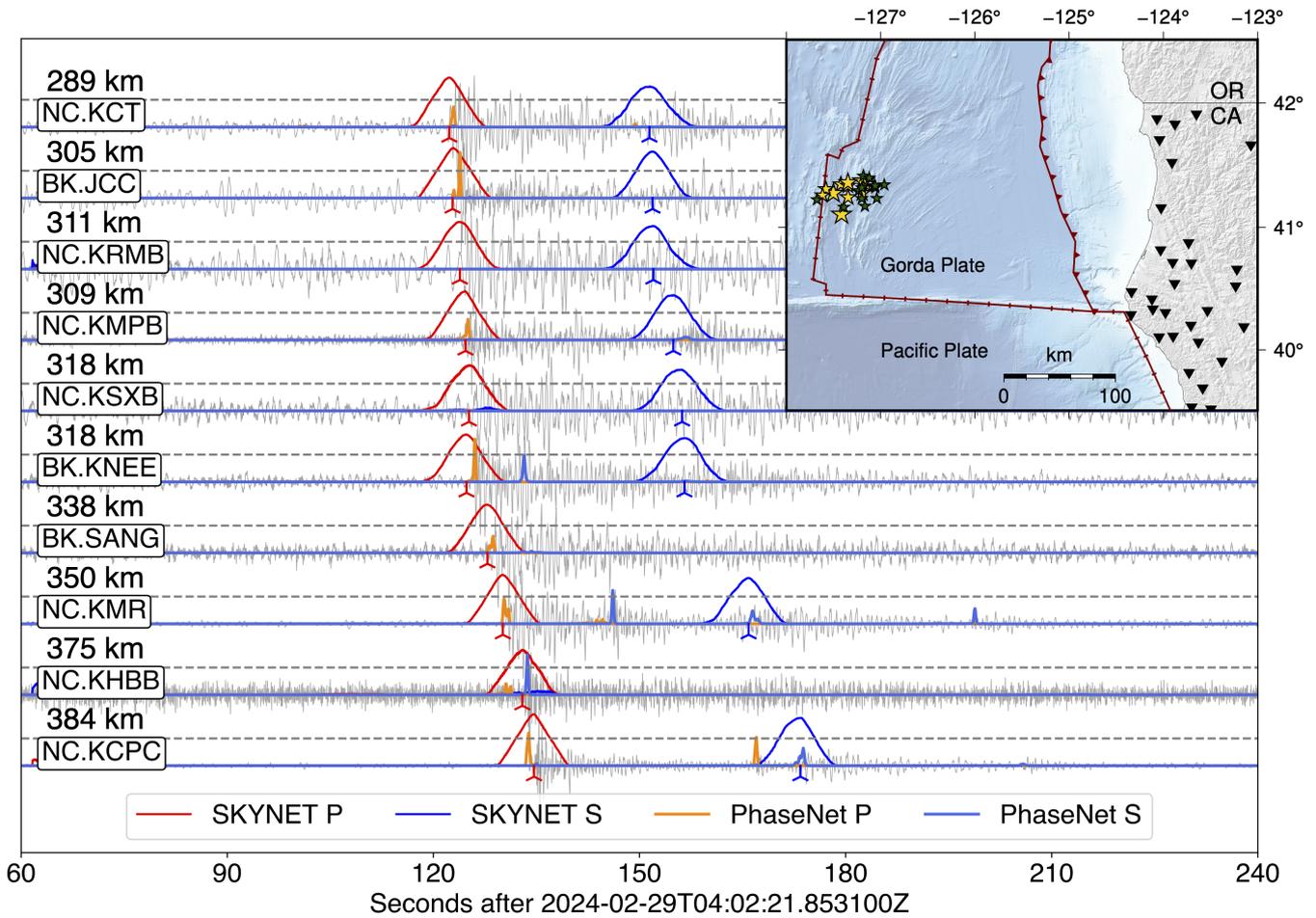


Figure 5 SKYNET and PhaseNet predictions and picks from an earthquake near the Gorda Ridge offshore Northern California. Red and blue solid lines are P and S predictions from SKYNET. Red and blue markers indicate the picked times. Dark orange and pale blue display PhaseNet predictions for P and S. Grey dotted lines indicate the 0.5 threshold for predictions. Waveforms are the HHZ channels of instruments of the BK and NC networks in Northern California. Waveforms are bandpass filtered for display, but predictions are performed on unfiltered data (see Figure S4 for raw waveforms). Station locations (triangles) and event locations (stars) for the swarm on 2024-02-09 are shown in the inset. Yellow stars show the USGS events and green stars the newly found earthquakes using our workflow, these correspond to the green markers in Figure 6. Red lines are plate boundaries.

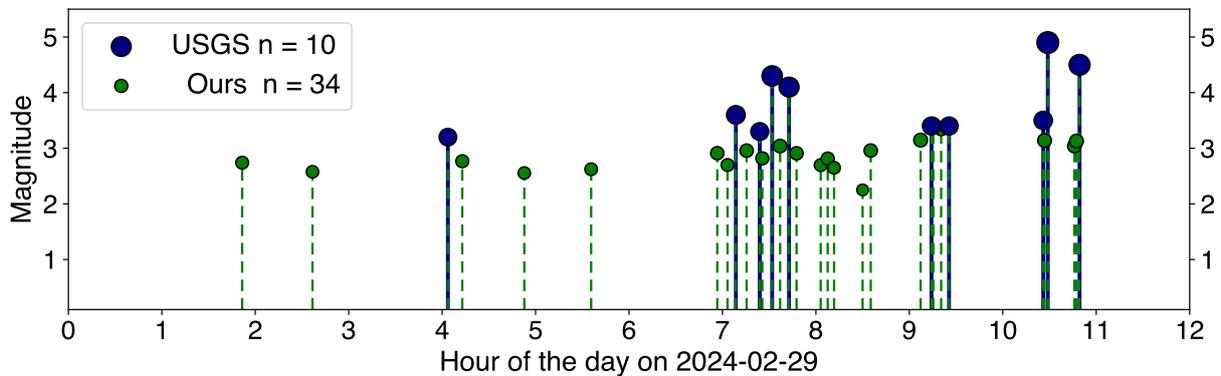


Figure 6 Our earthquake catalog compared to the USGS catalog. We detected 34 earthquakes during the day of February 29, 2024. The first 12 hours are shown, where all 10 USGS detections are, along with 32 out of the 34 in our catalog. Local magnitudes were estimated, and for those events in both the USGS and our catalog, we kept the USGS magnitudes. Green dotted lines on top of navy blue lines indicate that the events are in both our catalog and the USGS. Our new locations are displayed as green stars in the inset map of Figure 5, and the USGS locations are the yellow stars.

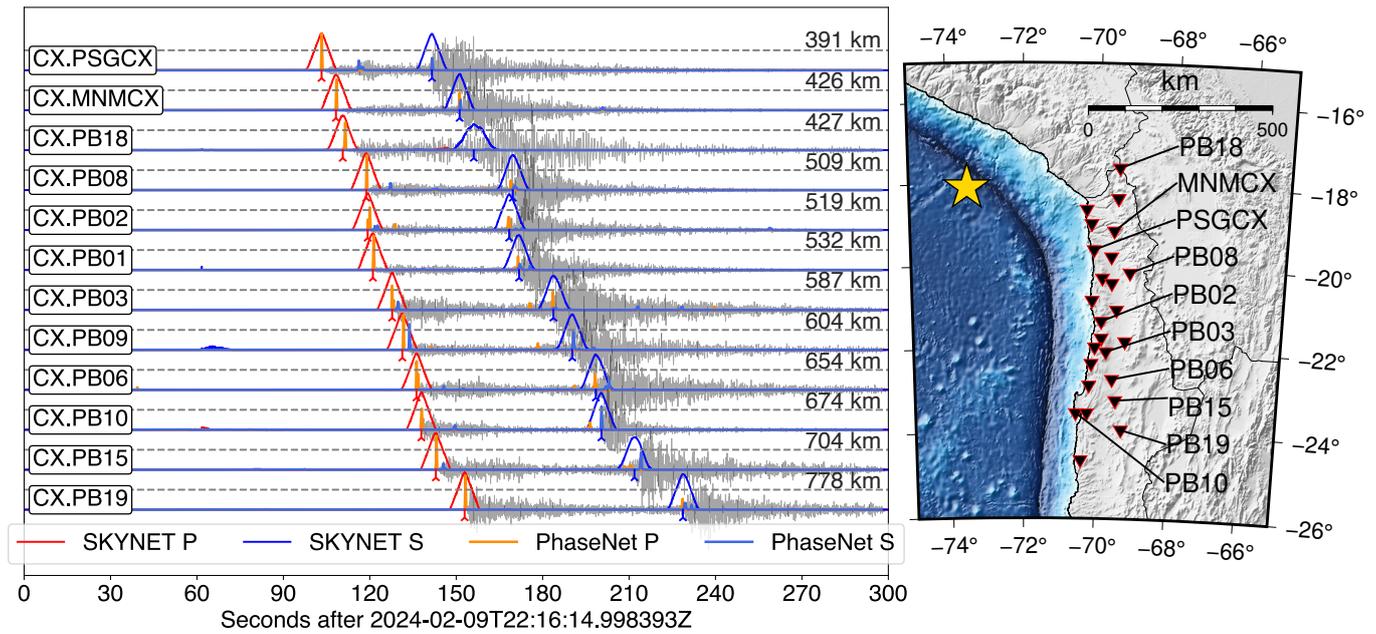


Figure 7 SKYNET and PhaseNet predictions for an event on the outer rise of the Nazca plate subducting under Northern Chile. HHZ components are displayed. Red and blue markers indicate picked times for P and S waves, respectively.

and magnitudes of the events in our catalog and the USGS catalog. The 34 events occur throughout the day, with 32 out of the 34 in the first half of the day, where the USGS events are reported. The smallest event in the USGS catalog is magnitude 3.2 and the smallest in ours is magnitude 2.3.

Another important setting for our applications is subduction zones, where the trench is often more than 100 km away from the coastline, so that seismicity near the trench or beyond the trench on the outer rise will only be recorded at distances over 1 degree by stations on land. Figure 7 shows predictions from PhaseNet and SKYNET in the CX network for an earthquake on the outer rise recorded in northern Chile and southern Peru. The closest station to the epicenter is approximately 391 km away. PhaseNet recovers P waves at all stations and produces S class activations a few seconds after the P picks, as seen in stations PB03 and PB09. SKYNET also produces P wave picks for all stations and S wave picks for all stations, in contrast to none for PhaseNet, which has higher activations for the P wave class at the time of arrival of S waves, in stations like PB06 and PB03. As noted above, SKYNET processes the 300 s in a single window, whereas PhaseNet needs several sliding windows to process the whole waveform. Intuitively, SKYNET sees the whole waveform in a window, whereas PhaseNet does not have such a broad receptive field to help make its predictions.

We created the regional picker to strengthen monitoring capabilities in places that have sparse instrumentation or that record seismicity from remote regions. We intend it to be a complement to the models that are available and widely used for densely instrumented regions. Figure S5 shows the model predictions from both SKYNET and PhaseNet on an earthquake near Lone Pine in eastern California. For the stations that are closest to the hypocenter, PhaseNet performs the

best, since it was trained on seismograms within one degree of source to receiver distance. As distance increases, the performance of SKYNET becomes better than PhaseNet, since it was trained specifically on data recorded at distances greater than one degree of source-receiver separation.

Figure 8 shows the performance of the picker on the seismograms of a recent very deep earthquake at 607 km depth under western Brazil as recorded by the CX network in northern Chile, as in Figure 7. In this case the station closest to the epicenter is over 1100 km away. P arrivals shown in red seem accurate, as do most S picks, but some of the later S arrivals seem to mark the onset of high frequency waves, rather than the low frequency and very large amplitude initial S arrival. This might be a consequence of the limited number of labeled waveforms of very deep earthquakes in the training data.

We trained our model on long waveforms, with data that contain long S minus P times. We engineered the model to have a long receptive field, which compared to PhaseNet comprises twice the number of trainable parameters. This is while also slightly reducing the computation time compared to PhaseNet (Figure S6).

6 Simultaneous picking of crustal and mantle phases

We collected the examples in the CREW dataset for which the four phases Pn, Pg, Sn, and Sg have labeled arrivals, leading to an initial dataset of 6705 examples, which makes up only 0.4 % of the whole CREW dataset. Figure 9(a) shows one of these examples. These type of waveforms where the four arrivals are distinguishable from each other are all in the distance range from 2 to 7 degrees source-receiver separation, with a peak at 3

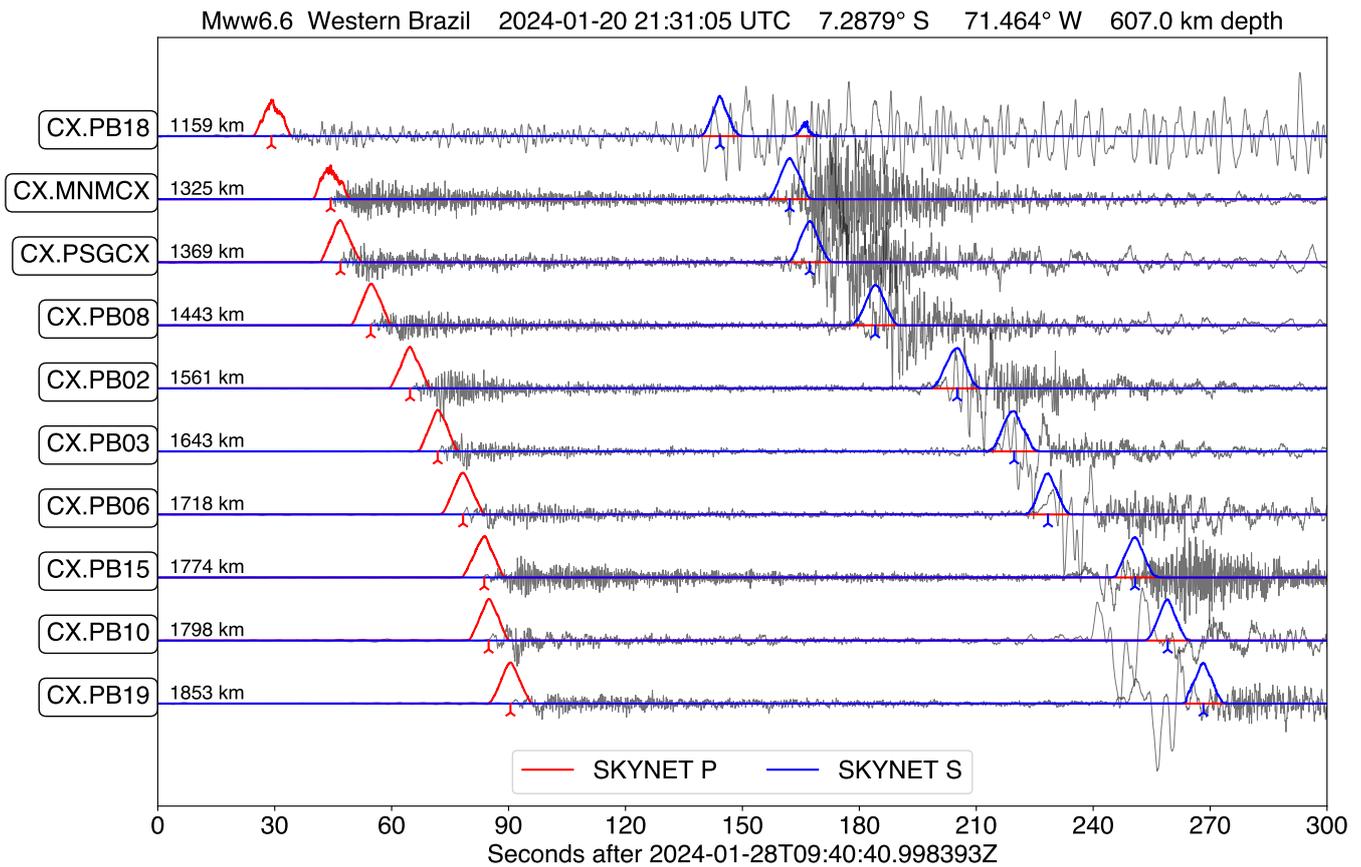


Figure 8 Very deep earthquake (607 km depth) under western Brazil recorded by the CX network in northern Chile. Stations are sorted by source-receiver distance. HHZ components are displayed. Distance from the epicenter to each station is indicated next to the station names. Model predictions are shown with solid red and blue lines, and the tripod markers indicate the derived arrival times.

degrees. In this subset of the CREW dataset, all selected examples have Pn and Sn as the first arrivals, and Pg and Sg as a secondary arrivals. We trained a picker on this small dataset, again with a U-Net architecture with skip connections, that takes as input the three-component waveforms and outputs 5 time series, one for each of the phases Pn, Pg, Sn, Sg and noise. It has more output channels than the regional picker, but is otherwise identical, with 79,442 trainable parameters, compared to 79,328 from the models presented above. We trained during 12,000 stages of samples of 200 randomly chosen examples. This model picks each one of the phases with an average residual of 0.2 seconds, Figure 10 shows the distribution of residuals for the training dataset. Compared to the residuals in Figure 4, the distributions are wider, with larger standard deviations.

Once we had a model that picks the four phases as in Figure 9(a) where both labels and model predictions are shown, we applied it to the CREW dataset to retrieve useful pick times for arrivals that were not previously labeled. Figure 9 (b-d) displays the waveforms and model predictions for examples that have three or two labeled arrivals, for which our model accurately picks the missing arrival time labels. In Figure 9(b) there were three labeled arrivals in the CREW dataset, and in Figure 9(c-d) there were two labels, and our model can retrieve four. We run model predictions for the 1.6 million ex-

amples in the CREW dataset and collected all the examples for which the four phases are picked with a score higher than 0.8 and both the first arriving P and S model predicted pick times differ by less than 2 seconds from the labeled arrivals. Using these metrics, we identified 22,595 examples that were later visually inspected for the presence of the four phases and the accuracy of the model picks. In the end, we kept 3,093 new examples for which the four phases are visible and accurately picked by our model. Examples that were visually reviewed but rejected are shown in Figure S8, where the model predicts picks with high probability, but either one of the phases is not present or it is not accurately picked. The newly found examples with both clear mantle and crustal phases are about 50% of the number of those in the training dataset. Even though we trained our multi-phase picker on very little data, its use led to a significant increase in the number of examples with four labeled phases. Further iterations of this procedure will lead to the discovery of more high quality examples, but is outside of the scope of this work.

7 Discussion

We have trained and deployed models that pick earthquake arrivals at regional distances, which is most use-

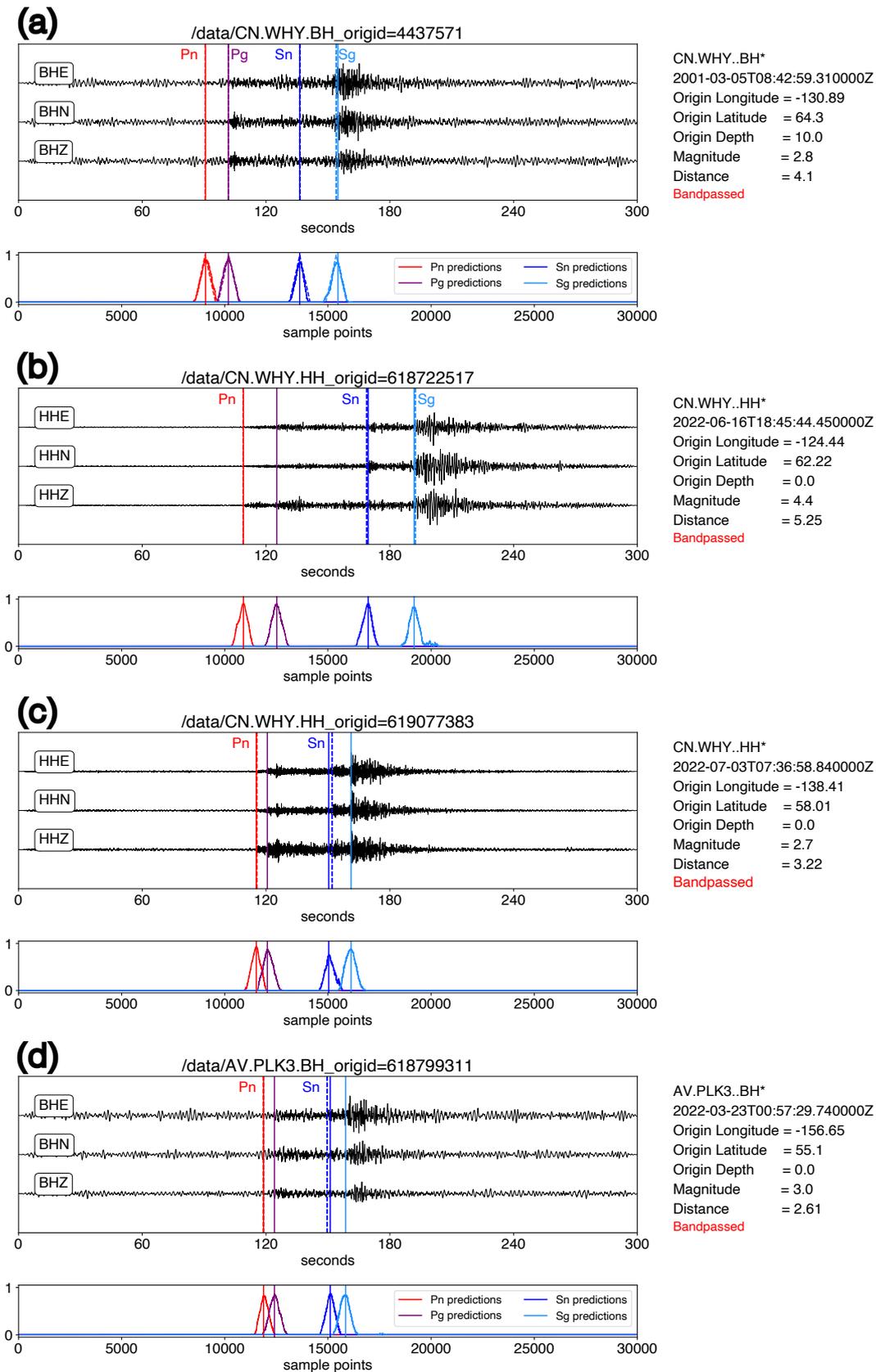


Figure 9 (a) Example from the CREW dataset that has four labeled phases Pn, Pg, Sn, and Sg, used for training. Labeled arrivals are signaled by the letters. Lower panel shows the labels with dotted lines and the predictions with solid lines alongside the inferred phase picks. (b) Example from CREW with three labeled arrivals (Pn, Sn, Sg), for which our previously trained model successfully recovered the three labeled picks and found the fourth pick for Pg. (c) Example with two labels (Pn and Sn) for which our model picked both Pg and Sg. (d) Another example with two labeled arrivals (Pn and Sn) for which our model picked both Pg and Sg. Note that all the waveforms displayed here are bandpass filtered from 0.5-20 Hz for visualization. Raw waveforms for (a) are shown in Figure S7.

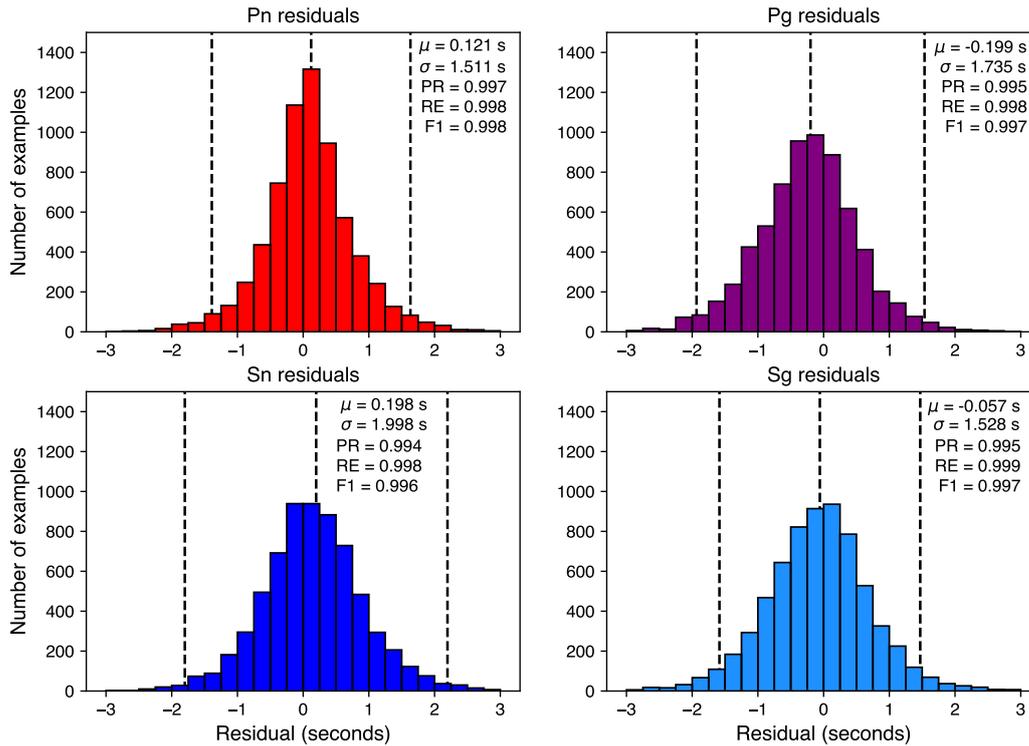


Figure 10 Residuals for each one of the four phases Pn, Pg, Sn and Sg in the training data. PR for precision, RE for recall and F1 for the F1-score.

ful when seismicity lies outside of the footprint of the network, or there is sparse instrumental coverage. Our deep learning picks are highly precise and accurate in the test data (Figure 4), regardless of source-receiver distance or signal to noise ratio (Figure S3). The strength of our SKYNET models lies in the data they were trained on, a global collection of quality controlled earthquake waveforms and labeled arrivals (Figure 1) complemented with data augmentation (Figure 3). Our models have long receptive fields and process long waveforms while maintaining comparable computational cost (Figure S6). We used these models to find numerous new earthquakes on the Gorda Ridge offshore Northern California (Figures 5, 6) and to pick earthquake arrivals in Northern Chile, from earthquakes near the trench (Figure 7), as well as very deep earthquakes under western Brazil (Figure 8). Our models unlock new regions to be seen under the lens of machine learning enhanced earthquake catalogs. There is value on improving monitoring in places like ocean transform faults, that show patterns unlike other tectonic settings (McGuire et al., 2005).

In addition to finding new earthquakes, the capabilities of phase picking at regional distances will help imaging studies. In a well-instrumented region such as California, data in the 0-100 km range is the key to an accurate location; however, having numerous S wave arrivals at distances of hundreds of kilometers (Figure S5) will help refine velocity models, particularly of the upper mantle. The shortcomings of our model for picking

near source seismograms (Figure S5) will be addressed in future work where we will integrate both local and regional data.

Benchmark datasets for seismology rely almost entirely on human labels. The majority of these labels are first arrival P waves. Datasets that contain secondary arrivals are an order of magnitude less numerous than those with first arrivals. The ISC_EHB_DepthPhases (Münchmeyer et al., 2023) dataset contains 174K examples, and the subset of the CREW dataset that contains both mantle and crustal arrivals is only 6,705 examples. This is in stark contrast with the 1.6 million examples in CREW and the 1.1 million examples in STEAD. Our approach of using the multi-phase model to increase the number of labeled secondary arrivals highlights the potential of label supplementation within established datasets. Improved capabilities for picking and characterizing later phases will aid in studies of Earth's structure, for instance in imaging slab structure (Shiina et al., 2021), creating refined tomographic images (Liu and Zhao, 2018; Zhao, 2019), improving earthquake locations and depths (Münchmeyer et al., 2023) for earthquakes recorded at long distances, and discriminating between mantle and crustal earthquakes based on Sn and Lg amplitude ratios (Wang and Klemperer, 2021; Song and Klemperer, 2024).

In both the CREW (Aguilar Suarez and Beroza, 2024) and STEAD (Mousavi et al., 2019) datasets only 10% of the processed metadata contains waveforms that include both P and S picks at the same station for the same

earthquake. Revisiting previously cataloged data to add S wave arrivals for a massive number of known earthquakes would increase the volume, variety, and rate of growth of labeled data that can be used to train future machine learning models. This approach was demonstrated by (Ni et al., 2023) in the assembly of the Curated Pacific Northwest AI-ready Seismic Dataset. Also, semisupervised approaches have proven effective in the Tonga region where small quantities of labeled data were available (Xi et al., 2024).

The road towards all-purpose, all-scale seismic phase picking will require balancing and filling gaps in existing datasets. Some examples include the CWA dataset (Tang et al., 2024) made of earthquakes with magnitude larger than five, as a complement to existing datasets that contain very few moderate and large events; the GTUNE dataset (Barama et al., 2022) of underground nuclear explosion waveforms; the VCSEIS dataset (Zhong and Tan, 2024) of volcano-tectonic earthquakes; the OBS PickBlue dataset (Bornstein et al., 2024); and the OBST2024 dataset (Niksejel and Zhang, 2024) dedicated to earthquakes recorded by ocean bottom seismometers. Clever data augmentation in the forms of rescaling and resampling proved useful in improving model performance (Shi et al., 2024). Also, the development of foundation models (Liu et al., 2024) that can be fine-tuned to domain specific tasks will help advance multiscale earthquake monitoring.

8 Future directions

We trained and deployed machine learning powered earthquake phase pickers dedicated to arrivals from 1 to 20 degrees of source-receiver distance. By doing so we aim to bridge the gap between dense comprehensive earthquake catalogs and sparse seismic networks and/or remote seismicity that require effective phase picking at regional distances. The comparatively higher number of picked S waves compared to well established ML models at regional distances, should improve the quality of earthquake locations and enable new insight into Earth structure at regional scales. Future models will incorporate capabilities to train and pick across the distance scales, covering local, regional and teleseismic arrivals.

8.1 Code

Our codes and models are contained within the **SKYNET** package, **Seismological Knowledge Yardstick Networks**. The output of our picker should take only a few lines of code to produce. We are working on making these models interoperable with other platforms to facilitate comparison and other downstream tasks.

Listing 1 Example use of skynet

```
import skynet
from obspy import read

model = skynet.load_model('regional_picker')
st = read('daylong_NC.mseed')
outname = 'NC_daypicks.csv'
picks = skynet.execute(st,model,0.5,outname)
plot = skynet.plot_picks(st,picks)
```

Acknowledgements

This work was supported by AFRL under contracts and PA04-S-D00243-12-TO-01-Stanford and PA04-S-D00243-12-TO-03-Stanford. We thank the comments from multiple people that helped improve this manuscript. We thank Jeff McGuire for providing the velocity model from the offshore region of northern California. We thank Ian McBrearty for assisting with GENIE. We also thank Bill Ellsworth and Rosie Ries for their comments which helped improve this manuscript. We thank editor Hongyu Sun and three anonymous reviewers for their help in improving this manuscript. The computing for this project was performed on the Stanford Sherlock cluster. We would like to thank Stanford University and the Stanford Research Computing Center for providing computational resources and support that contributed to these research results.

Data and code availability

Codes for SKYNET can be found here: <https://github.com/albertleonardo/skynet>. The Curated Regional Earthquake Waveforms (CREW) dataset is available at <https://redivis.com/datasets/1z6w-e1w70hpmt>. It is also accessible through Seisbench https://seisbench.readthedocs.io/en/stable/pages/benchmark_datasets.html#crew

The figures show waveforms from the networks BK Northern California Earthquake Data Center (2014), NC USGS Menlo Park (1966), CI California Institute of Technology and United States Geological Survey Pasadena (1926), NN University of Nevada, Reno (1971), and CX GFZ German Research Centre For Geosciences and Institut Des Sciences De L'Univers-Centre National De La Recherche CNRS-INSU (2006).

The earthquake catalog for the Gorda ridge comes from the USGS, last accessed March 9, 2024.

Competing interests

The authors declare no competing interests. This is a chatGPT free manuscript.

References

- Aguilar Suarez, A. L. and Beroza, G. Curated Regional Earthquake Waveforms (CREW) Dataset. *Seismica*, 3(1), may 2024. doi: 10.26443/seismica.v3i1.1049.
- Al-Damegh, K., Sandvol, E., Al-Lazki, A., and Barazangi, M. Regional seismic wave propagation (Lg and Sn) and Pn attenuation in the Arabian Plate and surrounding regions. *Geophysical Journal International*, 157(2):775–795, 2004. doi: 10.1111/j.1365-246X.2004.02246.x.
- Barama, L., Peng, Z., Newman, A. V., and Williams, J. GTUNE: An Assembled Global Seismic Dataset of Underground Nuclear Test Blasts. *Seismological Research Letters*, 93(6):3514–3523, 2022. doi: 10.1785/0220220036.
- Bornstein, T., Lange, D., Münchmeyer, J., Woollam, J., Rietbrock, A., Barcheck, G., Grevemeyer, I., and Tilmann, F. PickBlue: Seismic Phase Picking for Ocean Bottom Seismometers With

- Deep Learning. *Earth and Space Science*, 11(1), 2024. doi: 10.1029/2023EA003332.
- California Institute of Technology and United States Geological Survey Pasadena. Southern California Seismic Network, 1926. doi: 10.7914/SN/CI.
- Cole, H. M., Yeck, W. L., and Benz, H. M. MLAAPDE: A Machine Learning Dataset for Determining Global Earthquake Source Parameters. *Seismological Research Letters*, 20(20):1–11, 2023. doi: 10.1785/0220230021.
- GFZ German Research Centre For Geosciences and Institut Des Sciences De L'Univers-Centre National De La Recherche CNRS-INSU. IPOC Seismic Network, 2006. doi: 10.14470/PK615318.
- Isken, M. P., Niemz, P., Münchmeyer, J., Büyükkapınar, P., Cesca, S., Vasyura-bathke, H., and Dahm, T. Qseek : A Data-Driven Framework for Automated Earthquake Detection , Localization and Characterization. *Seismica*, 4, 2025. doi: 10.26443/seismica.v4i1.1283.
- Kong, H., Xiao, Z., Lü, Y., and Li, J. Seis-PnSn: A Global Million-Scale Benchmark Data Set of Pn and Sn Seismic Phases for Deep Learning. *Seismological Research Letters*, 07 2024. doi: 10.1785/0220230379.
- Liu, T., Münchmeyer, J., Laurenti, L., Marone, C., de Hoop, M. V., and Dokmanić, I. SeisLM: a Foundation Model for Seismic Waveforms. 140(NeurIPS), 2024. <http://arxiv.org/abs/2410.15765>.
- Liu, X. and Zhao, D. Upper and lower plate controls on the great 2011 Tohoku-oki earthquake. *Science Advances*, 4(6), 2018. doi: 10.1126/sciadv.aat4396.
- McBrearty, I. W. and Beroza, G. C. Earthquake Phase Association with Graph Neural Networks. *Bulletin of the Seismological Society of America*, 113(2):524–547, 2023. doi: 10.1785/0120220182.
- McGuire, J. J., Boettcher, M. S., and Jordan, T. H. Foreshock sequences and short-term earthquake predictability on East Pacific Rise transform faults (Nature (2005) 434 (457-461)). *Nature*, 435(7041):528, 2005. doi: 10.1038/nature03621.
- Mousavi, S. M. and Beroza, G. C. A Machine-Learning Approach for Earthquake Magnitude Estimation. *Geophysical Research Letters*, 47(1):1–7, 2020. doi: 10.1029/2019GL085976.
- Mousavi, S. M., Sheng, Y., Zhu, W., and Beroza, G. C. STanford Earthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access*, 7:179464–179476, 2019. doi: 10.1109/ACCESS.2019.2947848.
- Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., and Beroza, G. C. Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, 11(1):1–12, 2020. doi: 10.1038/s41467-020-17591-w.
- Münchmeyer, J., Saul, J., and Tilmann, F. Learning the Deep and the Shallow: Deep-Learning-Based Depth Phase Picking and Earthquake Depth Estimation. *Seismological Research Letters*, 2023. doi: 10.1785/0220230187.
- Ni, Y., Hutko, A. R., Skene, F., Denolle, M. A., Malone, S. D., Bodin, P., Hartog, J. R., and Wright, A. K. Curated Pacific Northwest Already Seismic Dataset. *Seismica*, 2, 2023. doi: 10.26443/seismica.v2i1.368.
- Niksejel, A. and Zhang, M. OBSTransformer: a deep-learning seismic phase picker for OBS data using automated labelling and transfer learning. *Geophysical Journal International*, 237(1): 485–505, feb 2024. doi: 10.1093/gji/ggae049.
- Northern California Earthquake Data Center. Berkeley Digital Seismic Network (BDSN), 2014. doi: 10.7932/BDSN.
- Novoselov, A., Balazs, P., and Bokelmann, G. SEDENOSS: SEparating and DENOising Seismic Signals With Dual-Path Recurrent Neural Network Architecture. *Journal of Geophysical Research: Solid Earth*, 127(3):1–22, 2022. doi: 10.1029/2021JB023183.
- Park, Y., Mousavi, S. M., Zhu, W., Ellsworth, W. L., and Beroza, G. C. Machine-Learning-Based Analysis of the Guy-Greenbrier, Arkansas Earthquakes: A Tale of Two Sequences. *Geophysical Research Letters*, 47(6):1–8, 2020. doi: 10.1029/2020GL087032.
- Pasyanos, M. E., Walter, W. R., and Matzel, E. M. A simultaneous multiphase approach to determine P-Wave and S-Wave attenuation of the crust and upper mantle. *Bulletin of the Seismological Society of America*, 99(6):3314–3325, 2009. doi: 10.1785/0120090061.
- Rodgers, A. J. and Walter, W. R. Seismic discrimination of the May 11, 1998 Indian nuclear test with short-period regional data from station NIL (Nilore, Pakistan). *Pure and Applied Geophysics*, 159(4):679–700, 2002. doi: 10.1007/s00024-002-8654-6.
- Ross, Z. E., Meier, M. A., and Hauksson, E. P Wave Arrival Picking and First-Motion Polarity Determination With Deep Learning. *Journal of Geophysical Research: Solid Earth*, 123(6):5120–5129, 2018. doi: 10.1029/2017JB015251.
- Schultz, R., Park, Y., Aguilar Suarez, A. L., Ellsworth, W. L., and Beroza, G. C. En echelon faults reactivated by wastewater disposal near Musreau Lake, Alberta. *Geophysical Journal International*, 235(1):417–429, 2023. doi: 10.1093/gji/ggad226.
- Shi, P., Meier, M.-A., Villiger, L., Tuinstra, K., Selvadural, P., Lanza, F., Yuan, S., Obermann, A., Mesimeri, M., Münchmeyer, J., and Others. From labquakes to megathrusts: Scaling deep learning based pickers over 15 orders of magnitude. *Journal of Geophysical Research: Machine Learning and Computation*, 1, 2024. doi: 10.1029/2024JH000220.
- Shiina, T., Katsumata, K., Yomogida, K., and Kato, A. Attenuation contrast in mantle wedge across the volcanic front of north-eastern Japan that controls propagations of high-frequency S-wave later phases. *Earth, Planets and Space*, 73(1), 2021. doi: 10.1186/s40623-021-01361-z.
- Smith, J. D., Ross, Z. E., Azizzadenesheli, K., and Muir, J. B. HypoSVI: Hypocentre inversion with Stein variational inference and physics informed neural networks. *Geophysical Journal International*, 228(1):698–710, 2022. doi: 10.1093/gji/ggab309.
- Song, X. and Klemperer, S. L. Numerous Tibetan lower-crustal and upper-mantle earthquakes, detected by Sn/Lg ratios, suggest crustal delamination or drip tectonics. *Earth and Planetary Science Letters*, 626(December 2023):118555, 2024. doi: 10.1016/j.epsl.2023.118555.
- Tan, Y. J., Waldhauser, F., Ellsworth, W. L., Zhang, M., Zhu, W., Michele, M., Chiaraluca, L., Beroza, G. C., and Segou, M. Machine-Learning-Based High-Resolution Earthquake Catalog Reveals How Complex Fault Structures Were Activated during the 2016–2017 Central Italy Sequence. *The Seismic Record*, 1(1): 11–19, 2021. doi: 10.1785/0320210001.
- Tang, K., Chen, K., Chen, D., Chin, T., and Hsu, T. The CWA Benchmark: A Seismic Dataset from Taiwan for Seismic Research. *Seismological Research Letters*, dec 2024. doi: 10.1785/0220230393.
- Uchide, T. Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning. *Geophysical Journal International*, 223(3): 1658–1671, 2020. doi: 10.1093/gji/ggaa401.
- University of Nevada, Reno. Nevada Seismic Network, 1971. doi: 10.7914/SN/NN.
- USGS Menlo Park. USGS Northern California Seismic Network, 1966. doi: 10.7914/SN/NC.
- Wang, S. and Klemperer, S. L. Love-wave normal modes discriminate between upper-mantle and crustal earthquakes: Simulation and demonstration in Tibet. *Earth and Planetary Science Letters*, 571:117089, 2021. doi: 10.1016/j.epsl.2021.117089.

- Wilding, J. D., Zhu, W., Ross, Z. E., and Jackson, J. M. The magmatic web beneath Hawai'i. *Science*, 379(6631):462–468, 2023. doi: 10.1126/science.ade5755.
- Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., Michelini, A., Saul, J., and Soto, H. SeisBench-A Toolbox for Machine Learning in Seismology. *Seismological Research Letters*, 93(3):1695–1709, 2022. doi: 10.1785/0220210324.
- Xi, Z., Wei, S. S., Zhu, W., Beroza, G. C., Jie, Y., and Saloor, N. Deep learning for deep earthquakes: insights from OBS observations of the Tonga subduction zone. *Geophysical Journal International*, 238(2):1073–1088, 2024. doi: 10.1093/gji/ggae200.
- Yeck, W. L., Patton, J. M., Ross, Z. E., Hayes, G. P., Guy, M. R., Nick, B., Shelly, D. R., Benz, H. M., and Earle, P. S. Leveraging Deep Learning in Global 24 / 7 Real-Time Earthquake Monitoring at the National Earthquake Information Center. *Seismological Research Letters*, 2020. doi: 10.1785/0220200178.
- Zhao, D. Importance of later phases in seismic tomography. *Physics of the Earth and Planetary Interiors*, 296(August):106314, 2019. doi: 10.1016/j.pepi.2019.106314.
- Zhong, Y. and Tan, Y. J. Deep-Learning-Based Phase Picking for Volcano-Tectonic and Long-Period Earthquakes. *Geophysical Research Letters*, 51(12), 2024. doi: 10.1029/2024GL108438.
- Zhu, W. and Beroza, G. C. PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, 216(1):261–273, 2019. doi: 10.1093/gji/ggy423.
- Zhu, W., Mousavi, S. M., and Beroza, G. C. Seismic signal denoising and decomposition using deep neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, 57(11):9476–9488, 2019.

The article *Picking Regional Seismic Phase Arrival Times with Deep Learning* © 2025 by A.L. Aguilar Suarez is licensed under CC BY 4.0.