Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking

Chengping Chai[1], Monica Maceira[2], Hector Santos-Villalobos[1], Singanallur Venkatakrishnan[1], Martin Schoenball[3], Weiqiang Zhu[4], Gregory C. Beroza[4], and Clifford H Thurber[5]

[1]Oak Ridge National Laboratory
[2]Oak Ridge National Laboratory (DOE)
[3]Lawrence Berkeley National Laboratory
[4]Stanford University
[5]University of Wisconsin-Madison

November 23, 2022

Abstract

The important task of tracking seismic activity requires both sensitive detection and accurate earthquake location. Approximate earthquake locations can be estimated promptly and automatically; however, accurate locations depend on precise seismic phase picking, which is a laborious and time-consuming task. We adapted a deep neural network (DNN) phase picker trained on local seismic data to meso-scale hydraulic fracturing experiments. We designed a novel workflow, transfer-learning aided double-difference tomography, to overcome the three orders of magnitude difference in both spatial and temporal scales between our data and data used to train the original DNN. Only 3,500 seismograms (0.45% of the original DNN data) were needed to re-train the original DNN model successfully. The phase picks obtained with transfer-learned model are at least as accurate as the analyst's, and lead to improved event locations. Moreover, the effort required for picking once the DNN is trained is a small fraction of the analyst's.

Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking

Enter authors here: Chengping Chai[1], Monica Maceira[1,2], Hector J. Santos-Villalobos[1], Singanallur V. Venkatakrishnan[1], Martin Schoenball[3], Weiqiang Zhu[4], Gregory C. Beroza[4], Clifford Thurber[5], and EGS Collab Team

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

[2]Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee, USA.

[3]Lawrence Berkeley National Laboratory, Berkeley, California, USA.

[4]Department of Geophysics, Stanford University, Stanford, California, USA.

[5]Department of Geology and Geophysics, University of Wisconsin-Madison, Madison, Wisconsin, USA

Corresponding author: Chengping Chai (chaic@ornl.gov)

Key Points:

- Successful transfer learning of a neural network to seismic data with orders of magnitude difference in spatial and temporal characteristics
- Created novel workflow combining deep learning and double-difference seismic imaging
- New workflow provides better seismic catalog and larger amount of phase picks compared to human analysts.

Abstract

The important task of tracking seismic activity requires both sensitive detection and accurate earthquake location. Approximate earthquake locations can be estimated promptly and automatically; however, accurate locations depend on precise seismic phase picking, which is a laborious and time-consuming task. We adapted a deep neural network (DNN) phase picker trained on local seismic data to meso-scale hydraulic fracturing experiments. We designed a novel workflow, transfer-learning aided double-difference tomography, to overcome the three orders of magnitude difference in both spatial and temporal scales between our data and data used to train the original DNN. Only 3,500 seismograms (0.45% of the original DNN data) were needed to re-train the original DNN model successfully. The phase picks obtained with transfer-learned model are at least as accurate as the analyst's, and lead to improved event locations. Moreover, the effort required for picking once the DNN is trained is a small fraction of the analyst's.

Plain Language Summary

Seismic sensors are widely used to monitor many energy-related systems. To monitor these systems effectively, we need to process a very large amount of data, which is very labor-intensive. A few deep learning models have been developed to perform these tasks for earthquake generated signals. We adopted one of these deep learning models developed for kilometer scale and updated it for signals recorded from a meter-scale project. This process not only allows us to overcome the significant spatial and temporal scale difference between our data and the data used by the original deep learning model, but also significantly reduces the amount of required training data. Our results show the updated model matches human performance, but with a much faster speed. A workflow that combines the deep learning algorithm with existing imaging technologies enables improvements for both monitoring small earthquakes and studying subsurface structure.

1 Introduction

Seismic monitoring plays a significant role in the oil and gas industry, underground mines, carbon capture and storage, and the geothermal industry due to its value for both reservoir management and for risk mitigation. Valuable information, such as fracture development and elastic properties of the subsurface, can be recovered from data recorded with seismic monitoring systems. For example, the spatial dimensions and temporal evolution of hydraulic and/or reactivated natural fractures are usually estimated by tracking seismic events. The location and origin time of these microseismic events are determined by arrival times of seismic phases at multiple seismic sensors. These same arrival times of primary (P) and secondary (S) waves are also used for subsurface seismic imaging to measure elastic properties of the subsurface. Manually picking arrival times of seismic phases is a very time-consuming task especially for small-scale projects since high temporal sampling rate is required. Therefore, reliable automatic phase pickers are essential for these projects. Traditional automatic pickers such as Short-Term Average/Long-Term Average (STA/LTA; Allen, 1978) and Auto Regression-Akaike Information Criterion (AR-AIC; Sleeman & van Eck, 1999) pickers require intensive human involvement and refinement, and they do not benefit from knowledge of previous picks because they treat each measurement individually. When applied to seismic data, the accuracy of traditional automatic pickers may not be satisfactory, particular for noisy data. Recent applications of deep-learning-based automatic seismic phase pickers (e.g. Y. Chen, 2018; Pardo et al., 2019; Ross et al., 2018; Zhou et al., 2019; L. Zhu et al., 2019; W. Zhu & Beroza, 2018) have shown remarkable accuracy and processing speed for seismic signals originated from natural earthquakes; however, whether these deep-learning phase pickers can be used for seismic monitoring remains unclear, and training such deep-learning phase pickers from scratch requires a huge amount of data.

We use seismic data from experiment 1 of the enhanced geothermal system (EGS) Collab project to test whether one of the deep-learning-based automatic phase pickers, PhaseNet (W. Zhu & Beroza, 2018), is useful for mesoscale monitoring systems. The experiment was conducted at the 4850-foot level of the Sanford Underground Research Facility (SURF) located in Lead, South Dakota (Kneafsey et al., 2019). The testbed consists of one injection, one production, and six 60 m-long monitoring boreholes. The seismic monitoring system was equipped with multiple types of geophysical instruments including 24 hydrophones and 12 accelerometers. An 8-core workstation with an automated processing flow was deployed at the experiment site. The processing scripts are capable of detecting seismic events (triggered), finding initial P-wave phase picks, and inverting for initial seismic event locations and origin times. Seismic event locations and origin times were improved with human reviewed and refined phase picks. The original seismic catalog was processed by Schoenball et al. (2020). Several hydraulic stimulations were performed since May 2018. We focused on seismic signals associated with stimulations between May 22nd and December 21st of 2018. These seismic signals have three orders of magnitude difference in spatial and temporal scales from the original training data used by the deep learning models mentioned earlier.

In this paper, we directly applied the PhaseNet model (W. Zhu & Beroza, 2018) to the seismic data from experiment 1 of the EGS Collab project. Although the results are reasonable, we show that retraining the PhaseNet model significantly boosts performance. The process, called transfer learning (TL), requires only a few thousand seismograms because the weights of the DNN were trained initially by a different dataset of 0.7 million seismograms (from natural earthquakes). The performance of the resulting TL model was compared with a traditional automatic picker, the original PhaseNet model, and human analysts. We then applied the TL model to all the seismograms from the triggered seismic events. We used the resulting TL-derived phase picks and double-difference tomography (tomoDD; Zhang & Thurber, 2003, 2006) to constrain subsurface seismic velocities and update seismic event locations. The results were compared with those using manual picks.

2 Data

Our data consist of seismograms from triggered microseismic events between May 2018 and December 2018 at the Experiment 1 site of the EGS Collab project, manually picked P-wave and S-wave arrival times, and the original seismic catalog from (Schoenball et al., 2020). We used 35 seismic sensors (one hydrophone was defective) with a 100 kHz sampling rate that were deployed in six 60m-long monitoring wells (Figure S1). We detected and located the microseismic events using a standard STA/LTA routine, the PhasePAPy package (C. Chen & Holland, 2016), and a modified version of Hypoinverse (Klein, 2002). We cut the triggered seismograms to 0.11 s long segments around the P-wave arrival times and filtered with a bandpass filter between 3 kHz and 20 kHz. The filtered seismograms show clear similarity with those used to train the original PhaseNet (Figure S2). We used a total of 69,444 waveform segments (from 1,932 seismic events). Initial P-wave arrival times were automatically measured with the PhasePAPy package. We manually reviewed and refined the P-wave arrival times. We picked all S-wave arrival times manually. Additional details about the monitoring system and data preprocessing procedures of the original seismic catalog can be found in (Schoenball et al., 2020).

3 Method

We designed a workflow (Figure 1), TL-aided double-difference tomography (TADT), that takes advantage of two existing technologies, deep neural networks (DNN) and seismic double-difference tomography. We started with the pre-trained DNN model, PhaseNet (W. Zhu & Beroza, 2018), that was trained with over 0.7 million seismic recordings from in and around northern California for natural earthquakes. The PhaseNet model was trained using 30 s long seismograms sampled at 100 Hz. The earthquake-station distance for these data is on the order of 10's of kilometers. Our monitoring system for our data samples at 100 kHz and the source-sensor distance is on the order of 10 meters. Despite the three orders of magnitude differences in both sampling rate and source-sensor distance between the PhaseNet data and our data, we found that PhaseNet produced acceptable results when applied to our data. To improve the performance further, we updated the PhaseNet model with a subset of seismic data that meets the training data requirements (three-component seismograms with both P- and S-wave picks) for PhaseNet. We then applied the resulting TL model to all the triggered seismograms (30 ms long) to obtain TL derived P- and S-wave phase picks. We used the tomoDD package (Zhang & Thurber, 2003, 2006) to update the seismic catalog and simultaneously image the subsurface tomographically. The TADT workflow allows us to reduce the human effort significantly.

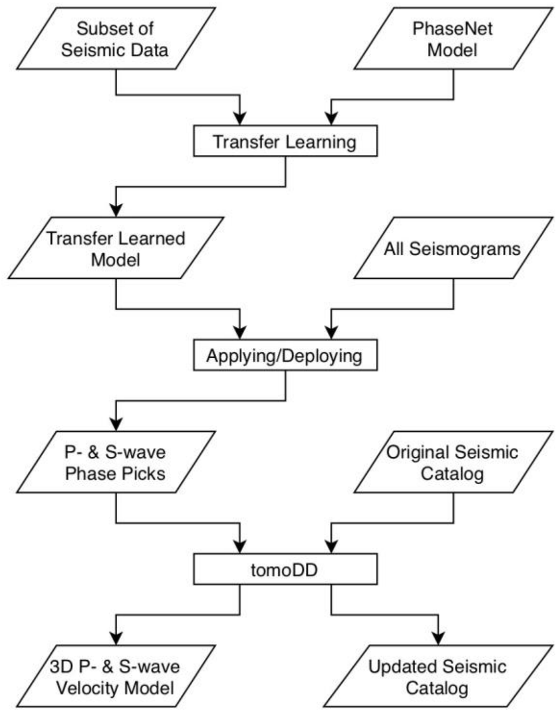

Figure 1. A flowchart of TL-aided seismic tomography using PhaseNet (W. Zhu & Beroza, 2018) and tomoDD (Zhang & Thurber, 2003, 2006).

3.1 Transfer learning

During the TL process, we use the same network architecture as PhaseNet and initialize the weights with the PhaseNet model. We visually inspected the selected seismograms and excluded 343 (9%) incorrect phase picks. The remaining 3,478 seismograms belong to 1,872 distinct seismic events. The total number of seismograms we used is only 0.45% of that used by the original PhaseNet. We randomly divided the seismic events into training, validation, and test sets. The training set (2,443 waveforms) was used to retrain the DNN model, the validation set (345 waveforms) was used to select the optimal model from different training runs, and the test set (690 waveforms) was used to evaluate performance (Figure S3). Similar to W. Zhu & Beroza (2018), we used a Gaussian distribution with a standard derivation of 0.1 ms centered on the manual picks to represent manual pick uncertainty. We allowed the entire neural network to change during TL, and used the Adam optimizer (Kingma & Ba, 2014). We used a learning rate (determines the step size of each iteration) of 0.01 and a batch size (number of training samples used each time) of 20. Our tests indicate using filtered data leads to better performance than raw data. We applied a bandpass filter with corner frequencies of 3 kHz and 20 kHz to the seismograms before feeding them into the neural network for training, which is different from W. Zhu & Beroza (2018). For a fair comparison, the bandpass filter was applied to all the seismograms throughout this study.

3.2 Double-difference tomography

The original seismic catalog was processed with a homogeneous seismic velocity model (Schoenball et al., 2019). Here we used the double-difference tomography package tomoDD (Zhang & Thurber, 2003, 2006) to simultaneously minimize the uncertainty in seismic event locations due to spatial seismic velocity variations and to constrain the 3D subsurface P-wave and S-wave velocity model for the seismically active region. Since the tomoDD package was originally designed for kilometer-scale problems, we made some modifications (e.g. input and output format, coordinate system) specifically for meter-scale projects. We estimated both P- and S-wave seismic velocity models. We relocated 1,743 seismic events and discretized a 3D volume of 77 m (easting), 83 m (northing), and 40 m (vertical) with nodes 1 m apart in each direction (then interpolated to 0.1 m by the tomoDD package). The initial model was homogeneous with a P-wave speed of 5.9 km/s and an S-wave speed of 3.5 km/s. These two velocities were obtained from curve fitting of travel-time observations (travel-time versus distance). Numerous previous studies (e.g. Chai et al., 2019; Syracuse et al., 2016) have shown that appropriate inversion parameters are required for a well-constrained seismic velocity model. We used an L-curve analysis (similar to Hansen, 1992) to find the optimal set of inversion parameters. An optimal weight of 10 was used for smoothing and 200 for damping (see Zhang & Thurber, 2003 for definition). We obtained the final velocity models and updated seismic catalog after eight iterations. The final models fit the observations better than the starting homogeneous model (Figure S4 and S5).

4 Results

Our results are new phase picks, updated seismic event locations, and 3D seismic velocity models. Hyper Text Markup Language (HTML) based interactive visualizations (similar to Chai et al., 2018) were used to inspect seismic event locations and seismic velocity models.

4.1 Phase picks

The TL model we obtained measures phase picks from seismograms with high accuracy. Some randomly selected waveforms and associated phase picks from the test dataset are shown in Figure S6. We can see that the TL results agree with manual picks even when the background noise level is high. Inspecting the data when TL results differ from manual picks, we noticed that the TL model is able to correct some human errors or skip difficult-to-pick signals (more often for P-waves than for S-waves). The difference between TL results and manual picks is just slightly larger than the threshold (0.1 milliseconds), for many cases in Figure S6. The TL model is more prone to error when signals are very complex.

We compared the TL results with those using the Obspy (Beyreuther et al., 2010) implementation of an AR picker (Akazawa, 2004), the original PhaseNet, and human analysts (Figure 2). We use precision, recall, and F1 score (see Text S1 for definition) to quantify and compare the performance. We estimated the performance of human analysts by having three analysts manually pick the phase arrival times from the same 100 three-component seismograms (to reduce time cost). For each seismogram, we considered the median of the three manual picks as the ground truth. We also measured the human performance for each analyst by comparing results from each analyst against the ground truth. The human performance in Figure 2a and 3b was the average of the three analysts. The original PhaseNet produced much better results than the AR picker for both P and S waves. The TL model outperformed the original PhaseNet with an improvement of roughly 0.1 in precision and 0.3 in recall, highlighting the importance of re-training the DNN with our data. The TL model is the only one among the three automatic pickers that has a performance comparable to human analysts. The TL model performed slightly better on S waves than human analysts, which could be due to the larger SNRs compared to P waves. The TL model achieved human performance in a fraction of the time.

We used a 5-fold cross-validation to measure the uncertainties in the performance matrices. We divided all the available waveforms (including training, validation, and test set) into five folds (equal parts). Five TL models were trained using one of the five combinations of four folds for each training and validated with the other fold. The performance of these five TL models was used to compute the uncertainty of the performance. The performance improvement is larger than the measurement uncertainty (Figure 2c). We also trained TL models with different number of training waveforms and tested the TL models with the same test dataset. As expected, F1 scores of the TL models for both P and S waves improve as more training data were included (Figure 2d).

When we apply the TL model to all the triggered seismograms, the TL model finds more S-wave picks than the human expert. We performed 3D double-difference tomography using manual picks and TL derived picks with the same inversion parameters. Although fewer P-wave picks were obtained by the TL model compared to the human analyst, the updated seismic locations using TL-derived picks show a more compact distribution compared to that using manual picks (see the next section for details). Specifically, we found 18,543 acceptable P-wave picks and 8,935 S-wave picks from the human expert using a total of 69,444 seismograms. The TL model identified 12,050 acceptable P-wave picks (20% of which were included in the training dataset) and 13,297 S-wave picks (18% of which were included in training dataset).

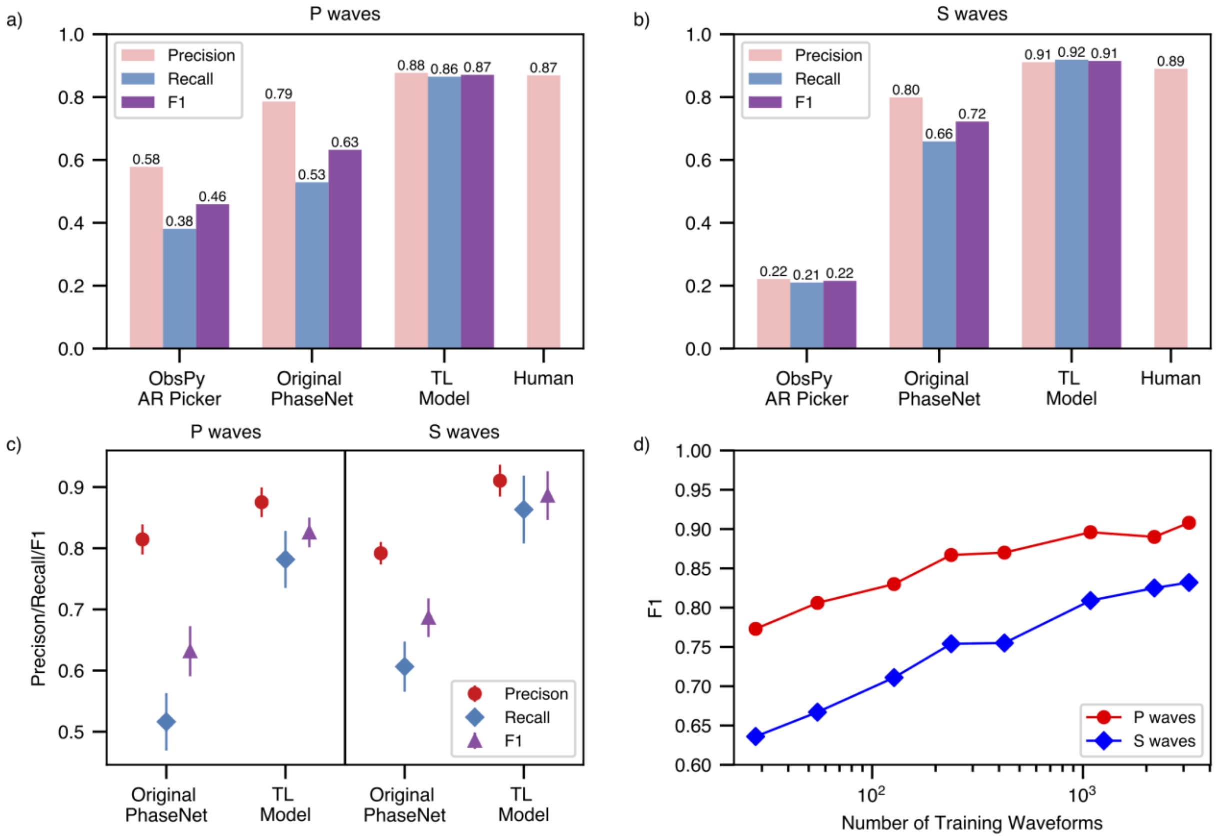

Figure 2. A comparison of performance between human (three analysts), the ObsPy AR picker, the original PhaseNet, and the TL model for (a) P waves and (b) S waves. The human performance was measured with a smaller data set (100 waveforms) due to cost. (c) shows uncertainty (2 sigma) of the performance matrices for the original PhaseNet model and the TL model measured with a 5-fold cross-validation. (d) shows the F1 score as a function of the number of training waveforms used for TL.

4.2 Updated seismic locations

We examine seismic locations associated with the May 2018 stimulations (May 22-25), June 2018 stimulation (June 25), and December 2018 stimulations (December 21 & 22) in Figure 3. Details of these stimulations can be found in Schoenball et al. (2020). Compared to the original seismic locations, the updated locations from double-difference tomography using either manual picks or the TL-derived picks show more detailed geometry of the fractures. For the May 2018 stimulations (Figure 3a-c), the updated seismic locations show two parallel fractures that are not obvious in the original locations. Since these two fractures intercepted one monitoring borehole, we were able to confirm these fractures with independent temperature data recorded in the borehole using distributed temperature sensing with 0.25 m spatial resolution (Fu et al., 2020). Using the TL-derived picks leads to more tightly clustered seismic locations. For the June 2018 stimulation (Figure 3d-f), both the original and updated seismic locations show two fractures. Compared to the original, the updated locations show a slightly tighter pattern delineating the activated fractures. For the December 2018 stimulations (Figure 3g-i), the original locations show two intersecting fractures, but the geometry of these fractures was not well constrained, especially near the two ends of the fractures. When we relocated the seismic event locations using the manual picks, these two fractures showed a tighter pattern. When TL-derived picks were used for relocation, these two fractures were constrained even better and we can see the two ends of the fractures more clearly. As indicated by the updated seismic event locations, the TL-derived picks are equivalent to or better than the manual picks in imaging the activated fractures.

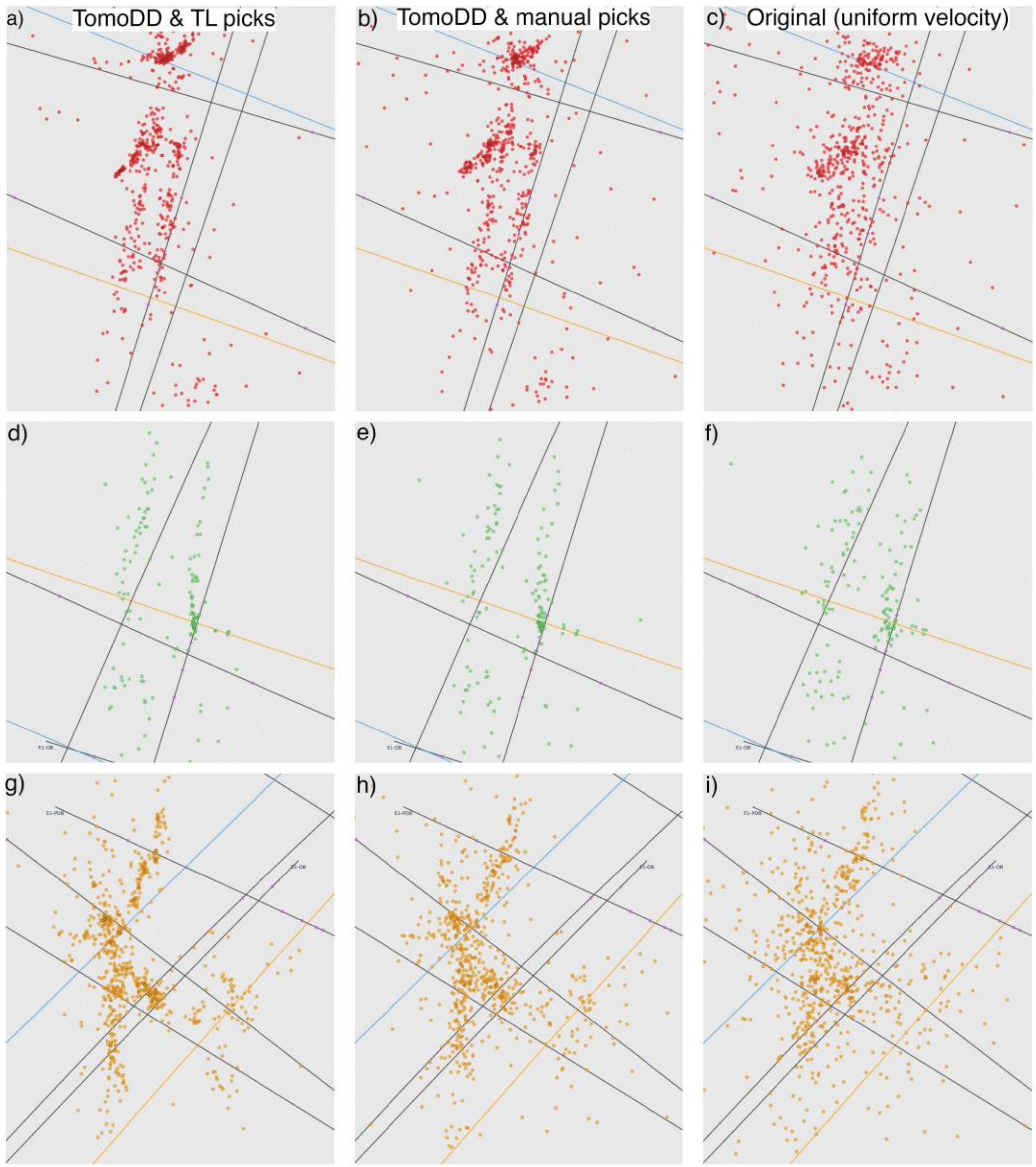

Figure 3. A comparison of original (c, f, and i) and updated (a, b, d, e, g, and h) microseismic event locations associated with stimulations in May 2018 (a, b, and c), June 2018 (d, e, and f), and December 2018 (g, h, and i). Seismic events in the left panel (a, d, and g) show locations updated with TL-derived phase picks using tomoDD. The middle panel (b, e, and h) shows locations updated with manual picks using tomoDD. Fractures are clearly visible with TL-derived phase picks. The lines are boreholes.

4.3 3D seismic velocity model

We constrained both P-wave and S-wave seismic velocities with double-difference tomography. The P-wave velocity model shows significant spatial heterogeneity when either manual picks or TL-derived picks were used in the tomographic inversion. Slices of the 3D P-wave and S-wave velocity models obtained using TL-derived picks are shown in Figure 4. The P-wave velocity model contains some small-scale high-velocity anomalies. To first order, the P-wave velocity is lower at a greater depth. The S-wave velocity model shows a smoother pattern with a low-velocity zone imaged at an elevation below 105 meters. The average P-wave velocity agrees with that obtained from active source surveys (Schoenball et al., 2020).

To identify the volume that we can reliably image, we performed checkerboard tests using data simulated according to both manual picks and TL-derived picks. For the checkerboard tests, we started with an artificial model with alternating high and low velocities. Synthetic P-wave and S-wave phase picks were computed using the artificial model matching the actual observations between seismic event and seismic sensor pairs. The synthetic phase picks were then used in tomography with a uniform starting model. The recovered (inverted) model is compared to the true model to identify the volume that is well-constrained by the data. To measure the volume, we first compute the absolute difference between the recovered model and the true model at each grid cell. A grid cell is considered well-constrained when the recovered seismic velocities are less than 0.1 km/s away for P waves or 0.06 km/s for S waves from the ground truth. The well-constrained volume is smoothed by applying a 3D spatial Gaussian filter with a standard derivation of 1 m in each direction to all of the well-constrained grid cells. Slices of the recovered P- and S-wave velocity models using manual picks and TL-derived picks are shown in Figure S7 and S8. For the P-wave velocity model, the well-constrained volume is 2,678 m^3 for manual picks and 2,465 m^3 (8% decrease) for TL-derived picks. For the S-wave velocity model, the well-constrained volume is 815 m^3 for manual picks and 1,895 m^3 (133% increase) for TL-derived picks.

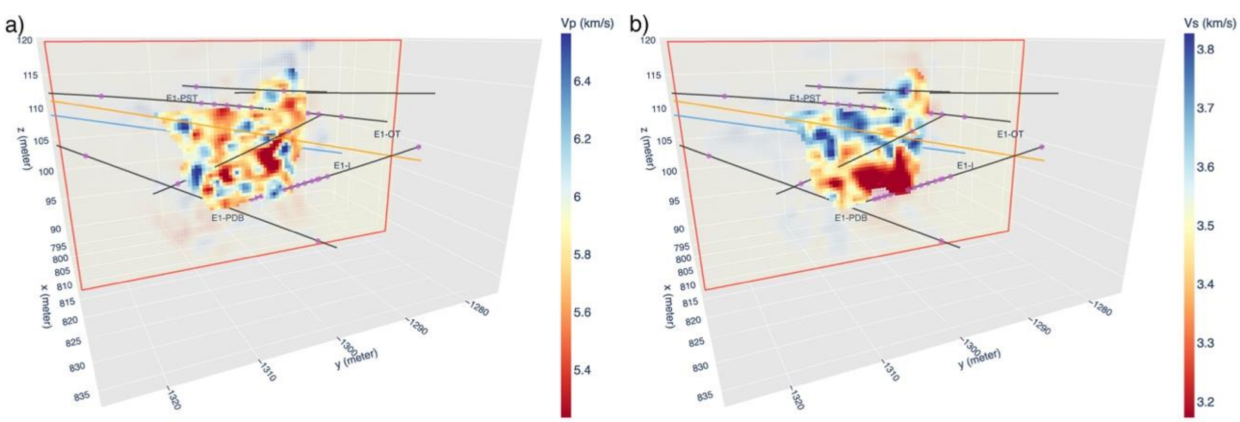

Figure 4. The final (a) P- and (b) S-wave velocity models inverted using TL picks. The lines are boreholes. The dots represent seismic sensors.

5 Discussion and Conclusions

We present a workflow that integrates TL and seismic double-difference tomography. As demonstrated with the EGS Collab data, the workflow can produce better seismic event locations, improve subsurface imaging capabilities, and reduce the overall time cost compared to the original labor-intensive workflow. Our results also show that the TL model obtained by retraining the PhaseNet DNN leads to human-level performance despite the significant differences in the study area size, sensor geometry, and sampling rate between the data used to develop and train PhaseNet and our data. Other types of geophysical observations such as P-wave receiver functions, surface-wave measurements, and gravity observations (e.g. Chai et al., 2015; Maceira & Ammon, 2009; Syracuse et al., 2016) can be inverted together with TL-derived phase picks when data are available.

Since phase picks are the basis for both locating seismic events and imaging the subsurface, it is valuable to determine seismic phase picks quickly and reliably. The PhaseNet model leads to better picks than many traditional auto-pickers such as the ObsPy implementation of the AR picker (W. Zhu & Beroza, 2018). A TL model initialized with the PhaseNet model and retrained with only around 3,600 three-component seismograms and associated manual picks outperforms the original PhaseNet model by over 10% in terms of precision and recall. The TL model performs equally to or slightly better than a human expert. The TL model found fewer (32%) P-wave picks but more (48%) S-wave picks than the human expert. Since the double-difference tomography results that used these TL-derived phase picks show better seismic event locations compared to those using manual picks, it is likely the TL model removed low-quality P-wave picks and added high-quality S-wave picks. The speed of the TL model (or PhaseNet) is about 1,900 times (excluded training time) faster than the human expert. Weights of the TL model show first-order similarities and small differences compared to the PhaseNet model (Figure S9-S11). A comparison (Figure S12 and S13) of convolutional features of hidden layers for an example input using the PhaseNet model and the TL model suggests that these small changes lead to better hidden features (more impulsive peaks).

Double-difference tomography tests using manual picks and TL derived picks show that the latter lead to better seismic event locations and a larger (133% increase) well-constrained volume for the S-wave velocity model. Even though we obtained fewer P-wave picks with the TL model compared to the human expert, the well-constrained volume for the P-wave velocity model only decreased slightly. The improved seismic event locations allow us to see detailed structures of the fracture planes, which in turn will help us better constrain the fracture geometry. Two parallel fracture planes were confirmed with independent borehole observations (Fu et al., 2020).

Our results show that we can reduce the time cost significantly, and improve results, by adding TL into the proposed workflow. Seismic phase picking is labor intensive and thus expensive. It took on the order of several days to determine all the seismic phase picks from the 69,444 seismograms recorded. For the presented workflow, the analyst would only need to manually pick around 3,500 high-quality seismograms. Retraining the PhaseNet model took around one hour (using 32 2.1 GHz Intel Xeon cores). Processing all the seismograms with the TL model took only nine minutes on a laptop computer (with six 2.9 GHz Intel i9 cores). Even including the retraining time, the presented workflow takes much less time than human labor. The speed can be increased with greater computational power. Moreover, the TL model can be directly used on future seismic data from the same recording system without retraining. The proposed workflow is an economical way to monitor subsurface fracture evolution and image subsurface seismic structure with high resolution. The workflow is also applicable to new study areas.

Acknowledgments and Data

This material was based upon work supported by the U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy (EERE), Office of Technology Development, Geothermal Technologies Office, under Award Number DE-AC05-00OR22725. The United States Government retains, and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The research supporting this work took place in whole or in part at the Sanford Underground Research Facility in Lead, South Dakota. The assistance of the Sanford Underground Research Facility and its personnel in providing physical access and general logistical and technical support is acknowledged. We thank Haijiang Zhang for sharing the tomoDD package. We thank Yarom Polsky, Amir Ziabari and Derek Rose for helpful discussions. We acknowledge comments and suggestions from Jacob Hinkle, Jeffrey Johnson, and Philip Bingham. The authors declare that there is no conflict of interest regarding the publication of this article. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Government.

Seismograms and initial microseismic catalog can be downloaded at https://gdr.openei.org/submissions/1166 (last accessed in February 2020). Final microseismic catalog, seismic velocity models, and associated visualizations are available at http://gdr.openei.org/submissions/1214 (last accessed in April 2020). Matplotlib (Hunter, 2007) and plotly (https://plot.ly, last accessed in March 2020) were used to generate figures. Some calculations were performed with Numpy (van der Walt et al., 2011). The original PhaseNet model can be accessed at https://github.com/wayneweiqiang/PhaseNet (las accessed in March 2020).

References

Akazawa, T. (2004). A technique for automatic detection of onset time of P-and S-Phases in strong motion records. In *13th World Conference on Earthquake Engineering* (p. 786).

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, *68*(5), 1521–1532.

Beyreuther, M., Barsch, R., Krischer, L., Megies, T., Behr, Y., & Wassermann, J. (2010). ObsPy: A Python Toolbox for Seismology. *Seismological Research Letters*, *81*(3), 530–533. https://doi.org/10.1785/gssrl.81.3.530

Chai, C., Ammon, C. J., Maceira, M., & Herrmann, R. B. (2015). Inverting interpolated receiver functions with surface wave dispersion and gravity: Application to the western U.S. and adjacent Canada and Mexico. *Geophysical Research Letters*, *42*(11), 4359–4366. https://doi.org/10.1002/2015GL063733

Chai, C., Ammon, C. J., Maceira, M., & Herrmann, R. B. (2018). Interactive Visualization of Complex Seismic Data and Models Using Bokeh. *Seismological Research Letters*, *89*(2A), 668–676. https://doi.org/10.1785/0220170132

Chai, C., Maceira, M., Santos-Villalobos, H., & Team, E. C. (2019). Subsurface Seismic Structure Around the Sanford Underground Research Facility. In *44th Workshop on*

Geothermal Reservoir Engineering (pp. 1364–1376).

Chen, C., & Holland, A. A. (2016). PhasePApy: A Robust Pure Python Package for Automatic Identification of Seismic Phases. *Seismological Research Letters*, *87*(6), 1384–1396. https://doi.org/10.1785/0220160019

Chen, Y. (2018). Automatic microseismic event picking via unsupervised machine learning. *Geophysical Journal International*, *212*(1), 88–102. https://doi.org/10.1093/gji/ggx420

Fu, P., Wu, H., Ju, X., & Morris, J. (2020). Analyzing Fracture Flow Channel Area in EGS Collab Experiment 1 Testbed. In *45th Workshop on Geothermal Reservoir Engineering* (pp. 1283–1292).

Hansen, P. C. (1992). Analysis of Discrete Ill-Posed Problems by Means of the L-Curve. *SIAM Review*, *34*(4), 561–580. https://doi.org/10.1137/1034115

Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Kingma, D. P., & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*. Retrieved from http://arxiv.org/abs/1412.6980

Klein, F. W. (2002). User's Guide to HYPOINVERSE-2000, a Fortran Program to Solve for Earthquake Locations and Magnitudes. *U.S. Geol. Surv. Open File Report 02-171*. https://doi.org/http://geopubs.wr.usgs.gov/open-file/of02-171/

Kneafsey, T. J., Blankenship, D., Knox, H. A., Johnson, T. C., Ajo-Franklin, J. B., Schwering, P. C., et al. (2019). EGS Collab Project: Status and Progress. In *44th Workshop on Geothermal Reservoir Engineering* (pp. 1335–1350).

Maceira, M., & Ammon, C. J. (2009). Joint inversion of surface wave velocity and gravity observations and its application to central Asian basins shear velocity structure. *Journal of Geophysical Research*, *114*(B2), B02314. https://doi.org/10.1029/2007JB005157

Pardo, E., Garfias, C., & Malpica, N. (2019). Seismic Phase Picking Using Convolutional Networks. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(9), 7086–7092. https://doi.org/10.1109/TGRS.2019.2911402

Ross, Z. E., Meier, M.-A., & Hauksson, E. (2018). P Wave Arrival Picking and First-Motion Polarity Determination With Deep Learning. *Journal of Geophysical Research: Solid Earth*, *123*(6), 5120–5129. https://doi.org/10.1029/2017JB015251

Schoenball, M., Ajo-Franklin, J., Blankenship, D., Cook, P., Dobson, P., Guglielmi, Y., et al. (2019). Microseismic monitoring of meso-scale stimulations for the DOE EGS Collab project at the Sanford Underground Research Facility. In *44th Workshop on Geothermal Reservoir Engineering* (pp. 1392–1399).

Schoenball, M., Ajo-franklin, J. B., Blankenship, D., Chai, C., Dobson, P., Kneafsey, T., et al. (2020). Creation of a mixed-mode fracture network at meso-scale through hydraulic fracturing and shear stimulation. https://doi.org/10.1002/essoar.10502549.1

Sleeman, R., & van Eck, T. (1999). Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. *Physics of the Earth* *and Planetary Interiors*, *113*(1–4), 265–275. https://doi.org/10.1016/S0031-9201(99)00007-2

Syracuse, E. M., Maceira, M., Prieto, G. A., Zhang, H., & Ammon, C. J. (2016). Multiple plates subducting beneath Colombia, as illuminated by seismicity and velocity from the joint inversion of seismic and gravity data. *Earth and Planetary Science Letters*, *444*, 139–149. https://doi.org/10.1016/j.epsl.2016.03.050

van der Walt, S., Colbert, S. C., & Varoquaux, G. (2011). The NumPy Array: A Structure for Efficient Numerical Computation. *Computing in Science & Engineering*, *13*(2), 22–30. https://doi.org/10.1109/MCSE.2011.37

Zhang, H., & Thurber, C. (2003). Double-Difference Tomography: The Method and Its Application to the Hayward Fault, California. *Bulletin of the Seismological Society of America*, *93*(5), 1875–1889. https://doi.org/10.1785/0120020190

Zhang, H., & Thurber, C. (2006). Development and Applications of Double-difference Seismic Tomography. *Pure and Applied Geophysics*, *163*(2–3), 373–403. https://doi.org/10.1007/s00024-005-0021-y

Zhou, Y., Yue, H., Zhou, S., & Kong, Q. (2019). Hybrid event detection and phase-picking algorithm using convolutional and recurrent neural networks. *Seismological Research Letters*, *90*(3), 1079–1087. https://doi.org/10.1785/0220180319

Zhu, L., Peng, Z., McClellan, J., Li, C., Yao, D., Li, Z., & Fang, L. (2019). Deep learning for seismic phase detection and picking in the aftershock zone of 2008 M7.9 Wenchuan Earthquake. *Physics of the Earth and Planetary Interiors*, *293*(May 2018), 106261. https://doi.org/10.1016/j.pepi.2019.05.004

Zhu, W., & Beroza, G. C. (2018). PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. *Geophysical Journal International*, *216*(1), 261–273. https://doi.org/10.1093/gji/ggy423

Appendix

EGS Collab Team includes J. Ajo-Franklin, T. Baumgartner, K. Beckers, D. Blankenship, A. Bonneville, L. Boyd, S. Brown, J.A. Burghardt, C. Chai, Y. Chen, B. Chi, K. Condon, P.J. Cook, D. Crandall, P.F. Dobson, T. Doe, C.A. Doughty, D. Elsworth, J. Feldman, Z. Feng, A. Foris, L.P. Frash, Z. Frone, P. Fu, K. Gao, A. Ghassemi, Y. Guglielmi, B. Haimson, A. Hawkins, J. Heise, C. Hopp, M. Horn, R.N. Horne, J. Horner, M. Hu, H. Huang, L. Huang, K.J. Im, M. Ingraham, E. Jafarov, R.S. Jayne, S.E. Johnson, T.C. Johnson, B. Johnston, K. Kim, D.K. King, T. Kneafsey, H. Knox, J. Knox, D. Kumar, M. Lee, K. Li, Z. Li, M. Maceira, P. Mackey, N. Makedonska, E. Mattson, M.W. McClure, J. McLennan, C. Medler, R.J. Mellors, E. Metcalfe, J. Moore, C.E. Morency, J.P. Morris, T. Myers, S. Nakagawa, G. Neupane, G. Newman, A. Nieto, C.M. Oldenburg, T. Paronish, R. Pawar, P. Petrov, B. Pietzyk, R. Podgorney, Y. Polsky, J. Pope, S. Porse, J.C. Primo, C. Reimers, B.Q. Roberts, M. Robertson, W. Roggenthen, J. Rutqvist, D. Rynders, M. Schoenball, P. Schwering, V. Sesetty, C.S. Sherman, A. Singh, M.M. Smith, H. Sone, E.L. Sonnenthal, F.A. Soom, P. Sprinkle, C.E. Strickland, J. Su, D. Templeton, J.N. Thomle, V.R. Tribaldos, C. Ulrich, N. Uzunlar, A. Vachaparampil, C.A. Valladao, W. Vandermeer, G. Vandine, D. Vardiman, V.R. Vermeul, J.L. Wagoner, H.F. Wang,

J. Weers, N. Welch, J. White, M.D. White, P. Winterfeld, T. Wood, S. Workman, H. Wu, Y.S. Wu, E.C. Yildirim, Y. Zhang, Y.Q. Zhang, Q. Zhou, and M.D. Zoback.

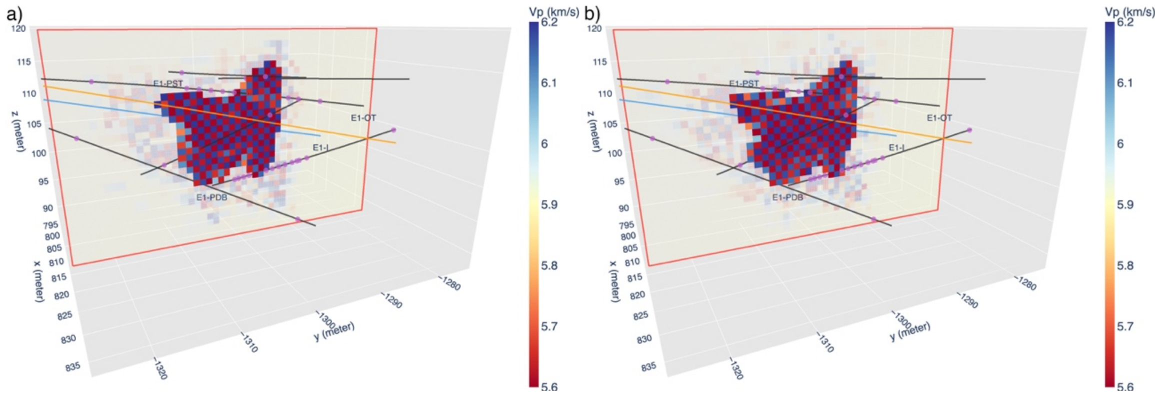

Figure S7. Checkerboard tests using (a) TL-derived phase picks and (b) manual phase picks for P-wave velocities. The inverted velocities are less than 0.1 km/s different from the true values in the highlighted area.

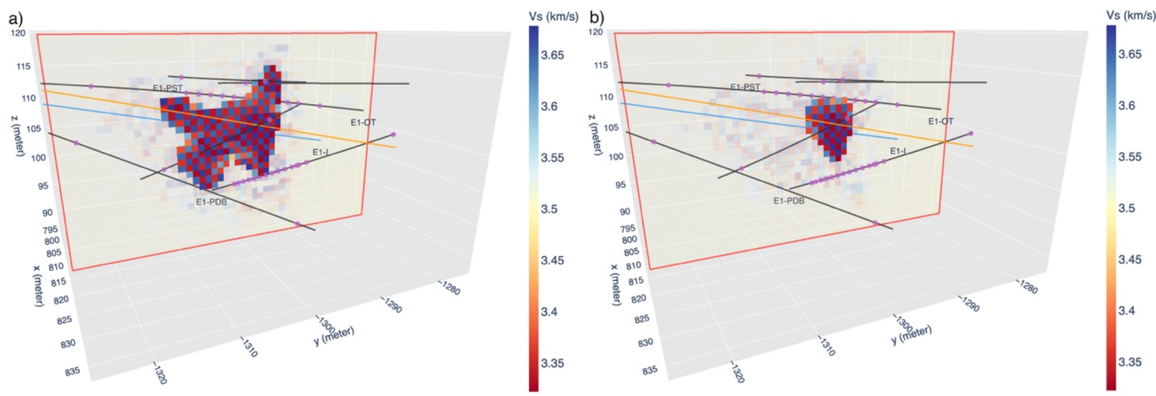

Figure S8. Checkerboard tests using (a) TL-derived phase picks and (b) manual phase picks for S-wave velocities. The inverted velocities are less than 0.06 km/s different from the true values in the highlighted area. The highlighted area in (a) is larger than that in (b), indicating that the TL-derived phase picks have better data coverage.

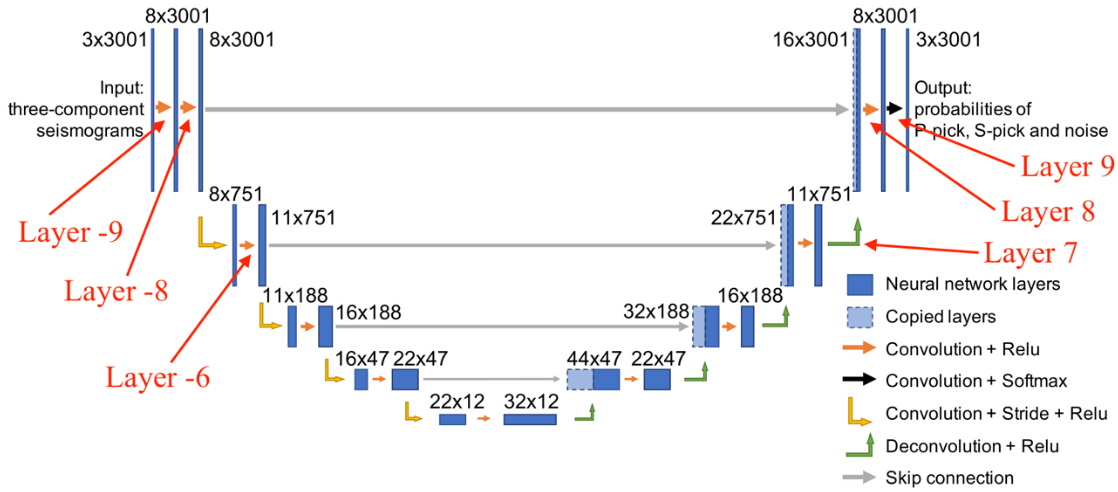

Figure S9. The neural network architecture as inherited from the PhaseNet model (W. Zhu & Beroza, 2018) with the layers used in Figure S10-S13 identified. The layers are numbered from -9 (left) to 9 (right) with layer 0 located at the bottom of the U shape.

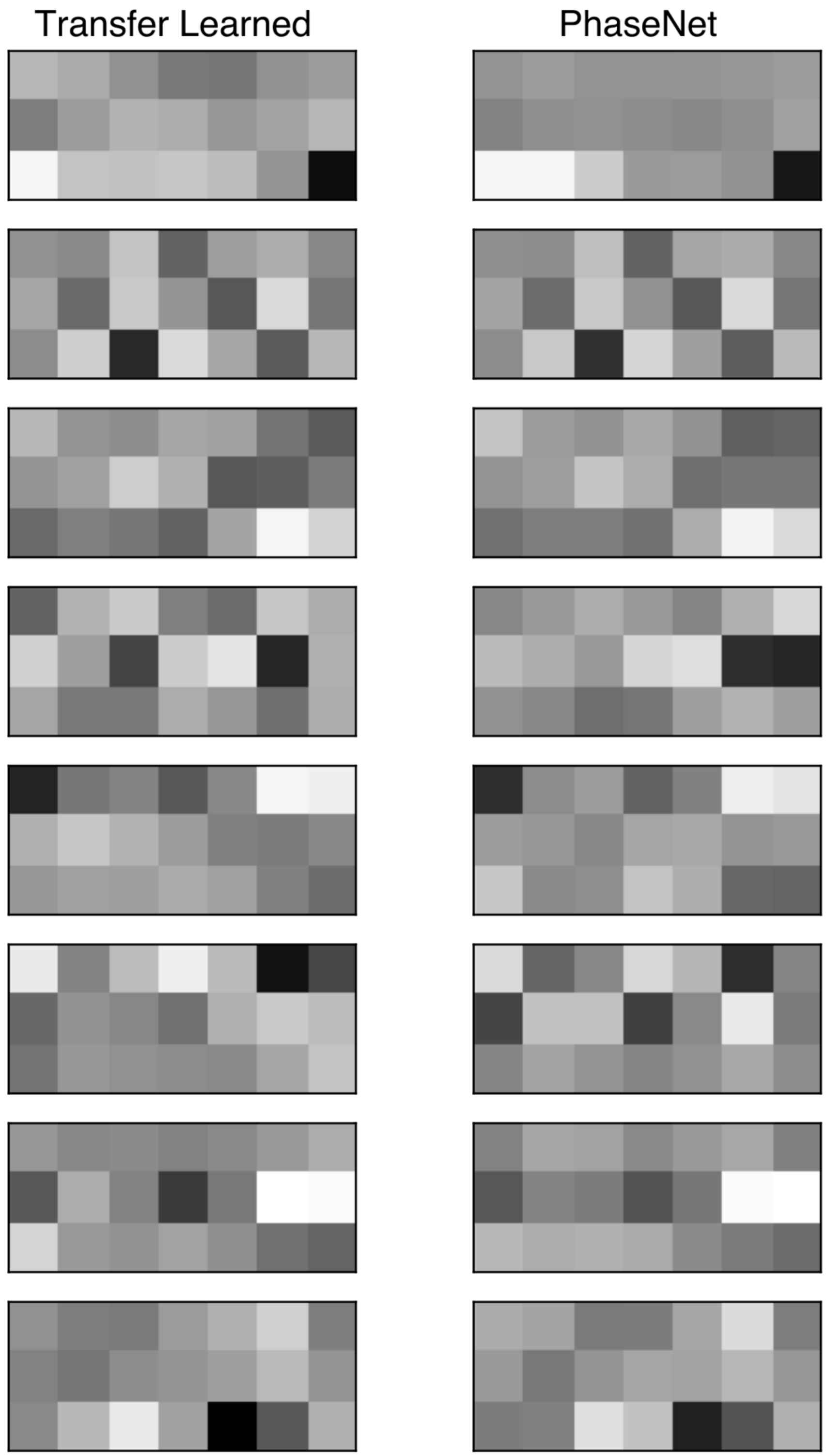

Figure S10. A comparison of convolutional filters of Layer -9 from (left) the TL model and (right) the PhaseNet model. The filters are two dimensional so that information from multiple channels is integrated. See Figure S9 for the layer location.

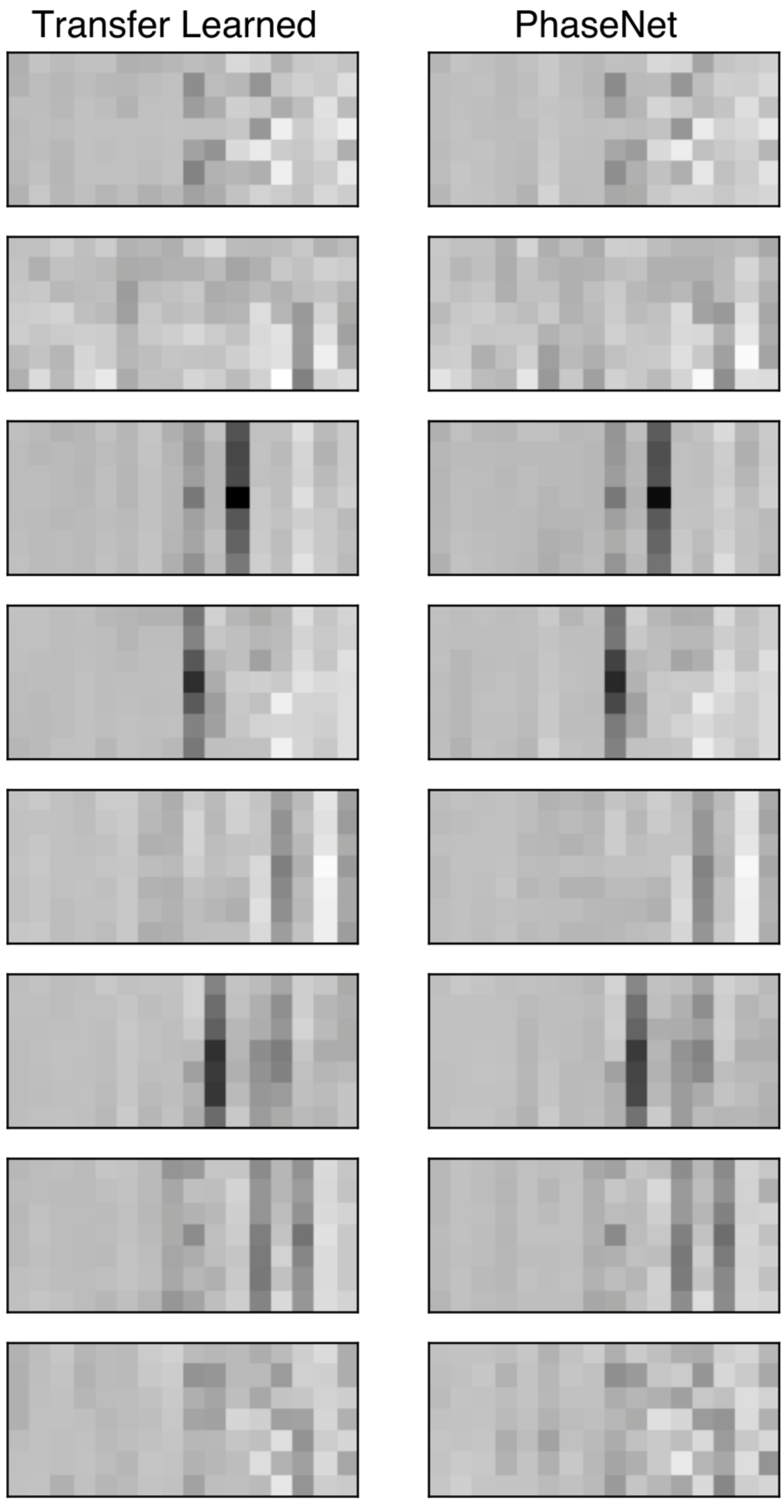

Figure S11. A comparison of convolutional filters of Layer 8 from (left) the TL model and (right) the PhaseNet model. See Figure S9 for the layer location.

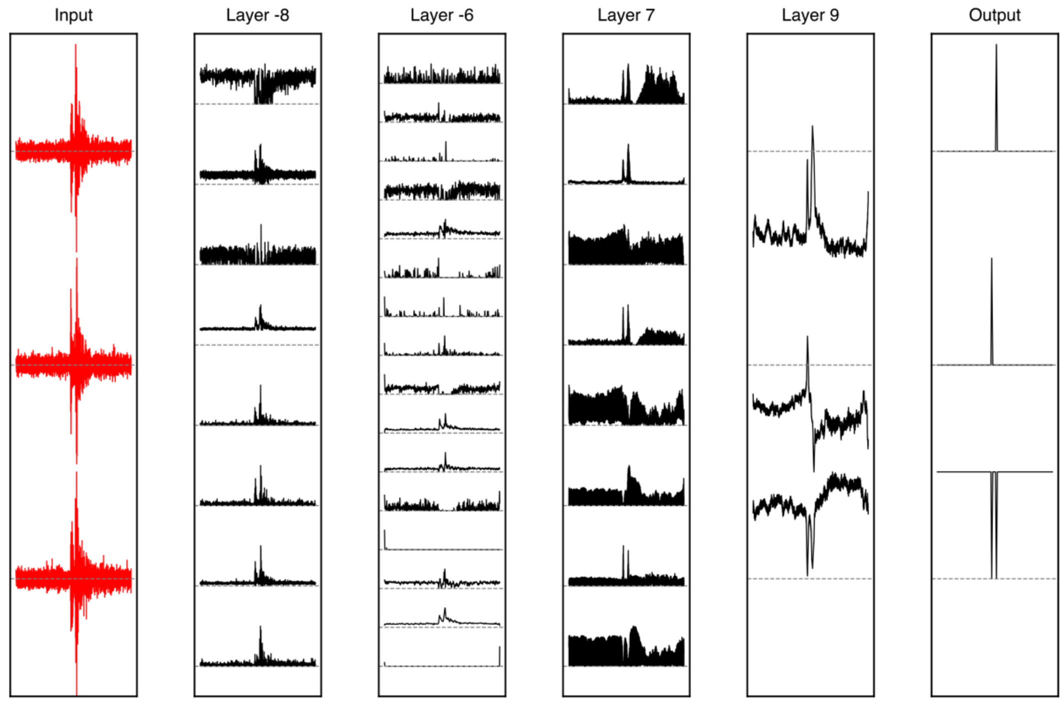

Figure S12. Example input, convolutional features of selected hidden layers, and output using the PhaseNet model. Dashed lines indicate zero. See Figure S9 for the layer location. In the output panel, the top curve corresponds to S waves; the middle curve corresponds to P waves; the bottom curve corresponds to “noise” (meaning neither P nor S waves).

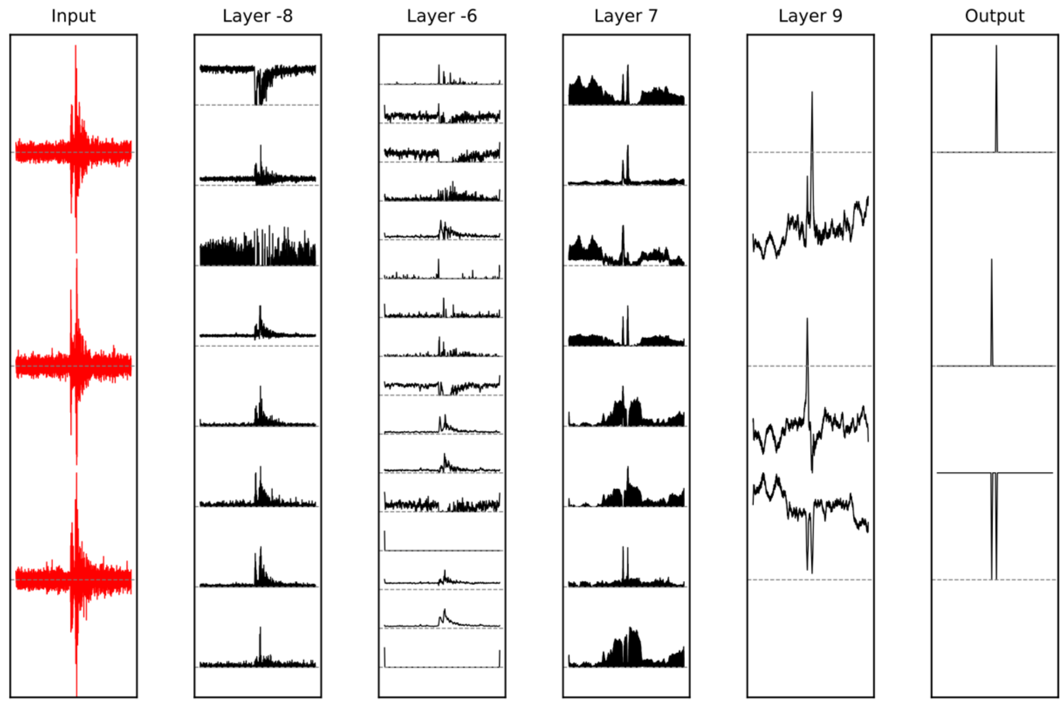

Figure S13. Example input, convolutional features of selected hidden layers, and output using the TL model. Dashed lines indicate zero. See Figure S9 for the layer location.

Geophysical Research Letters

Supporting Information for

Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking

Chengping Chai[1], Monica Maceira[1,2], Hector J. Santos-Villalobos[1], Singanallur V. Venkatakrishnan[1], Martin Schoenball[3], Weiqiang Zhu[4], Gregory C. Beroza[4], Clifford Thurber[5], and EGS Collab Team

[1]Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

[2]Department of Physics and Astronomy, University of Tennessee, Knoxville, Tennessee, USA.

[3]Lawrence Berkeley National Laboratory, Berkeley, California, USA.

[4]Department of Geophysics, Stanford University, Stanford, California, USA.

[5]Department of Geology and Geophysics, University of Wisconsin-Madison, Madison, Wisconsin, 53706, USA

Contents of this file

Introduction

The supporting information includes a paragraph (Text S1) that explains precision, recall, and F1 score, a figure (Figure S1) of sensor and borehole locations, a figure (Figure S2) with example seismograms, a figure (Figure S3) showing number of data used for transfer learning (TL), histograms (Figure S4-S5) that compare data fit for both P and S waves and for manual picks and TL-derived picks, example manual phase picks and TL-derived picks with corresponding seismograms (Figure S6), checkerboard test results for P and S waves (Figure S7-S8), a figure (Figure S9) showing the neural network architecture, two figures (Figure S10-S11) comparing weights of two convolution layers, a figure (Figure S12) showing hidden convolutional features using the PhaseNet model (W. Zhu & Beroza, 2018), and a figure (Figure S13) showing hidden features using the TL model.

Text S1.

For classification tasks, a prediction is defined as true positive when both the predicted label and ground truth are positive; as true negative when both the predicted label and ground truth are negative; as false positive when the predicted label is positive but the ground truth is negative; as false negative when the predicted label is negative but the ground truth is positive. For our case, true positive means the difference between a deep-learning-derived (or TL-derived) pick (P-wave or S-wave) and the corresponding manual pick is less than the assumed measurement error. True negative means both the deep learning model (or TL model) and the human analyst were not able to find a pick from a seismic waveform. False negative means that the human analyst found a pick from a seismic waveform but the deep learning model (or TL model) was not able to find a pick. False positive means the deep learning model (or TL model) found a pick from a seismic waveform but the human analyst did not, or the time difference is larger than the assumed error. Precision, recall, and F1 score were computed using the following

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

formulas.

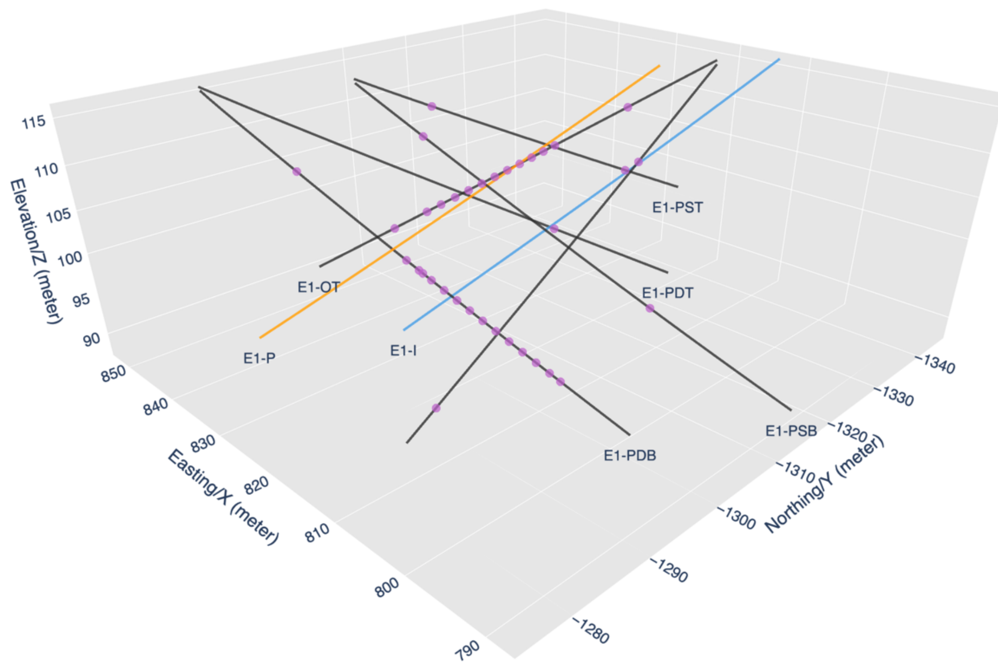

Figure S1. Seismic sensors (purple dots) and boreholes (lines). The blue line is the injection well (E1-I), and the yellow line is the production well (E1-P). Black lines represent monitoring wells.

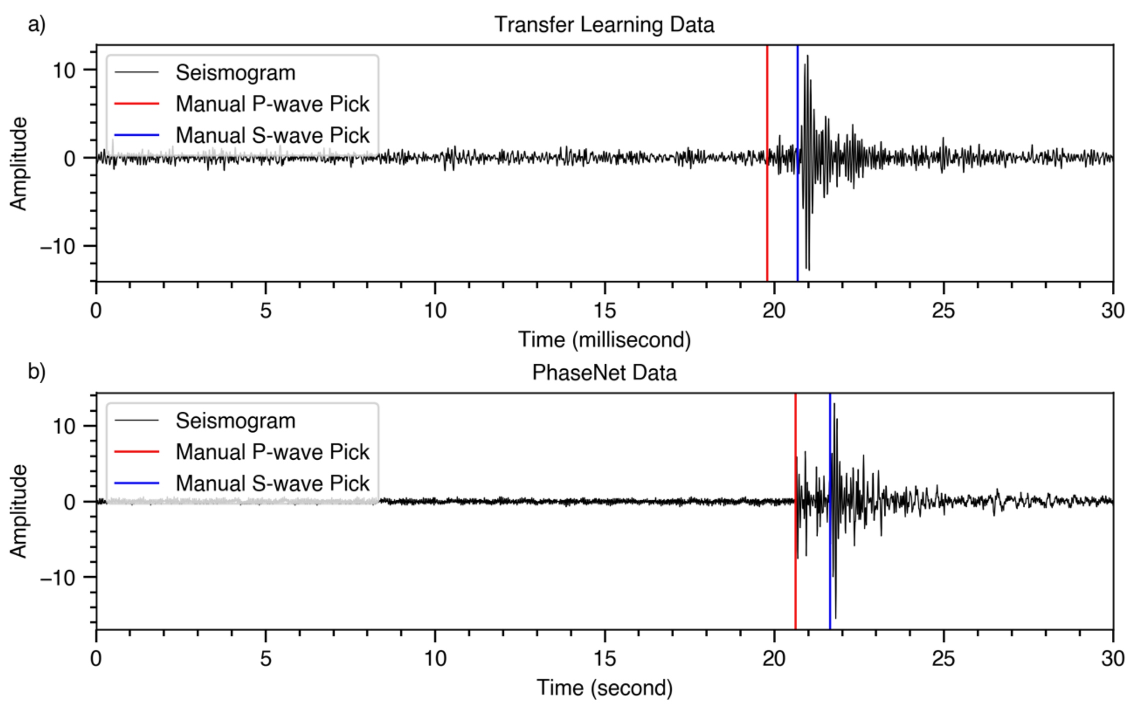

Figure S2. A comparison of seismograms used for (a) TL and (b) the original PhaseNet.

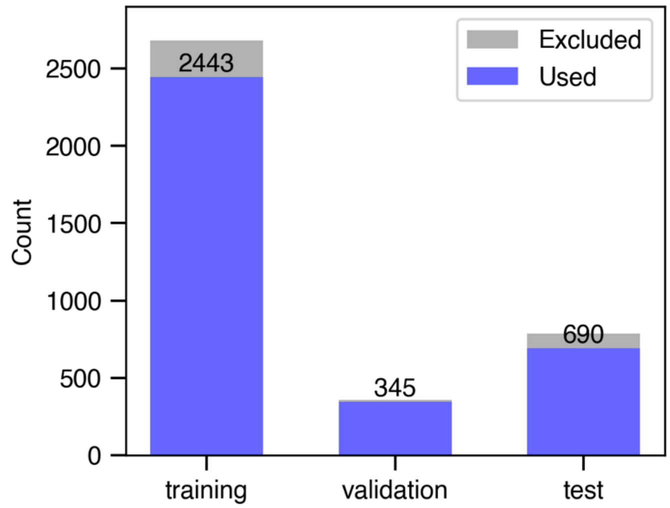

Figure S3. Number of waveform samples of training, validation and test set used for TL.

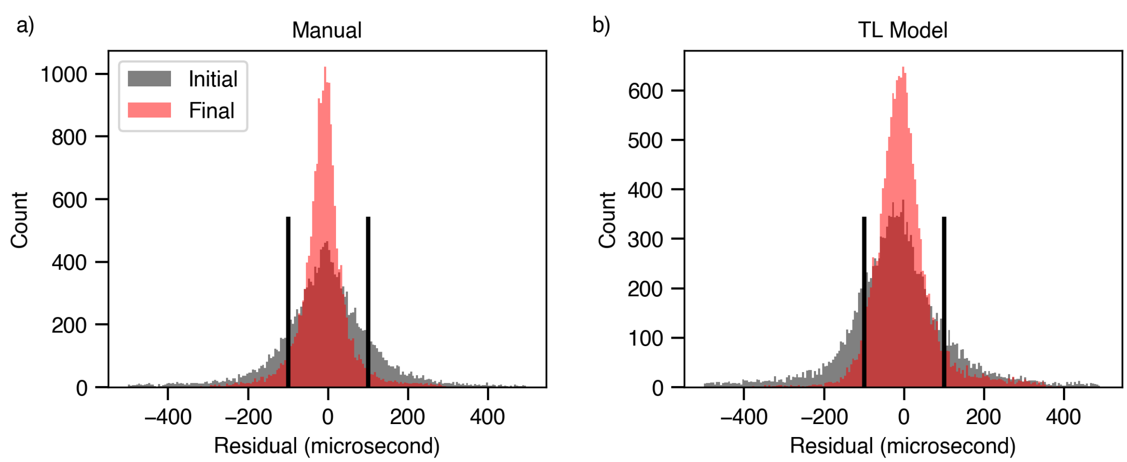

Figure S4. Histograms comparing the P-wave residual difference between the initial velocity model and the final velocity model for (a) manual picks and (b) TL model picks. The vertical lines indicate the assumed uncertainty of phase picks.

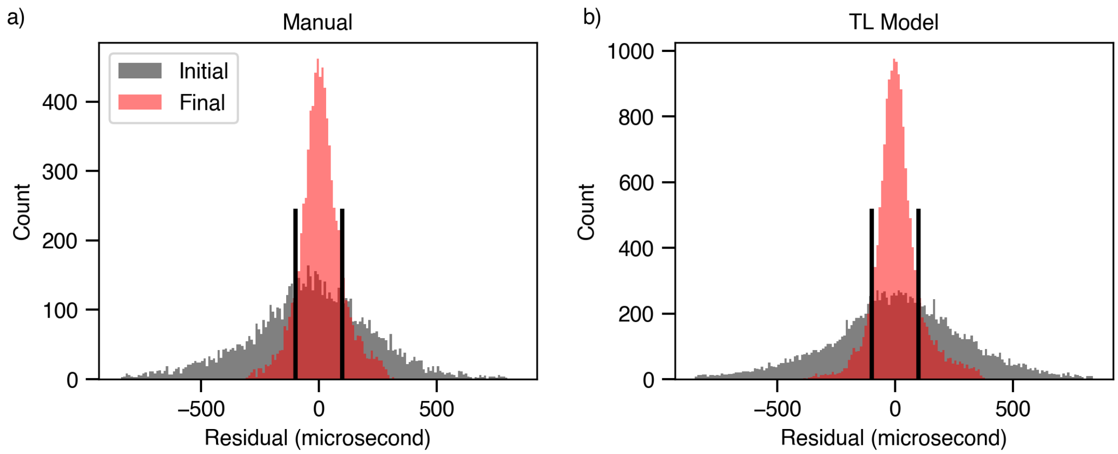

Figure S5. Histograms comparing the S-wave residual difference between the initial velocity model and the final velocity model for (a) manual picks and (b) TL model picks. The vertical lines indicate the assumed uncertainty of phase picks.

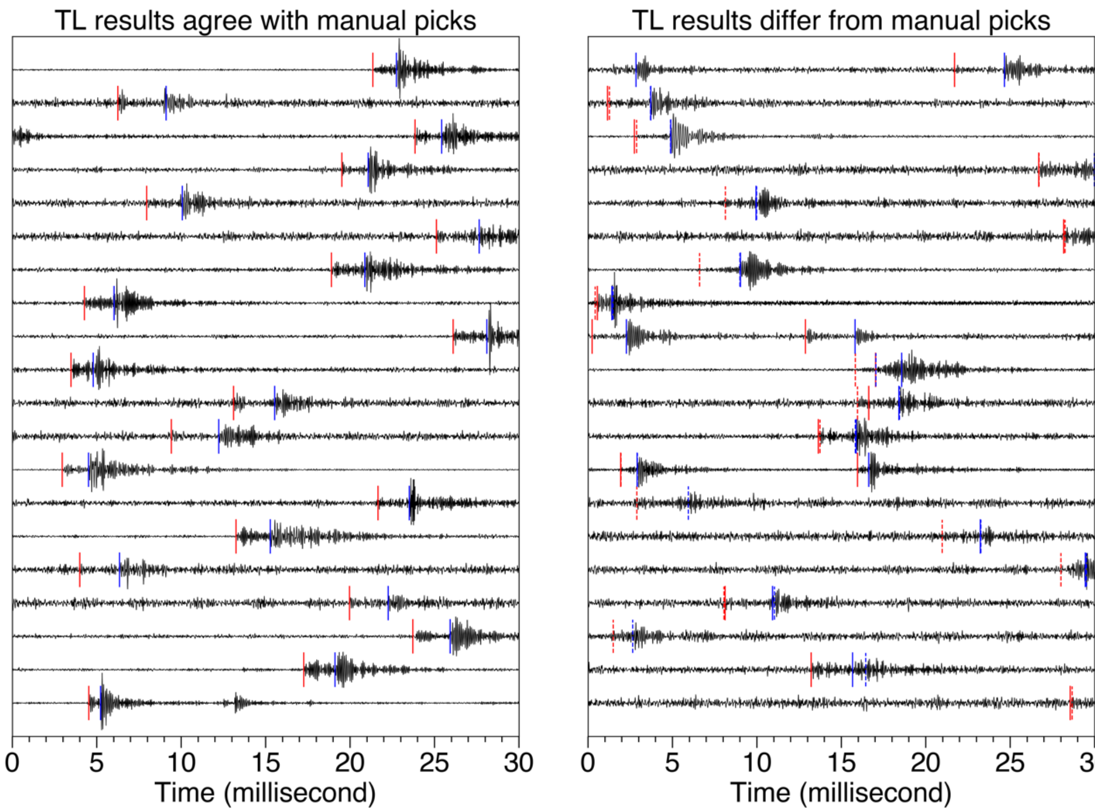

Figure S6. Randomly selected example waveforms and TL-derived phase picks that agree with (left) and differ from (right) the manual picks. Only one component seismogram is shown here, but three component seismograms were used for phase picking. Red vertical lines represent P-wave picks. Blue vertical lines are for S-waves. Solid vertical lines are TL-derived picks. Dashed lines are manual picks. When the time difference between TL results and manual picks is larger than 0.1 ms or one of the picks is missing, we consider the results to be different.