ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2021GL093157

Key Points:

- We trained different variations of a U-shaped convolutional neural network to detect low-frequency earthquakes on the San Andreas fault
- The model identifies known and new low-frequency earthquakes in continuous seismic data, including data from stations not used in training
- The preferred model has a peak accuracy of 85%, does not require repeating sources, and is a fast way to identify low-frequency earthquakes

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

A. M. Thomas,
amt.seismo@gmail.com

Citation:

Thomas, A. M., Inbal, A., Searcy, J., Shelly, D. R., & Bürgmann, R. (2021). Identification of low-frequency earthquakes on the San Andreas fault with deep learning. *Geophysical Research Letters*, *48*, e2021GL093157. https://doi.org/10.1029/2021GL093157

Received 1 MAR 2021
Accepted 26 MAY 2021

Identification of Low-Frequency Earthquakes on the San Andreas Fault With Deep Learning

Amanda M. Thomas[1], Asaf Inbal[2], Jacob Searcy[3], David R. Shelly[4], and Roland Bürgmann[5]

[1]Department of Earth Sciences, University of Oregon, Eugene, OR, USA, [2]Department of Geophysics, Tel Aviv University, Tel Aviv, Israel, [3]Research Advanced Computing Services, University of Oregon, Eugene, OR, USA, [4]Geologic Hazards Science Center, United States Geological Survey, Golden, CO, USA, [5]Department of Earth and Planetary Science, University of California, Berkeley, CA, USA

Abstract Low-frequency earthquakes are a seismic manifestation of slow fault slip. Their emergent onsets, low amplitudes, and unique frequency characteristics make these events difficult to detect in continuous seismic data. Here, we train a convolutional neural network to detect low-frequency earthquakes near Parkfield, CA using the catalog of Shelly (2017), https://doi.org/10.1002/2017jb014047 as training data. We explore how varying model size and targets influence the performance of the resulting network. Our preferred network has a peak accuracy of 85% and can reliably pick low-frequency earthquake (LFE) S-wave arrival times on single station records. We demonstrate the abilities of the network using data from permanent and temporary stations near Parkfield, and show that it detects new LFEs that are not part of the Shelly (2017), https://doi.org/10.1002/2017jb014047 catalog. Overall, machine-learning approaches show great promise for identifying additional low-frequency earthquake sources. The technique is fast, generalizable, and does not require sources to repeat.

Plain Language Summary Low-frequency earthquakes are an enigmatic seismic signal that often accompany slow earthquakes in many tectonic environments. They are weak seismic signals, and, as such, are challenging to find using routine approaches to earthquake detection. In this study, we use machine learning, a type of artificial intelligence, to discover low-frequency earthquakes that are hidden in continuous seismic data. This approach is computationally efficient and is capable of identifying both known and new earthquake sources. This new approach to low-frequency earthquake detection has the potential to discover many more earthquakes than traditional approaches which may help scientists better understand slow earthquakes.

1. Introduction

Slow slip events occur frequently on the shallow and deep extents of subduction megathrusts and on transform faults throughout the world (e.g., Behr & Bürgmann, 2021). In slow slip events, the fault slip velocity is ~1 to 2 orders of magnitude above the background plate rate and orders of magnitude below typical earthquakes, hence slow slip has only a very weak seismic manifestation. Commonly, slow slip is accompanied by low-frequency earthquakes (LFEs). LFEs manifest as low-amplitude, emergent signals. P- and S-wave arrivals are sometimes visible in individual traces but are most apparent in waveform stacks (Figure 1a). LFEs are also depleted in high-frequency content relative to traditional earthquakes of similar size. This depletion could be due to slip speeds that are slower than typical earthquakes, attenuation along the path, or some combination of the two (Bostock et al., 2015, 2017; Hawthorne et al., 2019; Littel et al., 2018; Thomas et al., 2016).

LFEs are commonly identified using variations of a network cross-correlation approach or template matching (e.g., Bostock et al., 2012; Chamberlain et al., 2014; Frank et al., 2013; Royer & Bostock, 2014; Shelly et al., 2007; Sweet et al., 2014; Tang et al., 2010; Thomas & Bostock, 2015), though some studies apply alternative network-based approaches (e.g., Frank & Shapiro, 2014; Poiata et al., 2018; Rubin & Armbruster, 2013). Network cross-correlation involves taking an LFE waveform and cross-correlating it through continuous data. This results in a cross-correlation function for each station and component that are then summed or averaged, resulting in a network cross-correlation function. When the network cross-correlation function

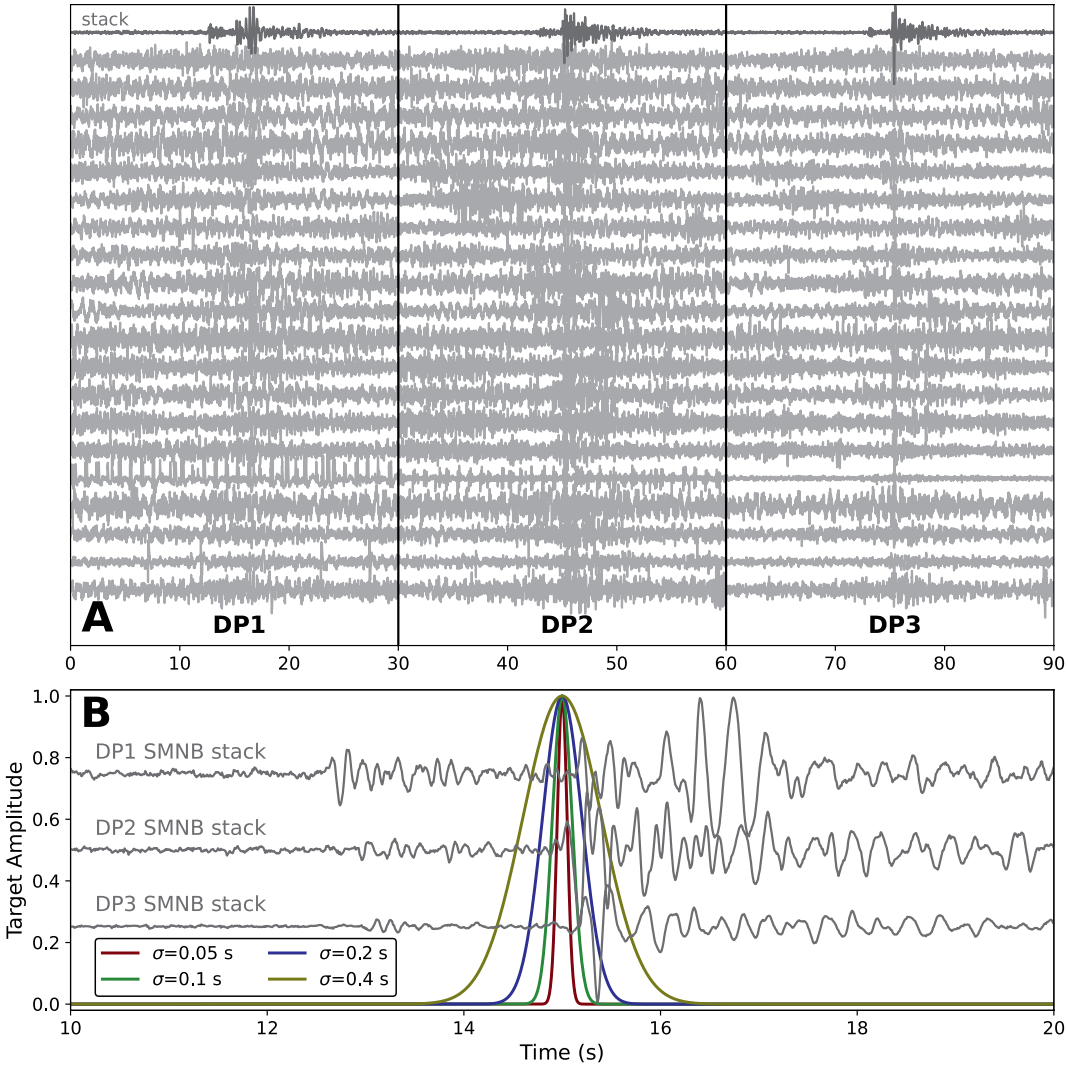

Figure 1. Example inputs and outputs of the convolutional neural network. Panel (a) shows example input data from high-resolution seismic network borehole station SMNB. Each trace represents an instance of an low-frequency earthquake (LFE) from the same family centered on the S-wave arrival on each component (channel DP1 is the vertical whereas DP2 and DP3 are horizontals). Note that the signal-to-noise ratio is low in the individual traces, and it would be difficult to identify LFEs from these records. However, when stacked ($N = 945$) P- and S-waves are visible on all three components (dark gray waveform at top). Panel (b) shows three target Gaussian distributions with different standard deviations centered on the S-wave arrival.

exceeds a given threshold, typically eight times the median absolute deviation (Shelly et al., 2007), the corresponding window pairs are considered a detection. Detections are then stacked on each station and channel to create a LFE template with a higher signal-to-noise ratio, and the process is repeated. This produces families of repeating LFEs identified by waveform similarity. An obvious limitation of this method is that it requires LFE sources to repeat. If we want to move toward producing a systematic, comprehensive characterization of the seismic radiation associated with slow slip and more thoroughly image the slow slip source region using LFEs, the detection method employed should be tolerant of repeating and non-repeating sources alike.

A number of recent studies have shown that machine learning, deep learning in particular, excels at common tasks in seismology such as earthquake detection and phase picking (e.g., Mousavi et al., 2020; Perol et al., 2018; Ross et al., 2018; Yeck et al., 2021; Zhu & Beroza, 2019), phase association (Ross et al., 2019), network operations (Walter et al., 2021), tremor detection (Rouet-Leduc et al., 2020), ground motion prediction

(Jozinović et al., 2020; Klimasewski et al., 2020), etc. Deep learning based approaches to earthquake detection and phase identification require large, labeled datasets (e.g., at least several hundred thousand examples of P-waves, S-waves, and noise) for training; often these datasets are analyst-reviewed phase picks cataloged by regional seismic networks. One limitation of applying a model trained on regional earthquakes to detect nontraditional seismicity is that the model may not generalize well to types of seismicity that are not well represented in the training data set. For example, Yoon (2021) recently showed that a commonly employed phase picker, EQ Transformer (Mousavi et al., 2020), failed to detect the January 2020 magnitude 6.4 earthquake in Puerto Rico, as well as several aftershocks greater than magnitude 3. This is likely because earthquake size distributions adhere to the Gutenberg-Richter frequency magnitude distribution and hence larger-magnitude earthquakes were not well represented in the training data set (Yeck et al., 2021). Given this result, existing earthquake detection algorithms developed using machine learning may not generalize well to nontraditional types of seismicity, such as LFEs, that are not represented in the training data.

Here, we use the 2001–2016 LFE catalog of Shelly (2017) from Parkfield, CA to train several variations of a U-Net (Ronneberger et al., 2015), a type of convolutional neural network (CNN), to detect LFEs on the deep extension of the San Andreas fault (SAF). Individual LFEs in the Shelly (2017) catalog were visually identified in continuous seismic data, and the LFE catalog was assembled using the network cross-correlation approach described above. We demonstrate that the best-performing CNN can detect LFEs in continuous seismic data from borehole stations, surface stations, and temporary stations such as nodal seismometers. The approach is fast, generalizeable, and has the potential to identify many more LFEs (beyond those represented in the training data). More complete LFE catalogs may help further elucidate the mechanism(s) responsible for slow fault slip.

2. Data and Methods

We use the catalog of LFEs assembled by Shelly (2017), which contains over 1 million LFEs recorded at multiple stations surrounding the central SAF. While this catalog contains only 88 families of events with highly similar waveforms, noise is also present in each individual trace, and, as such, each LFE record represents a unique piece of training data. For each LFE family, we attempt to download the 1,000 catalog events with the largest average template cross correlation values and calculate the S-wave arrival time at all stations with high-quality S-wave picks (with a pick quality of good or better assigned by Shelly & Hardebeck, 2010) in the stacked LFE templates located by Shelly and Hardebeck (2010). We download a 30-s window of time centered on the S-wave arrival on all three components. For one-component stations, we copy the single-component record to the remaining components. These records are then detrended, high-pass filtered using a corner frequency of 1 Hz, and resampled to a frequency of 100 Hz, if the sample rate is not 100 Hz. To train our network to distinguish between LFEs and noise, we download "noise" samples as well, which are randomly selected windows from time periods in which no LFEs are present in the catalog. These records are processed in the same manner as the LFE records. This process results in a data set of 1.73 million waveforms recorded on 65 different stations. Example LFE waveforms are shown in Figure 1. Note that LFE S-wave arrivals are often not visible in the individual traces (light gray traces in Figure 1a); only after stacking do P and S-wave arrivals emerge from the noise (stacked traces are in dark gray in Figure 1a).

After these processing steps, we input three-component seismic data to the model. We use a data generator during training, which randomly selects sets of traces from the training data, called batches, and applies the following modifications to the data prior to input. First, we randomly select a start time in the first half of the window and include only 15 s of data beginning at that time. This has the effect of randomly shifting the arrival in time such that it can occur at any time during the window. Second, to account for highly variable amplitudes in the training data, we apply a logarithmic transformation to the input data. This transformation maps each value, x, in the original traces to two numbers: the first is $sgn(x)$, while the second is the $ln(abs(x) + eps)$, where $eps = 1 \times 10^{-6}$. This has the effect of normalizing the data such that input amplitudes do not vary over orders of magnitude and preserving information on the sign. The data generator supplies six channels (three components with a normalized amplitude and sign for each) in batches to the CNN during training.

For our model, we employ a variation of a U-Net architecture (Ronneberger et al., 2015) shown in Table S1 and Figure S1. Similar architectures have been shown to be successful at earthquake phase identification (e.g., Zhu & Beroza, 2019). To explore how variations in network size affect performance, we include two additional variations in the network size shown in Table S1. In the first, we increase the number of filters in the convolutional layers by a factor of two (called the size 2 network). In the second, we reduce the number of filters in the convolutional layers by a factor of two (called the size 0.5 network). For all networks, the target output is a Gaussian window centered on the S-wave arrival. Since LFE arrivals are emergent and not impulsive, we explore the influence of varying the width (σ, the standard deviation) of the target; we explore $\sigma = 0.05$, 0.1, 0.2, and 0.4 s. Examples of targets with varying widths, and their alignment with the data, are shown in Figure 1b. For noise samples, we also take the 15 s of data beginning at a randomly selected start time however the target is set to zero.

We reserve 10% of our data for testing the resulting networks. Each network is trained for 50 epochs using Tensorflow (Abadi et al., 2016). For all networks, we use the Adam optimization algorithm (Kingma & Ba, 2014) with a learning rate of 0.0001 and a binary cross-entropy loss function.

As a simple test, we apply one of our best performing models to three nearby stations near the Cholame section of the SAF. The first station, THIS, is a permanent, broadband surface station, the second is a nodal seismometer, station 40, deployed near THIS as part of a 3-month temporary experiment in the fall of 2018, and the third is a borehole station B079 (Figure 2). We note that data from THIS and 40 was not included in our training data set. We select a time period in early August 2018 during which abundant LFE activity was recorded on stations near the southernmost LFE families shown in Figure 2. From the catalog of Shelly (2017), LFE activity during this time was largely confined to the LFE families in the SE section of the central SAF (horizontal bar in Figure 2). Families with the highest number of detections are located beneath the temporary array, station THIS, and B079, between 20 and 40 km along fault.

For each station, we apply the model to continuous three-component data generating a prediction or output functional form for each time window. Since the CNN was designed to output a target Gaussian distribution when it is confident that an LFE is present, we define detections as times when a peak with a minimum amplitude of 0.1 is present in the CNN prediction. We take the detection time as the time corresponding to the peak height and require that detections occur >2 s apart. To compare the CNN results to the LFE detections in the same time period, we take all LFEs in the Shelly (2017) catalog from families that locate in the southeastern section of the central SAF, determine S-wave travel times between the source and THIS using stacked LFE detections or interpolated travel times from Shelly (2017), and add the two to determine the arrival time.

3. Results

The training loss (a measure of misfit on the training data) and validation loss (measure of misfit on the testing data) for all $\sigma = 0.2$ s networks are shown in Figure 3a. In machine learning, a balance must be struck between allowing the network to be sufficiently large that it can perform effectively, but not so large that it becomes overfit and does not generalize to new datasets well. The ratio of validation loss to training loss provides a measure of how overfit the network is. Ratios of one are desirable, indicating that the network fits the testing data about as well as the training data, whereas ratios significantly greater than one are an indicator that the network is likely overfit. Figure 3a shows the training loss and validation loss decrease as a function of epochs and begin to plateau around 50 epochs. The size 0.5 network has lower training loss and validation loss followed by the size 1 network and the size 2 network. The loss ratio for three networks at 50 epochs are all near 1, suggesting that the networks do not suffer significantly from overfitting. After exploring other target widths and the influence of a drop layer, we find that the smaller size networks generally have lower validation loss and loss ratios than the size 2 networks. This suggests they are sufficiently complex to perform well without being overfit.

Figure 3b shows the accuracy, precision, and recall for the three different variations of the size 0.5 network. We define true positives (TP) as correct signal predictions, true negatives (TN) as correct noise predictions, false positives (FP) as incorrect signal predictions, false negatives (FN) as incorrect noise predictions, and $N = TP + FP + TN + FN$ as the total number of signal and noise predictions. Accuracy is the fraction of

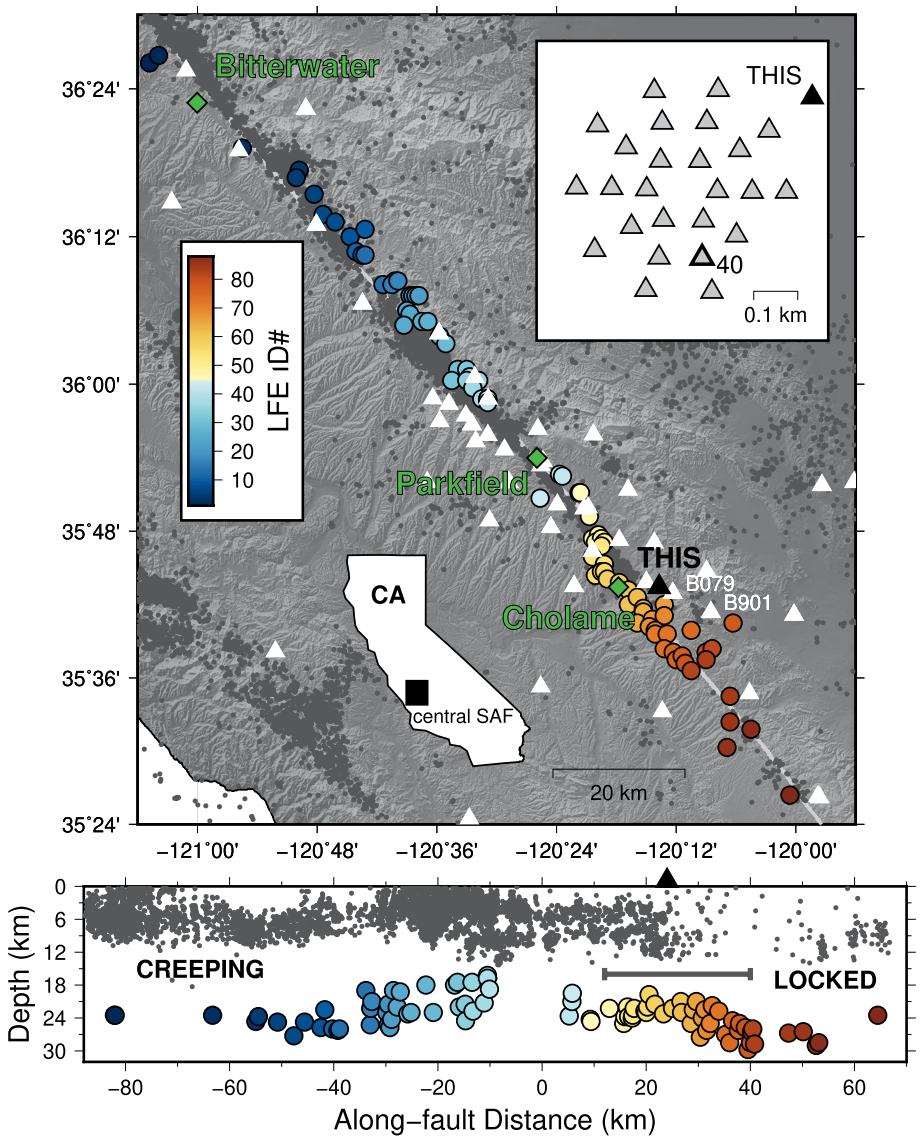

Figure 2. Map and cross section of the central San Andreas fault (SAF). Upper panel shows regional seismicity as black dots, low-frequency earthquake (LFE) locations are color coded by their family ID number (from Thomas et al., 2018), station THIS is indicated by a black triangle. White triangles indicate locations of stations that recorded LFEs and noise samples used to train the convolutional neural network (CNN). The lower inset shows the location of the central SAF. The upper right inset shows the geometry of a subarray of stations, indicated by gray triangles that operated as part of a temporary deployment in 2018. The array is nearly colocated with station THIS (THIS is shown in black). The data shown in Figure 4b below is from station 40. The cross section shows LFE families and shallow seismicity along the SAF going from the NW (left) to SE (right). Black dots are regional seismicity occurring within 10 km of the SAF. The along-strike extent of the 2018 slow-slip event is denoted by the horizontal bar.

correct model predictions and is defined as $(TP + TN)/N$. Precision is defined as $TP/(TP + FP)$ and describes the fraction of positive predictions that are correct. Recall is defined as $TP/(TP + FN)$ and describes the fraction of positive instances that are correctly classified. If the maximum values of the model output exceeds the decision threshold, we classify the window as signal and otherwise we classify the window as noise. Recall decreases as the decision threshold increases, reflecting how increasing the decision threshold reduces the fraction of true LFEs that are detected. Increasing precision with decision threshold reflects an increasing fraction of true LFEs among the CNN positive predictions. The maximum accuracy for all $\sigma = 0.2$ snetworks is between 82% and 87%. The accuracy of the $\sigma = 0.05$ s network decreases monotonically, whereas the $\sigma = 0.1$ s, $\sigma = 0.2$ s, and $\sigma = 0.4$ s networks peak at decision thresholds of 0.02, 0.04, and 0.08,

10.1029/2021GL093157

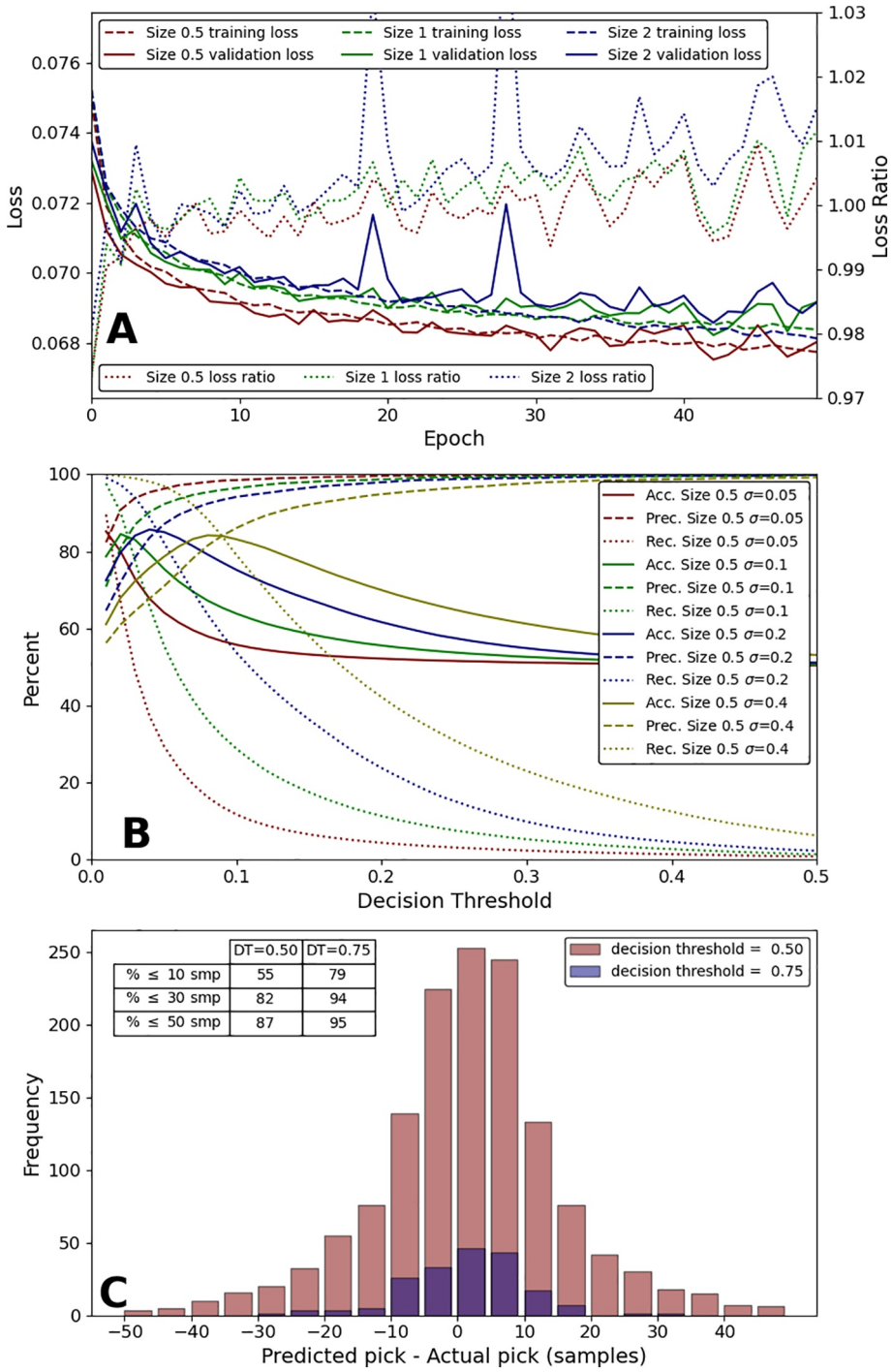

Figure 3. Convolutional neural network (CNN) performance metrics. Panel (a) shows the training loss, validation loss, and loss ratio (validation loss/training loss) as a function of epoch for the three different size networks for $\sigma = 0.2$ s. Panel (b) shows the precision, recall, and accuracy as a function of decision threshold for the size 0.5 networks. Panel (c) shows a histogram of the picking performance, that is, the difference between the predicted and actual pick, for the size 0.5, $\sigma = 0.2$ s network. The inset table presents additional picking metrics, that is, the percentage of picks that are less than or equal to 10, 30, and 50 samples for the two decision thresholds (DT).

respectively. Overall, precision, recall, and accuracy are high for all networks at very low decision thresholds, with the size 0.5, $\sigma = 0.4$ s having the highest accuracy at a decision threshold of 0.08. Increasing the decision threshold has the effect of reducing the number of false positive (i.e., spurious) detections at the expense of permitting more false negatives (i.e., missing true LFEs). Distributions of maximum model outputs for populations of LFEs and noise windows are shown in Figure S2.

Figure 3c shows the picking performance of the size 0.5, $\sigma = 0.2$ s network for two different decision thresholds, 0.5 and 0.75. We chose these relatively high decision thresholds because we only want to assess pick performance when the model is very confident there is an LFE present. The inset table shows the fraction of picks that are within 10, 30, and 50 samples (0.1, 0.3, and 0.5 s) of the true pick. Picking performance improves with higher decision thresholds, that is, the larger decision threshold results in a larger fraction of picks within a fixed number of samples. Higher decision thresholds also result in fewer overall picks as evidenced by the increased event frequency for a decision threshold of 0.5. After exploring the picking performance for all networks we find there is a tradeoff between σ the total number of identified LFEs (i.e., TP + FP). The $\sigma = 0.4$ s networks generally identify more LFEs but have worse picking performance. Picking performance improves with decreasing σ but this also results in lower recall and fewer LFEs being identified overall. As such, our preferred networks are the size 0.5 or 1 networks with, $\sigma = 0.2$ s. Performance statistics for the 24 networks explored in this study are shown in Table S2.

Figure 4 compares the timing of CNN detections and LFE S-wave arrival times on THIS, the nodal seismometer, and nearby borehole station B079. CNN detections are indicated by colored circles in Figure 4 (where the color indicates the CNN output amplitude); LFE arrival times from the Shelly (2017) catalog are shown as black diamonds. Data from the east and north components of each station are shown in gray. As shown in Figures 4a, 4c, and 4e the CNN has several detections during this time period, many of which coincide with known LFE detections. There are also LFE arrivals that are not detected by the model. Lowering the decision threshold may result in detecting these known arrivals at the expense of increasing the number of false positives. When comparing detections on THIS and B079 to those on 40, shown in Figure 4c, we find that the nodal sensor has many fewer daily detections, which are generally lower amplitude, owing to the fact that, this is a temporary station that has a higher noise floor, different instrument response, and records more environmental signals than the permanent stations used to train the CNN. Some of the CNN detections registered on THIS and B079 are also detected on station 40, suggesting that event association, which would be required to locate LFE sources, could be used to cull false detections. Figures 4b, 4d, and 4f show a time period in which the CNN identifies an LFE on all three stations that is not registered in the catalog of Shelly (2017). We believe this event is an LFE because it is low amplitude, has a dominant frequency content between 4 and 7 hz, typical of LFEs in Parkfield (Thomas et al., 2016), and there is no earthquake in the NCEDC catalog at this time. Additional examples of detections are shown in Figures S3–S5.

4. Discussion

Initially, it was unclear whether the CNN approach to LFE detection would be successful given that the underlying signal in each LFE family is similar (facilitating detection via cross-correlation) and because the signal-to-noise ratio is extremely low; arrivals are not readily identified on individual traces (Figure 1a). Despite these potential limitations, the resulting model has surprisingly good performance. Our preferred networks are the size 0.5 or 1, $\sigma = 0.2$ s networks because they are not overfit, generally have the lowest training and validation loss (Figure 3a), and balance picking performance (Figure 3c) and LFE identification ability (Figure 3b). For a decision threshold of 0.04, this network has an accuracy of 85.7%, a recall of 88.8%, and a precision of 83.6%. The $\sigma = 0.4$ s network has maximum accuracy at a higher decision threshold (0.08) but has worse picking performance.

For the networks we present here, peak accuracy occurs at extremely low decision thresholds. Increasing σ generally increases the network accuracy (Figure 3b) and the decision threshold at which peak accuracy occurs, but results in worse picking performance. Additionally, the picking performance is inferior to PhaseNet, which is unsurprizing given the low signal-to-noise ratios, the emergent onsets of LFEs, and small differences between the true LFE arrival time and the catalog time resulting from making picks on stacked waveforms and data decimation (see Shelly, 2017). If the CNN approach we present here were to be used to

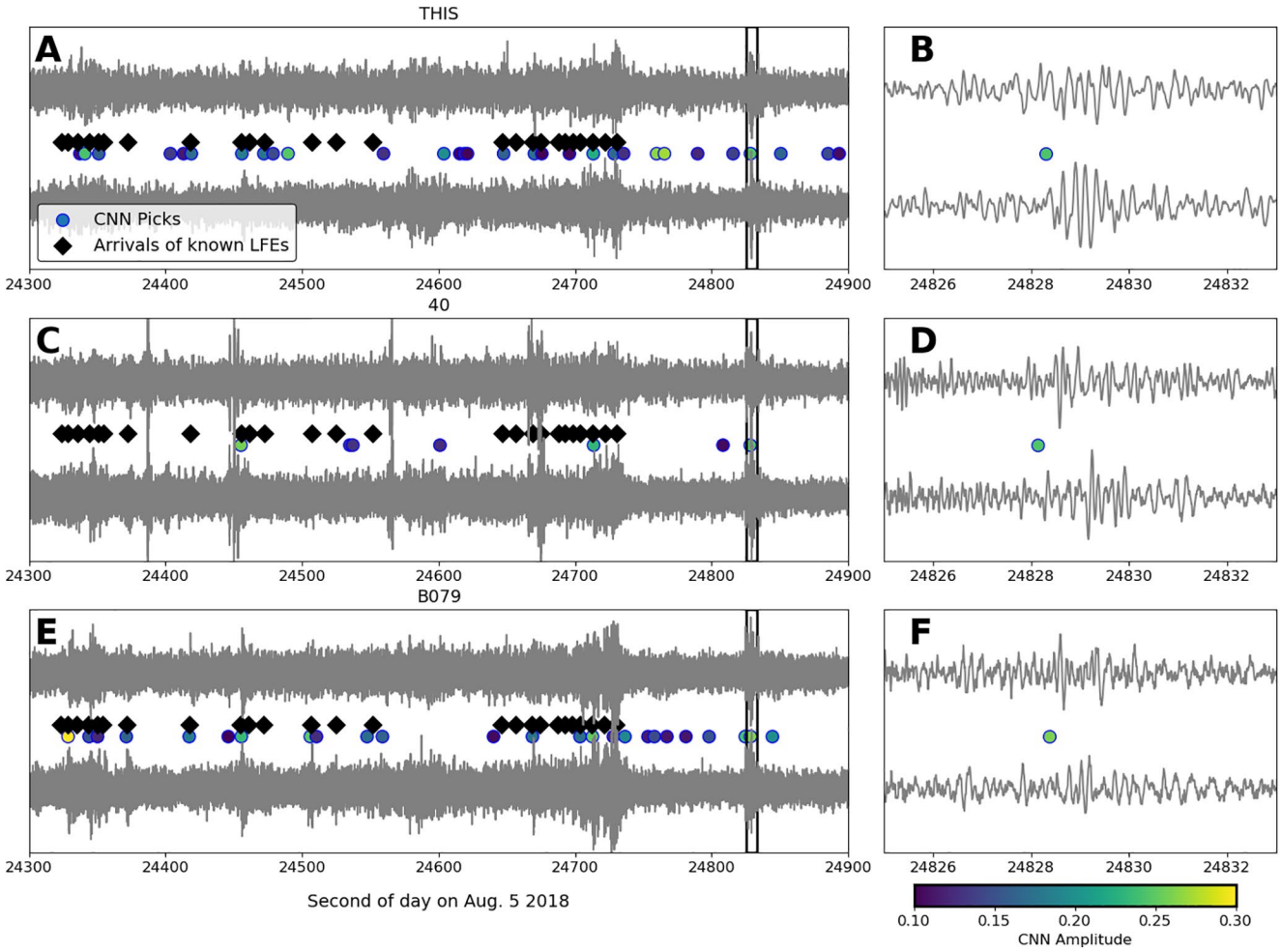

Figure 4. Comparison of convolutional neural network (CNN) (size 0.5, $\sigma = 0.2$ s) detections with S-wave arrival times during a 10-min-long window. Panels (a), (c), and (e) show the east (upper) and north (lower) component recordings on station THIS, the nodal seismometer, and B079 respectively. CNN detections for a decision threshold of 0.1 and above are indicated by circles; their color corresponds to the peak amplitude of the CNN output at the time of the detection. Computed S-wave arrival times from LFEs in the Shelly (2017) catalog are shown as black diamonds representing five different LFE families. Panels (b), (d), and (f) are detailed views (8 s) of the region outlined by the black box in Panels (a), (c), and (e), respectively. An LFE arrival is present that is identified with high confidence by the CNN on THIS and lower confidence on the nodal seismometer, but not included in the Shelly (2017) catalog.

mine continuous seismic data for LFEs, the low decision thresholds required to achieve peak accuracy will likely result in a significant number of false detections. Increasing the decision threshold from the value used in Figure 4 (0.1) would have the effect of reducing the fraction of false positive detections. However, characterizing LFEs, that is, determining locations, focal mechanisms, magnitudes, and other routinely cataloged earthquake properties, necessitates association of detections, which will undoubtedly cull many of the false detections that do not have consistent move out on stations across the network. Figures S3–S5 show examples of nearly simultaneous CNN detections on nearby stations suggesting that associating detections may be a viable way to cull false positive picks as would cross-correlation and/or backprojection.

5. Conclusions

We trained a U-Net to detect LFEs on the Parkfield section of the San Andreas fault using the LFE catalog of Shelly (2017). We show that the resulting network has detection and phase picking capabilities that approach those of PhaseNet (Zhu & Beroza, 2019), despite the extremely low-amplitudes and emergent onsets of LFEs. We demonstrate that the CNN is able to detect LFEs in continuous seismic data recorded on stations that were not used as part of the original training and testing datasets. Additionally, the CNN

performs well on temporary surface stations despite higher amplitude noise and frequency content that is distinct from the regional surface and borehole stations used to train the network.

Machine learning based detection of LFEs, combined with association and network cross-correlation, have the potential to dramatically increase the number of LFEs identified in continuous seismic data. This approach is fast (it takes 42 s to make detections on a full day of data at one station on a 2.3 GHz processor) and has the potential to significantly increase the number of identified and verified LFEs owing to its ability to identify non-repeating LFEs and operate on data with very low signal-to-noise ratio (Figure 4b). Finally, the CNN approach we apply on the SAF could be easily adapted to any region where a sizable LFE catalog has been assembled (e.g., Cascadia [Bostock et al., 2012], Mexico [Frank & Shapiro, 2014], New Zealand [Chamberlain et al., 2014], and Japan [Kato & Nakagawa, 2020]). In the absence of an established LFE catalog, machine learning models trained to detect LFEs in other environments, such as the CNN we develop here, may generalize well such that they are able to identify LFEs in other tectonic settings (Rouet-Leduc et al., 2020).

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

This study utilized data from the the Cholame Nodal Array (A. Thomas, 2018), the Transportable Array (IRIS Transportable Array, 2003), the Plate Boundary Observatory, the Southern California Seismic Network (California Institute of Technology and United States Geological Survey Pasadena, 1926; SCEDC, 2013), the Berkeley Digital Seismic Network (BDSN, 2014), the Northern California Earthquake Data Center (NCEDC, 2014), and the USGS Northern California Network (USGS Menlo Park, 1967). Waveform data, metadata, or data products for this study were accessed through the Northern California Earthquake Data Center, the Southern California Earthquake Data Center, and the IRIS Data Management Center. The authors use Tensorflow (Abadi et al., 2016) and Keras (Chollet, 2015). The trained models are available at https://zenodo.org/record/4777132#.YKbi7WZKheg (A. M. Thomas, 2021). Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Acknowledgments
This work was enabled by the Python programming language (Van Rossum & Drake, 1995) and Obspy package (Krischer et al., 2015). Some figures were made with the Generic Mapping Tools software (Wessel et al., 2019); others were made using the matplotlib software package (Hunter, 2007). This material is based upon work supported by the U.S. Geological Survey under grant No. G18AP00045, the US-Israel Binational Science Foundation grant No. 385760, and National Science Foundation grant No. 1848302. IRIS Data Services are funded through the Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048. The authors thank Clara Yoon and two anonymous reviewers for excellent feedback that improved this work.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., et al. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. In *Proceedings of the 12th USENIX conference on operating systems design and implementation*, (pp. 265–283).

BDSN. (2014). *Berkeley digital seismic network [BK]*. Dataset: UC Berkeley Seismological Laboratory. https://doi.org/10.7932/BDSN

Behr, W. M., & Bürgmann, R. (2021). What's down there? The structures, materials and environment of deep-seated slow slip and tremor. *Philosophical Transactions of the Royal Society A*, *379*, 2193. https://doi.org/10.1098/rsta/379/2193

Bostock, M. G., Royer, A. A., Hearn, E. H., & Peacock, S. M. (2012). Low frequency earthquakes below southern Vancouver Island. *Geochemistry, Geophysics, Geosystems*, *13*, Q11007. https://doi.org/10.1029/2012GC004391

Bostock, M. G., Thomas, A. M., Rubin, A. M., & Christensen, N. I. (2017). On corner frequencies, attenuation, and low-frequency earthquakes. *Journal of Geophysical Research: Solid Earth*, *122*, 543–557. https://doi.org/10.1002/2016JB013405

Bostock, M. G., Thomas, A. M., Savard, G., Chuang, L., & Rubin, A. M. (2015). Magnitudes and moment-duration scaling of low-frequency earthquakes beneath southern Vancouver Island. *Journal of Geophysical Research: Solid Earth*, *120*(9), 6329–6350. https://doi.org/10.1002/2015JB012195

California Institute of Technology and United States Geological Survey Pasadena. (1926). *Southern California seismic network [CI]*. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/CI

Chamberlain, C. J., Shelly, D. R., Townend, J., & Stern, T. A. (2014). Low-frequency earthquakes reveal punctuated slow slip on the deep extent of the Alpine fault, New Zealand. *Geochemistry, Geophysics, Geosystems*, *15*, 2984–2999. https://doi.org/10.1002/2014GC005436

Chollet, F. (2015). *Keras: Deep learning for Python version 2.4.0*. Retrieved from https://github.com/fchollet/keras

Frank, W. B., & Shapiro, N. M. (2014). Automatic detection of low-frequency earthquakes (LFEs) based on a beamformed network response. *Geophysical Journal International*, *197*, 1215–1223. https://doi.org/10.1093/gji/ggu058

Frank, W. B., Shapiro, N. M., Kostoglodov, V., Husker, A. L., Campillo, M., Payero, J. S., & Prieto, G. A. (2013). Low-frequency earthquakes in the Mexican Sweet Spot. *Geophysical Research Letters*, *40*, 2661–2666. https://doi.org/10.1002/grl.50561

Hawthorne, J. C., Thomas, A. M., & Ampuero, J. P. (2019). The rupture extent of low frequency earthquakes near Parkfield, CA. *Geophysical Journal International*, *216*, 621–639. https://doi.org/10.1093/gji/ggy429

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, *9*, 90–95. https://doi.org/10.1109/MCSE.2007.55

IRIS Transportable Array. (2003). *USArray transportable array [TA]*. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/TA

Jozinović, D., Lomax, A., Štajduhar, I., & Michelini, A. (2020). Rapid prediction of earthquake ground shaking intensity using raw waveform data and a convolutional neural network. *Geophysical Journal International*, *222*, 1379–1389. https://doi.org/10.1093/gji/ggaa233

Kato, A., & Nakagawa, S. (2020). Detection of deep low-frequency earthquakes in the Nankai subduction zone over 11 years using a matched filter technique. *Earth Planets and Space*, *72*. https://doi.org/10.1186/s40623-020-01257-4

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *Proceedings of the 3rd international conference for learning representations*, (p. 15).

Klimasewski, A., Sahakian, V., & Thomas, A. (2020). Comparing artificial neural networks with traditional ground-motion models for small-magnitude earthquakes in southern California. *Bulletin of the Seismological Society of America*, *111*, 1577–1589. https://doi.org/10.1785/0120200200

Krischer, L., Megies, T., Barsch, R., Beyreuther, M., Lecocq, T., Caudron, C., & Wassermann, J. (2015). ObsPy: A bridge for seismology into the scientific Python ecosystem. *Computational Science & Discovery*, *8*, 014003. https://doi.org/10.1088/1749-4699/8/1/014003

Littel, G. F., Thomas, A. M., & Baltay, A. S. (2018). Using tectonic tremor to constrain seismic wave attenuation in Cascadia. *Geophysical Research Letters*, *45*, 9579–9587. https://doi.org/10.1029/2018GL079344

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer: An attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*, 3592. https://doi.org/10.1038/s41467-020-17591-w

NCEDC. (2014). *Northern California earthquake data center*. Dataset: UC Berkeley Seismological Laboratory. https://doi.org/10.7932/NCEDC

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*, e1700578. https://doi.org/10.1126/sciadv.1700578

Poiata, N., Vilotte, J. P., Bernard, P., Satriano, C., & Obara, K. (2018). Imaging different components of a tectonic tremor sequence in southwestern Japan using an automatic statistical detection and location method. *Geophysical Journal International*, *213*, 2193–2213. https://doi.org/10.1093/gji/ggy070

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Proceedings of international conference on medical image computing and computer-assisted intervention*, (pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28

Ross, Z. E., Meier, M. A., Hauksson, E., & Heaton, T. H. (2018). Generalized seismic phase detection with deep learning. *Bulletin of the Seismological Society of America*, *108*, 2894–2901. https://doi.org/10.1785/0120180080

Ross, Z. E., Yue, Y., Meier, M. A., Hauksson, E., & Heaton, T. H. (2019). PhaseLink: A deep learning approach to seismic phase association. *Journal of Geophysical Research: Solid Earth*, *124*, 856–869. https://doi.org/10.1029/2018JB016674

Rouet-Leduc, B., Hulbert, C., McBrearty, I. W., & Johnson, P. A. (2020). Probing slow earthquakes with deep learning. *Geophysical Research Letters*, *47*, e2019GL085870. https://doi.org/10.1029/2019GL085870

Royer, A. A., & Bostock, M. G. (2014). A comparative study of low frequency earthquake templates in northern Cascadia. *Earth and Planetary Science Letters*, *402*, 247–256. https://doi.org/10.1016/j.epsl.2013.08.040

Rubin, A. M., & Armbruster, J. G. (2013). Imaging slow slip fronts in Cascadia with high precision cross-station tremor locations. *Geochemistry, Geophysics, Geosystems*, *14*, 5371–5392. https://doi.org/10.1002/2013GC005031

SCEDC. (2013). *Southern California earthquake center*. Dataset: California Institute of Technology. https://doi.org/10.7909/C3WD3xH1

Shelly, D. R. (2017). A 15 year catalog of more than 1 million low-frequency earthquakes: Tracking tremor and slip along the deep San Andreas Fault. *Journal of Geophysical Research: Solid Earth*, *122*, 3739–3753. https://doi.org/10.1002/2017JB014047

Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, *446*, 305–307. https://doi.org/10.1038/nature05666

Shelly, D. R., & Hardebeck, J. L. (2010). Precise tremor source locations and amplitude variations along the lower-crustal central San Andreas Fault. *Geophysical Research Letters*, *37*, L14301. https://doi.org/10.1029/2010GL043672

Sweet, J. R., Creager, K. C., & Houston, H. (2014). A family of repeating low-frequency earthquakes at the downdip edge of tremor and slip. *Geochemistry, Geophysics, Geosystems*, *15*, 3713–3721. https://doi.org/10.1002/2014GC005449

Tang, C. C., Peng, Z., Chao, K., Chen, C. H., & Lin, C. H. (2010). Detecting low-frequency earthquakes within non-volcanic tremor in southern Taiwan triggered by the 2005 Mw8.6 Nias earthquake. *Geophysical Research Letters*, *37*, L16307. https://doi.org/10.1029/2010GL043918

Thomas, A. M. (2021). *amtseismo/Parkfield-LFE-CNN: Parkfield-LFE-CNNv1.0 (Version v1.0)*. Zenodo. https://doi.org/10.5281/zenodo.4777132

Thomas, A. M., Beeler, N. M., Bletery, Q., Burgmann, R., & Shelly, D. R. (2018). Using low-frequency earthquake families on the San Andreas Fault as deep creepmeters. *Journal of Geophysical Research: Solid Earth*, *123*, 457–475. https://doi.org/10.1002/2017JB014404

Thomas, A. M., Beroza, G. C., & Shelly, D. R. (2016). Constraints on the source parameters of low-frequency earthquakes on the San Andreas Fault. *Geophysical Research Letters*, *43*, 1464–1471. https://doi.org/10.1002/2015GL067173

Thomas, A. M., & Bostock, M. G. (2015). Identifying low-frequency earthquakes in central Cascadia using cross-station correlation. *Tectonophysics*, *658*, 111–116. https://doi.org/10.1016/j.tecto.2015.07.013

Thomas, A. M., Inbal, A., & Newton, T. (2018). *Cholame nodal array 2018 (PH5) [1B_2018]*. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/1B_2018

USGS Menlo Park. (1967). *USGS Northern California network [NC]*. International Federation of Digital Seismograph Networks. https://doi.org/10.7914/SN/NC

Van Rossum, G., & Drake, F. L., Jr. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., & Woelfel, I. (2021). EasyQuake: Putting machine learning to work for your regional seismic network or local earthquake study. *Seismological Research Letters*, *92*, 555–563. https://doi.org/10.1785/0220200226

Wessel, P., Luis, J. F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W. H. F., & Tian, D. (2019). The generic mapping tools version 6. *Geochemistry, Geophysics, Geosystems*, *20*, 5556–5564. https://doi.org/10.1029/2019GC008515

Yeck, W. L., Patton, J. M., Ross, Z. E., Hayes, G. P., Guy, M. R., Ambruz, N. B., et al. (2021). Leveraging deep learning in global 24/7 real-time earthquake monitoring at the National Earthquake Information Center. *Seismological Research Letters*, *92*, 469–480. https://doi.org/10.1785/0220200178

Yoon, C. (2021). A high-resolution view of the 2020 Puerto Rico earthquake sequence with machine learning. *Seismological Research Letters*, *92*, 1261. https://doi.org/10.1785/0220210025

Zhu, W., & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*, 261–273. https://doi.org/10.1093/gji/ggy423