Detecting Low-frequency Earthquakes with a Deep Convolutional Neural Network

Principal Investigator: Gregory C. Beroza

Graduate student: Fantine Huot

Department of Geophysics, Stanford University

Abstract. Tectonic tremor is composed of many repeating low-frequency earthquakes (LFEs, LFE analysis provides valuable insights into slow slip processes. LFEs have seismic amplitudes slightly above background noise, making them difficult to detect. In most locations, they are detected using prior waveform templates. We use a 15-year catalog of more than 1 million LFEs along the San Andreas Fault (Shelly, 2017), to train a convolutional neural network (CNN) for LFE detection. We detect new LFEs with low signal amplitude, even below the noise level, without a prior template, with an accuracy of 95.9%. We cluster the detections by waveform similarity to characterize new templates. Our method is scalable and computationally efficient.

Introduction. Tectonic tremor is a low-amplitude, extended-duration seismic signal on the deep extension of some major faults. It was discovered in SW Japan (*Obara*, 2002) and later identified in the Cascadia subduction zone (*Rogers & Dragert*, 2003) and along the San Andreas Fault (*Nadeau & Dolenc*, 2005). Tremor occurs as a swarm of small low-frequency earthquakes (LFEs) (*Shelly et al*., 2007) that lack high-frequency content compared to earthquakes of similar magnitude. LFEs have emergent onsets and seismic amplitudes little above background noise levels, making them difficult to detect reliably. Their arrivals are not always visible in individual traces and only become apparent in waveform stacks. Therefore, in most locations, they are detected only by targeted analysis – usually through network cross-correlation and template matching (*Shelly et al*., 2007); however, these methods require an initial LFE waveform template.

Creating initial LFE waveform templates is labor-intensive. As LFEs tend to repeat over time, events assigned to the same source location are considered to belong to the same LFE family. Each family has a similar seismic response, for which it is possible to create a waveform template. Network cross-correlation relies on taking a candidate waveform and cross-correlating it through continuous data. The waveforms are then stacked to create a template. This process is repeated multiple times until the template achieves a higher signal-to-noise ratio. Since the template candidates are initially derived by visually identifying single LFE candidates from the data, templates only exist for sources that at least occasionally generate tremor above the noise level.

An attractive alternative to this approach is the use of supervised learning. Neural networks have been shown to be successful at earthquake detection and phase picking. *Thomas et al.* (2021) use a CNN for LFE detection along the San Andreas fault, training it on over a 88,000 LFEs and achieving 85% accuracy. We take this approach further – using the LFE catalog compiled by Shelly (2017) – to create a labeled data set of LFEs and background noise waveforms containing over 685,000 LFEs, with over 50 million unique seismic traces across stations. We train a CNN and optimize its architecture and its training hyperparameters jointly by Bayesian optimization, resulting in a model that achieves over 95% accuracy at detecting LFE families not present in the training data. We apply the CNN to continuous data and validate the results by clustering detections by waveform similarity. We identify uncataloged LFE clusters and use the clustered LFEs to create new LFE family templates.

Data Processing. We use a 15-year catalog of more than 1 million LFEs established by Shelly (2017). This catalog contains 1,045,627 events belonging to 88 different LFE families recorded between April 2001 and September 2016. The locations of these LFEs spread over ~ 150 km along the San Andreas Fault, and range in depth from 16 to 30 km. While the catalog contains 88 LFE families, each with a seismic waveform signature, each waveform is unique due to varying noise levels present in each seismic trace. For all these LFEs, we extract the corresponding seismic data from the High-Resolution Seismic Network (HRSN). We use data from 12 3-component stations, resulting in 36 seismic traces.

To train the model, we create a labeled data set of LFE and background noise waveforms. For each LFE entry in the catalog, we pull the corresponding waveforms from the HRSN database. In addition to the LFE data, we extract an equal number of data windows with background noise to provide negative examples for supervised learning. We select background noise at random times over the same time span as the LFE catalog. To prevent noise samples from containing LFEs, we select them from periods at least 5 minutes from any LFE in the catalog. While the presence of uncataloged LFEs inside the background noise windows cannot be totally excluded, this approach ensures that the selected background noise windows have low probability of containing LFEs (Shelly, 2017). The HRSN has gone through many alterations. Some stations failed, were replaced, or removed. The noise and gain levels vary significantly over the years. Some of these modifications were documented, but many were not. To avoid missing data, we only keep data windows for which at least 8 of the 12 stations were operational.

We split the dataset into a training, evaluating, and testing data set. To ensure that we do not introduce prior knowledge of LFE family templates, we separate the data by LFE family. Of the 88 families from the LFE catalog, we dedicate 69 of the families to the training set, 10 to the evaluation set, and the 9 remaining are kept aside for the test set. The split between LFE families is random and the resulting training, evaluation, and test sets each contain various examples of both low and high template cross-correlation coefficient in the catalog. We also separate the data windows in time, taking only data between 2001 and 2014 for training, from 2015 for evaluation, and 2016 for testing. The dataset contains over 50 million seismic traces across different stations.

Machine learning model. We frame LFE detection as a binary classification between windows containing LFEs and those that do not. We implement a 1D CNN for this task, performing the convolutions along the time-axis of the data, treating the 36 traces as input channels for the CNN. We explore different network architectures by hyperparameter tuning using the methodology described by Huot (2022); Huot and Biondi (2020); Huot et al. (2021). This allows us to explore different networks while following heuristics from state-of-the-art classifiers. The hyperparameters related to network architecture are the number of convolution blocks, the number of filters, the type of activation (Ramachandran et al., 2017), the type of downsampling, and the dropout rate (Srivastava et al., 2014). The hyperparameters related to training are batch size and learning rate. We optimize the architecture and training hyperparameters by Bayesian optimization (Snoek et al., 2012). We use the loss computed over the evaluation dataset as the performance metric for tuning. The model is trained with an Adam optimizer (Kingma & Ba, 2014) over 500 passes through the dataset and stops if there is no improvement in the loss after 50 epochs. Our best-performing network is a VGG-type network (Simonyan & Zisserman, 2015) with 5 convolutional blocks with filters [64, 128, 256, 512, 512]. The convolution layers have a kernel of size 3 and a stride of 1, while strided convolutions perform the downsampling with a kernel of size 3 and a stride of 2. It uses Leaky ReLU (Ramachandran et al., 2017) as activation, batch normalization, and has a dropout rate of 20%. There is no regularization of the weights.

Results. The results (Table 1) on the test data demonstrate the trained CNN performs remarkably well at detecting LFEs. IT achieves over 95% accuracy, yet none of the LFE families in the test data are present in the training or evaluation data. It has a precision of 98%, which is important

for detection over continuous data, where slight differences in precision can translate to thousands of false detections. This demonstrates the CNN's ability to detect new LFE families without a prior waveform template. The result also confirms the CNN's performance in low signal-to-noise settings since most LFEs in the test data are under the background noise level.

Table 1. Classification metrics for LFE detection computed against our benchmark test dataset. The trained CNN achieves 95.91% accuracy. The first row shows the metrics obtained with the best-performing model obtained by hyperparameter tuning. The second row reports the mean and standard error for each of these metrics obtained by re-training and re-evaluating this model five times.

	Accuracy	AUC	Precision	Recall
Benchmark test	95.91%	98.94%	98.02%	94.40%
Mean (Standard error)	95.79% (0.26)	98.80% (0.05)	97.81% (0.23)	93.68% (0.31)

We cluster the detections by waveform similarity using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al., 1996), which groups closely packed examples (with many nearby neighbors), and marks as outliers examples that lie in low-density regions. Ignoring isolated examples makes the DBSCAN algorithm robust to noise. We use 1−C as the distance metric for clustering, where C is the normalized cross-correlation coefficient between two waveforms. We set the minimum number of elements per cluster to 5. The resulting clusters are shown projected onto the principal components of the normalized cross-correlation coefficients between all waveforms in Figure 1. The clustering is performed without prior knowledge of LFE families, but we find that distinct clusters correspond to distinct LFE families. While some clusters visually overlap in the 2-dimensional projection, each computed cluster only contains waveforms belonging to one LFE family. This demonstrates that clustering by waveform similarity can automatically group the CNN detections by LFE family.

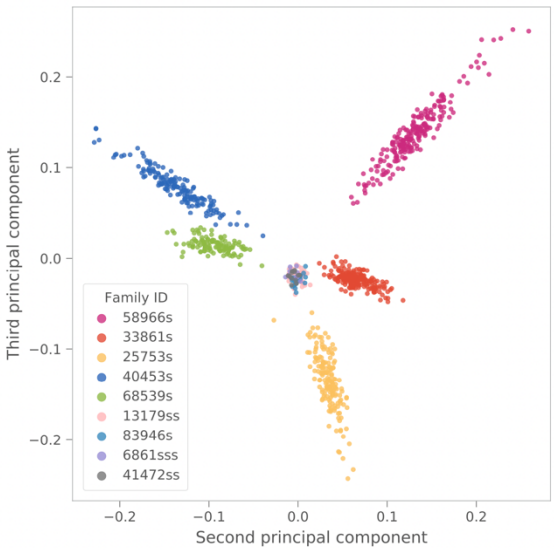

Figure 1. *Clusters obtained by DBSCAN on the test dataset projected on the principal components of the computed waveform similarity. Each LFE family is plotted in a different color. We note that despite the clustering being performed without prior knowledge of LFE families, the obtained clusters correspond to distinct LFE families. The second and third principal components were selected for visualization purposes. Some clusters visually overlap in this 2-dimensional projection.*

We apply the trained CNN to a month of continuous data, February 2016. After consolidating duplicate detections, our CNN detects 10,821 events over February 2016. The LFE catalog contains 4,468 entries over the same month. The CNN detects nearly all the cataloged LFEs; but, it also detects events not in the catalog. The high precision of the trained CNN on the test data suggests that many of these are likely to be LFEs. We apply the clustering to validate these uncataloged detections. Many of the obtained clusters correspond to previously cataloged LFE

families, but 9 clusters of uncataloged events do not. Since these events get clustered according to the same criteria as the cataloged LFEs, the uncataloged events are likely to be LFEs.

To validate whether these correspond to LFEs, we use these stacked waveforms as a template for matched-filter search over the year 2016. We apply the matched-filter search as described in Shelly (2017), using the same detection threshold of eight times the median absolute deviation, and obtain hundreds of detections. This result con- firms that the uncataloged events detected by the CNN and clustered by DBSCAN are actual LFEs. The CNN detection combined with clustering by similarity allows us to generate new candidate LFE family templates.

References

Brown, J. R., Prejean, S. G., Beroza, G. C., Gomberg, J. S., & Haeussler, P. J. (2013). Deep low-frequency earthquakes in tectonic tremor along the alaska- aleutian subduction zone. Journal of Geophysical Research: solid earth, 118(3), 1079–1090.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In kdd (Vol. 96, pp. 226–231).

Huot, F. (2022). Machine learning for seismic event detection. a story in three parts: Earthquakes, microseismic, and tectonic tremors, doctoral dissertation, Stanford University.

Huot, F., & Biondi, B. L. (2020). Detecting earthquakes through telecom fiber using a convolutional neural network. In Seg technical program expanded abstracts 2020 (pp. 3452–3456). Society of Exploration Geophysicists.

Huot, F., Lellouch, A., Given, P., Clapp, R. G., Biondi, B. L., Nemeth, T., & Nihei, K. (2021). Detecting microseismic events on das fiber with super-human accuracy. In Seg/aapg/sepm first international meeting for applied geoscience & energy.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Nadeau, R. M., & Dolenc, D. (2005). Nonvolcanic tremors deep beneath the san an- dreas fault. Science, 307(5708), 389–389.

Obara, K. (2002). Nonvolcanic deep tremor associated with subduction in southwest Japan. Science, 296(5573), 1679–1681.

Ramachandran, P., Zoph, B., & Le, Q. V. (2017). Searching for activation functions. arXiv preprint arXiv:1710.05941.

Rogers, G., & Dragert, H. (2003). Episodic tremor and slip on the Cascadia subduction zone: The chatter of silent slip. Science, 300(5627), 1942–1943.

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(5), 3739–3753.

Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low- frequency earthquake swarms. Nature, 446(7133), 305–307.

Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition.

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In Advances in neural information processing sys- tems (pp. 2951–2959).

Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.

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(13), e2021GL093157.