EQDetect: Earthquake phase arrivals and first motion polarity applying deep learning

Johnson Christopher W[1] and Johnson Paul A.[2]

[1]Los Alamos National Laboratory
[2]Los Alamos National Laboratory (DOE)

November 16, 2022

Abstract

Earthquake detection is critical for tracking fracture networks and fault zone deformation, particularly microseismicity that produces weak ground motions. We develop deep learning models to detect seismic phase arrivals and first motion polarities. The detection model is a convolutional encoder-decoder with a multi-head attention latent space that assigns a softmax value to each data point in continuous seismic records for classifying earthquake waveforms and the phase arrivals. The multi-output classification model utilizes weighted categorical cross entropy for the different softmax predictions to account for the unbalanced number of signal points compared to noise. The model training uses a benchmark data set of global seismic waveforms and the events are augmented using various techniques to reduce the signal-to-noise ratio, simulate multiple events arrivals, and channel failures. Detected p-waves are passed through a second model to obtain the first motion polarity. The phase arrivals, first motions, arrival waveforms, and additional metrics needed for catalog development are saved in a detection table. A neural network phase associator is used with the detection table to build an event arrival table. Locations are calculated and double difference locations are produced using correlation metrics from the waveforms retained in the detection table. The analysis is wrapped in a multiprocessing workflow to efficiently analyze large data sets. As a case study the workflow is applied to southern Kansas, a region with increased seismic activity related to hydrocarbon-production and waste water injection. The deep learning seismicity and focal mechanism catalogs show immensely more seismic activity than standard processing.

EQDetect: Earthquake phase arrivals and first motion polarity applying deep learning

Christopher W. Johnson[1] and Paul A. Johnson[1]

[1]Los Alamos National Laboratory, Geophysics Group, Los Alamos, N.M., USA

Corresponding author: Christopher W Johnson (cwj@lanl.gov)

Key Points:

- Deep learning earthquake phase detection and first motion polarity models are developed and extensively tested for model design performance
- Model layers in neural network are stress tested to determine where performance gains are attainable while simplifying the design
- Case study for southern Kansas shows widespread activity with twice the number of relocated events and a dense focal mechanism catalog

Abstract

Earthquake detection is critical for tracking fracture networks and fault zone deformation, particularly microseismicity that produces weak ground motions. We develop deep learning models to detect seismic phase arrivals and first motion polarities. The detection model is a convolutional encoder-decoder with a multi-head attention latent space that assigns a softmax value to each data point in continuous seismic records for classifying earthquake waveforms and the phase arrivals. The multi-output classification model utilizes weighted categorical cross entropy for the different softmax predictions to account for the unbalanced number of signal points compared to noise. The model training uses a benchmark data set of global seismic waveforms and the events are augmented using various techniques to reduce the signal-to-noise ratio, simulate multiple events arrivals, and channel failures. Detected p-waves are passed through a second model to obtain the first motion polarity. The phase arrivals, first motions, arrival waveforms, and additional metrics needed for catalog development are saved in a detection table. A neural network phase associator is used with the detection table to build an event arrival table. Locations are calculated and double difference locations are produced using correlation metrics from the waveforms retained in the detection table. The analysis is wrapped in a multiprocessing workflow to efficiently analyze large data sets. As a case study the workflow is applied to southern Kansas, a region with increased seismic activity related to hydrocarbon-production and waste water injection. The deep learning seismicity and focal mechanism catalogs show immensely more seismic activity than standard processing.

Plain Language Summary

EQDetect is designed for scanning continuous daily waveforms to detect earthquake phase arrivals from local to regional (150 km) events. The detections are made with a deep learning encoder-decoder model. When the model detects an earthquake in the waveforms, a second model is implemented to determine the first arriving motions. Both deep learning models are trained with the open source Tensorflow package using publicly available benchmark waveform data sets. The model output is a data table of time stamped detections, signal amplitude, signal-

to-noise ratio, and softmax probability of the detection in a generic format applicable to post-processing association algorithms for event locations. Additionally, the p-wave and s-wave waveforms are saved in a data table for rapid access when producing improved locations using correlation-based techniques. The workflow is designed for multiprocessing with multiple GPU's for rapid processing of large data sets. The model performance is shown for a case study in southern Kansas to detect event associated with wastewater injection activities.

1 Introduction

Earthquake detection relies on signal processing techniques to identify an emergent or impulsive signal in continuous waveforms. Seismic phase arrival times are the fundamental pieces of information required to identify and locate sources of ground motion in a heterogeneous volume of material and track subsurface deformation. Deep learning detection models specifically designed to identify seismic phase arrivals in continuous waveform data excel at identifying weak, low-amplitude regional earthquakes in noisy data, and vastly increase the number of event arrivals to track the microseismicity (Ross et al., 2020). Proof-of-concept machine learning architectures demonstrate these tools are driving the next phase of large-scale seismic data processing (Aguiar & Beroza, 2014; Hammer et al., 2012; Perol et al., 2018; Yoon et al., 2015) and more advanced model designs with a wide range of the number of layers and trainable parameters are beginning to showcase the full potential of these for application to earthquake detection (Chai et al., 2020; Mousavi et al., 2020; Mousavi, Zhu, et al., 2019; Reynen & Audet, 2017; Ross, Meier, Hauksson, et al., 2018; Saad et al., 2021; Woollam et al., 2019; Zhou et al., 2019; Zhu & Beroza, 2018).

Despite the differences in model design and data used to train the model, each application shows an increase in detection capability when compared to traditional signal processing techniques. Efforts are in progress to standardize the training data and test multiple algorithms for strengths and weaknesses (Münchmeyer et al., 2022; Woollam et al., 2022). As model designs continue to evolve it is necessary to dissect the architectures and determine what model layers are contributing most to the output decision and determine if modifying the implementation produces equivalent or improved results. This is an important consideration when the model task is to determine the ground motions from very small earthquakes, which can be masked or contain similar processes to surface processes producing noise (Johnson, Meng, et al., 2019; Johnson, Vernon, et al., 2019)

The focus here is to test model designs to improve microseismicity detections for a high-resolution catalog of events and focal mechanisms. For this effort we systematically test and redesign two existing deep learning models; EQTransformer for earthquake phase arrivals (Mousavi et al., 2020) and a classification model for first motion polarity (Ross, Meier, & Hauksson, 2018). The models are tested for performance using seismic data from a 5-year temporary deployment in southern Kansas where active wastewater injection occurs and elevated seismicity rates are reported (Rubinstein et al., 2018). Results are presented for the development and testing of the deep learning models and details of the performance when producing a microseismicity catalog with focal mechanisms.

2 Earthquake detection and first motion deep learning models

2.1 Earthquake detection

2.1.1 Detection model architecture

The convolutional neural network applied here combines earthquake detection and precise phase arrival time selection into a single model (Figure 1). The design is conceptually similar to Mousavi et al. (2020) in that it is a very deep encoder-decoder network with attention layers, but portions of that model are removed and redesigned, and moreover implemented differently. Changes to the Mousavi et al. (2020) architecture design were motived by systemically testing the addition and removal of layers to measure the performance of different model components, and ultimately many were removed. The deep learning model presented here includes two inputs—waveform time series and spectrogram, that improve detection of low signal-to-noise ratio (SNR) events. The model output is two classification probabilities that define the entire waveform for the presence of an earthquake and the timing of the seismic phase arrivals. The input and output were implemented to increase the amount of information provided to the model and to merge the phase arrival outputs into a single classification branch of the model. The choice to combine the phase arrivals is motivated by the need to weight the data labels due to the unbalanced nature of the continuous waveform.

As shown in Figure 1, the model inputs are 3 component waveforms and the corresponding time-frequency amplitude spectrograms. The waveform encoder branch contains seven 1-dimensional convolutional operators applying Leaky Rectified Linear Unit (ReLU) activation functions and maxpooling (size = 2). The spectrogram encoder branch contains four 2-dimensional convolutional operators with a Leaky ReLU activation function and maxpooling (size = [2,2]). The final time-frequency, high-dimensional representation is flattened, padded with zeros for reshaping, and concatenated to the final layer in the waveform encoder. The combined high-dimensional representation is passed to one residual connection layer with 10% drop out, then to a multi-head self-attention model (4 heads), with feed-forward neural network, skip connections, and 10% drop out (Vaswani et al., 2017). The output of the transformer-encoder is passed to the earthquake decoder using seven 1-dimensional convolutional transpose operators. The earthquake decoder output layer has 2 filters (size = [6000,2]) and uses a pointwise softmax activation function. All points in the waveforms are classified as noise or earthquake in this output branch of the model. The phase-arrival-time decoder branch is similar except the transformer-encoder is passed to a second transformer model, then seven 1-dimensional convolutional transpose operators. The final layer has 3 filters (size = [6000, 3]) and a pointwise softmax activation function. Each point in the waveform is classified as noise, p-wave, or s-wave. The maximum classification value is the phase arrival time.

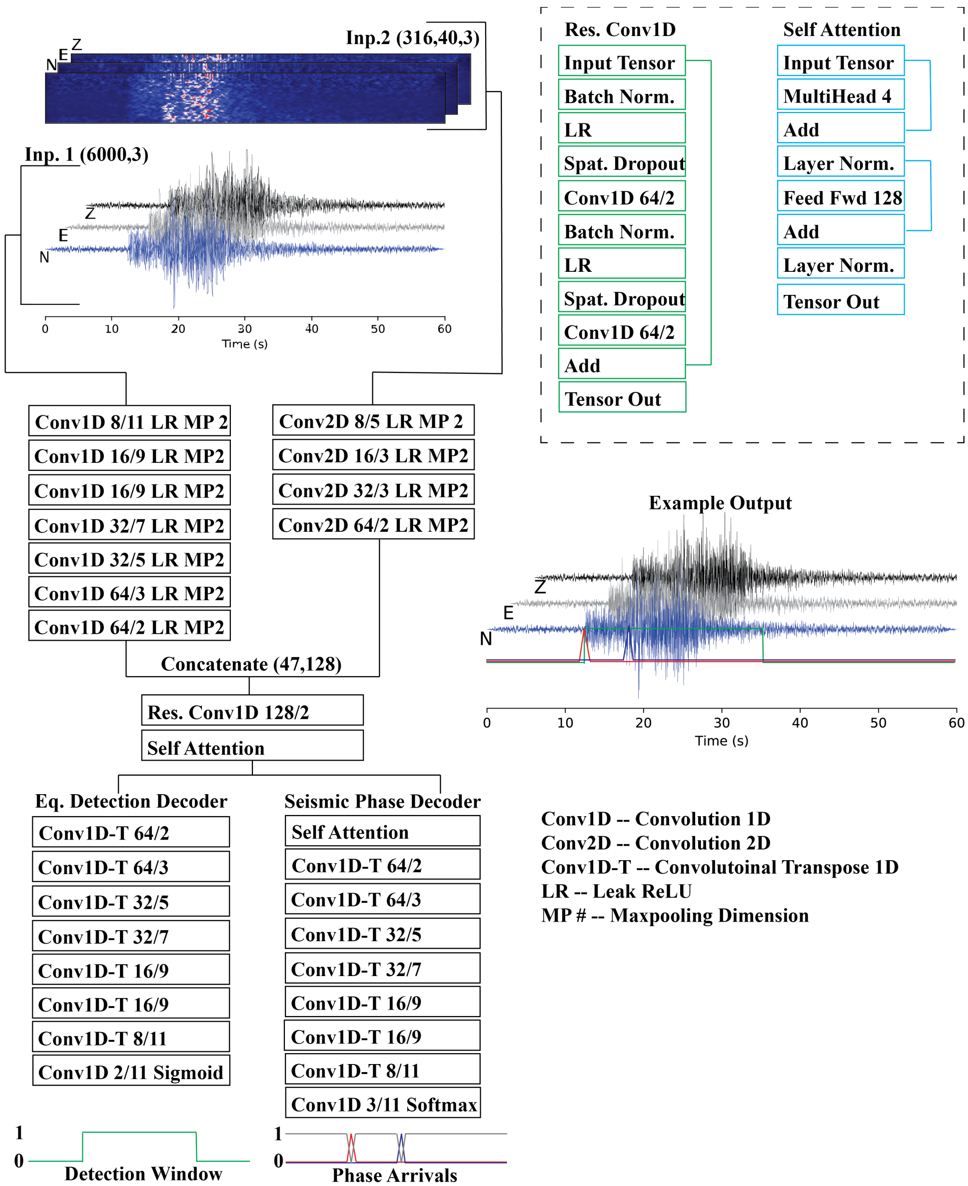

Figure 1. Earthquake detection and phase arrival deep learning model architecture. The inputs include 3-component normalized waveforms and amplitude spectrograms. The inputs are passed to 2 convolutional encoder branches that are then concatenated and passed to a residual network, attention layer, then to 2 decoder branches. Layer descriptions include the number of filters/kernel size. Inset on top right shows the layer design for the residual network and attention layer. The example output shows the waveform with the corresponding softmax output arrays.

2.1.2 Training data

High quality data with generalizable characteristics is essential to training an unbiased deep learning model. We use the seismic waveforms in the STanford EArthquake Dataset (STEAD; Mousavi, Sheng, et al., 2019), an established benchmark data set containing seismic phase arrivals, to assemble the training data. The data set contains about 1 million regional earthquake examples with p- and s-wave arrival times labeled and about 300,000 examples of noise waveforms. Each waveform is 1 minute duration at 100 samples per second (size = [6000, 3]) and contains the metadata for the station, events, and arrival time.

We develop a training data set as follows. The p-wave arrival times are randomly shifted by up to 30 seconds and limited to within 1 second of the waveform. A 2.5% (150 points) cosine taper is applied. The original data set is augmented using each set of three component waveforms by selecting a new time-shift for the p-wave arrival and reducing the SNR. To reduce the SNR the Fourier transform is applied, the real and imaginary components are independently shuffled, and the inverse transform is performed. The result is a noise signal for each channel that has nearly identical scaling properties without any energy-based impulsive signals. The noise waveform is scaled by a value selected from a uniform distribution between (0, 0.25] if the original SNR ratio is <1.5, otherwise it is scaled between (0.25, 0.5]. The scaled noise is added to the waveforms to produce a lower SNR phase arrival. For 8.33% of randomly selected waveforms, 1 or 2 channels are randomly selected and dropped, i.e., replaced with zeros. For another 8.33%, 1, 2, or 3 channels are randomly selected and intermitted channel drops between 15-30 seconds are simulated in the data. Additionally, to simulate multiple phase arrivals within a 1-minute window, 8.33% of the waveforms are augmented with a second phase arrival with a different SNR ratio than the original event in the time series. The final data set contains a total of 2,513,000 examples comprised of the original STEAD earthquake and noise and the augmented waveforms to represent regularly encountered scenarios in continuous daily seismic records.

The model design has 2 inputs that pass through separate encoder branches. The first is the 3-component waveforms normalized by the standard deviation with a tensor shape of (size = [6000, 3]). For the second input the waveforms are used to calculate the time-frequency amplitude spectrum with short-time Fourier transform using 80 samples per window (40 Hz Nyquist) with 76.25% overlap (size = [40, 316, 3]). The mean amplitude is removed and the spectrum is normalized to unit variance with the standard deviation.

The earthquake labels are assigned using a boxcar function between 0-1 starting at the p-wave arrival through the coda wave end time; the noise label is the opposite. The p- and s-waves are labeled using a normalized Gaussian function with a halfwidth of 0.1 seconds (10 points) centered on the arrival times. Each seismic phase is a separate vector and the noise label is the opposite of the combined arrival times. The output tensors are the earthquake duration labels (size = [6000, 2]) and the seismic phase arrivals (size = [6000, 3]).

The waveforms are split into 2,041,000 for training (81.2%), 314,000 for validation (12.5%), and 158,000 for testing (6.3%). The data is pre-processed and serialized with the TensorFlow *TFRecordDataset* module, which allows rapid access to thousands of files using optimized IO functions that run in parallel, shuffle the data, and map the tensors to the model input using the selected batch size. This is ideal when training on a GPU cluster with each node having 8 NVIDIA RTX6000 GPUs and 40 CPUs to reduce IO performance issues. The preprocessed data requires about 1 Tb of disk storage. This approach allows us to evaluate model performance using the same training and validation data while applying changes to the network design.

The model contains 354,221 trainable parameters and the Adam optimizer is applied with a learning rate of 1.0e-3 for epochs 0-50, 5.0e-4 for epochs 50-300, and 1.0e-4 for epochs >300. The loss function applied to each decoder branch is a weighted categorical cross-entropy using weights of 0.25 and 1.00 for the noise and earthquake signals, respectively, and 0.006, 1.0 and 0.8 for the noise, p-wave, and s-waves, respectively. The weights are selected to balance the large number of points that are noise, i.e., the phase arrivals are labeled using a total of 40 points in a vector of length 6000. The total loss is scaled to 0.05 for the earthquake detector branch and 0.95 for the phase arrival times. The F1-score for each output is also calculated to monitor training performance. The batch size is set to 32 per GPU for a total 63,781 batches per epoch. Training is terminated when the combined loss function shows no improvement for 50 epochs. The model training takes approximately one day if only using 1 node with 8 GPUs.

2.1.3 Evaluating the deep learning architecture

Extensive testing was completed to better understand the model design presented by Mousavi et al. (2020) to determine the following: (1) What contributes to the model performance? (2) Does adding additional input information produce gains? (3) Can the model design be simplified? The final model design presented here contains approximately 354k trainable parameters, about 18k less than Mousavi et al. (2020), and contains additional input data with a second encoder branch. Without the second encoder branch the model has about 200k trainable parameters. The model design (Figure 1) evolved by adding and removing specific layers to evaluate changes in training performance. Because the training and validation data did not change, we can infer the performance differences arise from the subtle changes applied with each training iteration.

A systematic grid search of hyperparameters was not performed because the interest was learning how changes to the input and altering or removing specific layers would impact performance. For example, a model that included only 3 component waveforms as inputs correctly identified 96% of the phase arrivals in the test data. Altering the model to include a parallel encoder branch that included the spectral amplitudes increased the performance by 2% and was adopted in future iterations. We found using long-short term-memory (LSTM) layers subsequent to the encoders produced no improvement in the loss function or validation data detection metrics. The addition of LSTM layers did increase the model training time by a factor of 3x-5x due to the sequential calculations required for this operation. Implementing repeating residual convolutional layers did not improve performance beyond using a single residual convolutional layer. The model architecture, specifically the encoder and decoder layers, were stress tested by systematically reducing the layers and the number of filters so we could determine the lowest model dimensions, i.e., the lowest number of trainable parameters, before the results began to degrade. Rectified linear units (ReLUs) were used in many of the tests, but ultimately changed to a Leaky ReLU activation function which improved the model results. This suggests previous model designs suffered from vanishing gradients, possible due to the very deep architecture. A stochastic gradient descent optimizer was tested and produced poorer results than the Adam optimizer. Model results using a constant learning rate of 10^{-3} produced nearly identical results when compared to applying the learning rate schedule. However, without the scheduler the number of training epochs increased by a factor of 2 without any performance gains, so the scheduler was adopted.

The largest improvements were realized when testing the addition of multi-head self-attention layers (Vaswani et al., 2017). This is different from previous implementations that use a restricted width self-attention model to isolate the signal of interest (Mousavi et al., 2020). Instead, we apply multi-head self-attention to solve multiple attention scores simultaneously and isolate the signal of interest without a predefined width parameter. The final model design contains 2 multi-head self-attention models, each with 4 heads. Including more than 4 heads did not improve performance. The first attention layer is applied following the encoders and residual convolutional layer and isolates the high-dimensional waveform representation to include only the portion of the signal relevant to the earthquake detection. Removing this layer produced many incorrect earthquake detection probabilities where the background noise would produce softmax values between 0.1-0.3 instead of being near zero. With the attention layer included, the softmax probabilities during background noise, with no earthquake present, were almost always near zero. Implementing a second multi-head self-attention layer prior to the phase arrival decoder branch increased the ability to isolate the different phases without producing ambiguous results with multiple peaks in the softmax probability curve. Including multi-head self-attention layers is found to be very important, but incorporating a sequence of attention layers did not improve model performance.

Another source of improved performance is the phase arrival decoder branch that contains the p-wave, s-wave, and noise signal classification in a single softmax probability. The design allows the use of a weighted categorical cross-entropy loss function, which is found to greatly increase the predicted probabilities at the arrival times to values near 1 for much of the validation data, which is quite good for low SNR phase arrivals. The weighting was selected based on the width of the p- and s-waves (~0.2 s each) within the 1-minute time window. The same approach is applied to earthquake detection and the weight of the noise signal is set to 25% assuming a duration of about 15 seconds is common for a small earthquake arrival.

2.1.4 Training, validation, and testing

The final model is trained for 495 epochs and no improvement in the validation data loss function is observed after 445 epochs (Figure S1). The weighed loss function for the best model is 0.0019 obtained from a combination of a loss function of 0.028 for the earthquake detection and 0.0005 for the phase arrivals. The f1-scores for the earthquake detection and phase arrivals are 0.97 and 0.98, respectively. Sudden improvements are observed in the training metrics which correspond to the learning rate schedule that decreases at set intervals.

The model is applied to the testing data and evaluated using a softmax threshold of 0.9 for earthquake detection and 0.5 for p- and s-wave phases. A mask is developed with the earthquake detection probably by making a boxcar function with a 2 second padding window. This boxcar function is multiplied with the p- and s- phase arrival probability functions to remove any phase detections that do not correspond to an earthquake detection. Of the 158,000 waveform examples, 133,944 contain p- and s-waves while the others are noise. The model correctly identifies 98.66% and 98.78% of the p- and s-waves, respectively, within 0.5 seconds of the true value. The residual pick times indicate most detections are within 0.1 second (10 data points) of the true value (Figure 2). The mean p-wave residual, reported as the pick time less the actual time, is -0.02 seconds and the s-wave is -0.01 seconds. Both residuals are negative indicating the model is picking the arrival time slightly later than the true value, which is an average of 1-2 data points for this sample rate. The results show 20 false positives for p-waves and 15 for s-waves. There were 406 false negatives for p-waves and 254 for s-waves. The model performance is encouraging since the evaluation metrics are set to a high threshold that can be reduced for network data to increase the number of detections.

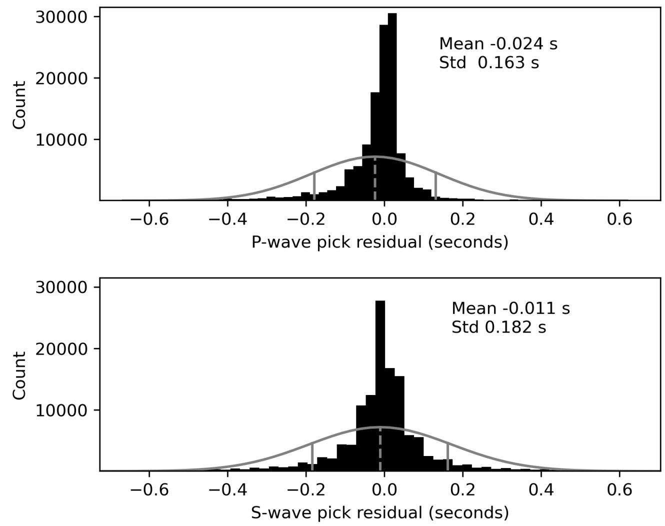

Figure 2. Histogram (black bars) of detection time residuals for (top) p-waves and (bottom) s-waves. The gray lines show the expected normal distribution and 1-standard deviation for the metrics from each set of residuals.

Waveform examples for the testing data show precise arrival picks, as expected from the performance metrics. The model correctly identifies the arrival window and selects a p- and s-wave arrival times of 0.03 and 0.00 seconds, respectively, from the actual arrival time (Figure 3a). For an example with a lower SNR, the model correctly identifies the arrival window and selects a p- and s-wave arrival time at -0.02 and 0.08 seconds, respectively, from the actual arrival time (Figure 3b). The 2 examples highlight the ability of the model to correctly identify the window containing the p-wave to the end of the coda wave for a range of augmented SNR situations.

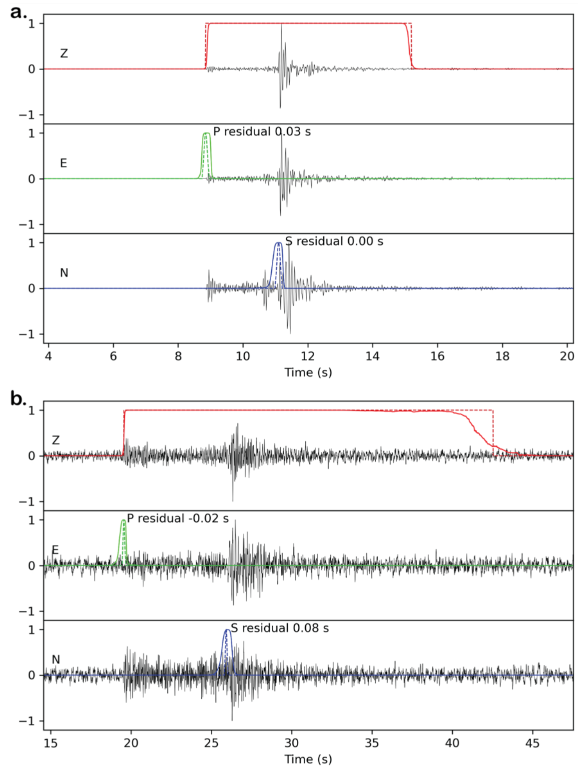

Figure 3. Model predictions for two examples of different durations and noise levels showing the Z, E, and N waveform components in black with the earthquake detection probability in red (Z frame), the p-wave in green (E frame), and the s-wave in blue (N frame). The solid lines are the model prediction and the dashed line is the training label. (**a.**) Earthquake waveform with duration of 7 seconds from the p-wave arrival to the end of the coda waves. (**b.**) Earthquake waveform with longer duration of 22 seconds from the p-wave arrival to the end of the coda waves and a lower SNR.

Another example demonstrates the model performance when multiple arrivals are present. The predictions are correct with the testing label, with about a 2 second spacing between arrivals (Figure 4a). However, when a p-wave arrives concurrently with an s-wave, the model does not predict the correct p-wave arrival, but does select the s-wave (Figure 4b). The model correctly identifies the first p-wave but does not estimate the arrival of the second that coincides with the s-wave. Both s-wave arrivals were correctly identified.

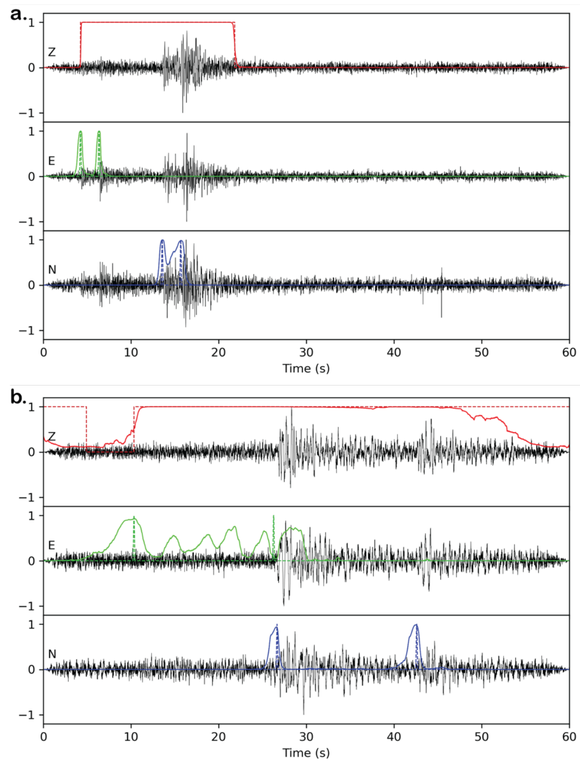

Figure 4. Examples of detection performance for multiple earthquake arrivals. The earthquake detection probability in red (Z frame), the p-wave in green (E frame), and the s-wave in blue (N frame). (**a.**) Two arrivals within 2 seconds are correctly identified for both the arrival window and the p-waves and s-waves. (**b.**) Two arrivals with the second p-wave arriving nearly simultaneous to the first s-wave.

An example of how the model performs when no earthquakes are present is also shown (Figure 5). The earthquake arrival probability is near zero for the entire time window. Both examples show a high softmax probability prediction for s-wave arrivals, but since the earthquake detection window is used as a mask the low probability p-wave and s-wave predictions will not be reported as false detections.

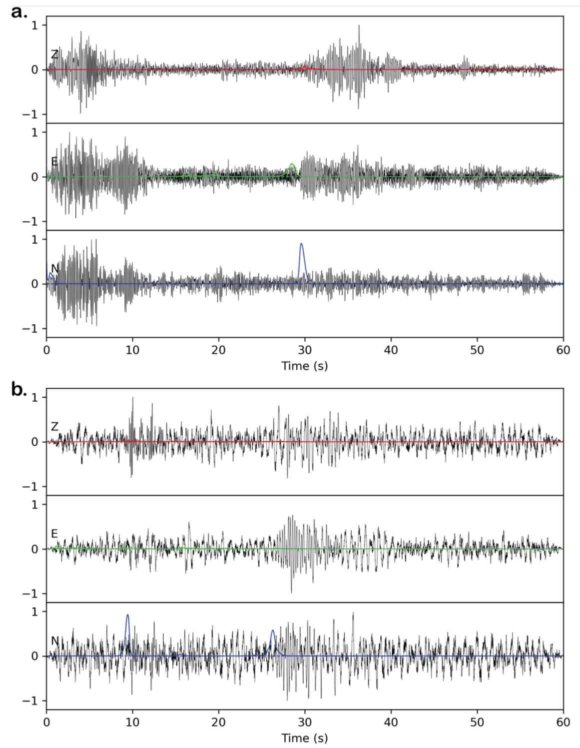

Figure 5. Examples of waveforms with no earthquake detections as indicated by the red line at zero for the entire 60 seconds. False phase arrival predictions with low softmax probabilities are observed for p-waves in panel (a) but not in panel (b). In both examples, the false s-waves predictions are observed with a high probability, however, the lack of a coinciding earthquake detection window would prevent these to be reported as false detections.

2.2 First motions polarity

2.2.1 First motions model architecture

The first motion for each vertical component seismic record is determined using a binary (up or down) classification model. The model input is the p-wave with ±1 second at 100 samples per second to 2 convolutional layers, each containing 16 filters with a Leaky ReLU activation function. No pooling or downsampling is applied. The high-dimensional p-wave representation (size = [200, 16]) is passed to a multi-head self-attention model (4 heads) then a single filter convolutional layer (size = [200,1]) with a liner activation. The transformed p-wave is passed to a fully connected neural network containing 50 neurons with a Leaky ReLU activation function and a binary classification layer using a softmax activation function (Figure 6). The model is a simple architecture that is designed to harness the multi-dimensional filtering of the convolutional layers as the input to the attention model. The model contains 16,325 trainable parameters. The primary assumption is that an up or down classification is always attainable from the data and no ambiguity from emergent signals is accounted for in the model output.

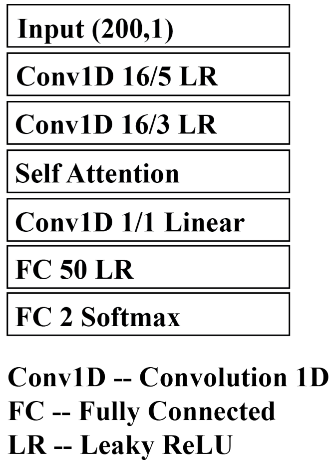

Figure 6. First motion polarity classification model design. The input is a vertical component p-wave arrival that is passed through 2 convolutional layers then a self-attention layer that is described in Figure 1. Layer descriptions include the number of filters/kernel size. The fully connected layers are followed by the number of neurons.

2.2.2 Training data

We use the first motion training data set procured from the Southern California Seismic Network (Ross, Meier, & Hauksson, 2018) that contains 2,494,194 vertical component waveforms with a duration of 6 seconds at 100 samples per second and labeled as up, down, or undetermined. We use ±1 s around the p-wave arrival as the input (size = [200,1]) and calculate the SNR. Note, the SNR is available in the meta data, but we recalculate it to ensure consistency since only 2 seconds of the waveform are used. To mimic the p-wave arrivals obtained from the earthquake detector model, the arrival time is randomly shifted off-center by an error bound of ±0.1 second. The signal is normalized by the maximum amplitude and no additional filtering is applied. After training many models with 3 classes and extensive manual waveform inspection, we decided to discard all waveforms with a SNR <1 to remove many incorrectly labeled examples. The decision to remove the waveforms with an undetermined label is to prevent false negative predictions if the model correctly identifies the first motion as up or down, but is labeled as undetermined which was found during testing. This issue is noted in the original study that produced the data set (Ross, Meier, & Hauksson, 2018) and we opt to eliminate as many waveform examples as possible that have the potential of an incorrect label.

For a balanced data set, the training data is split into 768,265 up and 767,735 down labels for a total of 1,536,000 examples, and the validation is split into 61,643 up and 62,122 down labels for a total of 123,765 examples. The data is pre-processed and serialized with the TensorFlow *TFRecordDataset* module. The Adam optimizer is applied using a learning rate of 1.0e-4 for training. The data set is balanced and no weighting is applied. The batch size is set to 32 per GPU (48,000 batches per epoch) and the training is terminated when the loss function shows no improvement for 25 epochs.

2.2.3 Evaluating the model design

The model architecture is a simple design and systematic testing showed increasing the model complexity did not improve results. The initial modeling tests included a 3-class prediction for up, down, and undetermined that included architectures mimicking (Ross, Meier, & Hauksson, 2018) with a series of convolutional layers passed to a fully connected classification network. During model training the loss function and f1-score plateau quickly with no additional improvement gains from hyperparameter tuning. This initial observation prompted testing of the training data by increasing the SNR and rebalancing the data set with equal numbers of up, down, and undetermined. The final decision was to remove the undermined labels since many were found to increase the model loss function by correctly identifying the first motion. Removing as many incorrectly labeled training data produces a more generalizable and robust model.

The next suite of models tested, and used for the final model, contain a simplified encoder design to utilize the self-attention network and isolate the important components in the signal for classification. The final model contains 2 encoder layers with 16 filters, which was reduced from 5 layer to determine when training performance decreased. This is followed by a self-attention layer that is passed to a fully connected network, then to a softmax classification layer. Utilizing a hidden layer improved performance and increasing the number of neurons above 50 showed no improvement. The final model has relatively few trainable parameters (~16k), or <1% when compared to the original network presented in (Ross, Meier, & Hauksson, 2018) that has about 2.4M trainable parameters to obtain very similar testing data metrics. This demonstrates the performance improvements that are possible when applying a self-attention network to simplify a classification problem.

2.2.4 Training, validation, and testing

The model is trained for 125 epochs and no improvement in the validation data loss function is observed after 100 epochs (Figure S2). Model training takes approximately 2 hours with a single node on the GPU cluster. The best model has a loss function of 0.18 and a f1-score of 0.94 using the validation data.

The test data contains 2,353,054 examples (Ross, Meier, & Hauksson, 2018) and we remove the waveforms with an undetermined label. The waveforms are selected as ±1 second around the p-wave arrival for the input and to calculate the SNR. The distance from the source and event magnitude are taken from the meta data. The model results are shown for these 3 metrics and all show a higher recall value indicating the false-negative predictions are slightly less impactful than the false-positives, as shown by the reduction in precision (Figure 7). The SNR indicates improved precision and recall for values >1 and is consistent with the cutoff SNR applied in the model training data. At the highest SNR the precision and recall are both approximately 0.97 with a consistent f1-score. The 3 metrics each decay with distance and events <10 km show the highest values around 0.9 indicating the SNR, regardless of distance to station, is more important. When evaluated against the magnitude the results show decreased performance for events with M<1, most likely due to low SNR at greater distances, and values that increase up to M2 before decreasing for the high magnitudes events. At higher magnitudes one might expect increased performance, however, this does not account for the increased number of stations detecting first motions at larger distances, which are shown to have reduced performance. Overall, the results indicate the model is successfully identifying the first motions and the choice to force the model to classify as up or down without an undetermined option is sufficient when applying additional constraints on the required information to constrain a robust focal mechanism (Uchide, 2020).

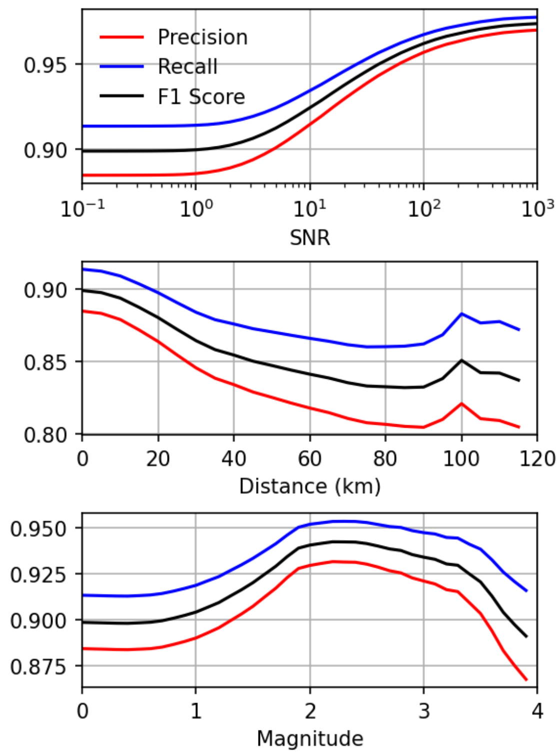

Figure 7. Precision, recall, and f1-score curves for the first motion polarity test data shown as a function of signal-to-noise ratio (SNR), distance, and earthquake magnitude.

3 Application to data from southern Kansas

3.1 Seismic waveform data

Harper and Sumner counties in southern Kansas are the location of increased seismic activity in recent years due to the deep-injection hydrocarbon-production-activities in the Arbuckle formation. Daily seismic waveforms are obtained for 19 stations (Table S1) deployed by the U.S. Geological Survey (Rubinstein et al., 2018). The initial monitoring began in mid-2014 and included 5 accelerometers. The network expanded in 2015 with broadband and accelerometer sensors operating until mid-2019. Here we utilize 2 accelerometers and all broadband data from 17 stations in the GS and OK networks that coincide with previous studies (Cochran et al., 2018; Rubinstein et al., 2018). The daily records are more complete for the GS stations and many temporal gaps are in the OK network data (Figure S3). The analysis used 26,976 daily waveforms to develop a seismicity catalog for about 5 years in southern Kansas.

3.2 Continuous waveform processing

The continuous waveforms are applied to the processing workflow to build detection and arrival tables used for developing the event and focal mechanism catalog (Figure 8). The waveforms are preprocessed to obtain 3-component daily waveforms compatible with the deep learning models. The stations in the GS network are recorded at 200 samples per second and are resampled to 100 samples per second. The accelerometer data is integrated to obtain velocity. A 1 Hz highpass filter is applied to all data. If a daily trace is not complete, gaps are filled with zeros to produce 3 traces with an equal number of points. The model input uses 1-minute windows with a 30 second overlap to avoid edge effects.

The normalized waveforms and corresponding short-time Fourier transform amplitude spectrum are input to the phase detection model. The model outputs 2 sets of softmax probability vectors describing the earthquake detection and the phase arrivals times. The overlapping windows are removed by averaging the softmax outputs, which are equal except for the tapered ends of the waveforms. The onset of an earthquake detection is set to a softmax probability threshold of 0.3 and extents until a value of 0.05 is observed. This selection is based on setting a low threshold to allow removal of detections during postprocessing. The onset and offset detection times are extended by 1 second and used to produce a boxcar function [0,1] that is applied as a mask to the phase arrival model softmax probabilities. This restricts all phase arrivals selected to be within a window of high probability of an earthquake in the waveforms. The 1 second padding allows for the increasing probability of the p-wave arrivals, which is coincidental with the onset time, and not removing it with the mask. A threshold of 0.3 for p-waves and 0.3 for s-waves is applied for the phase arrival probabilities. If the softmax probability exceeds the threshold, the maximum value in the earthquake detection window is recorded. Multiple arrivals within an arrival window are limited to a minimum of 2 second separation, this follows the criteria used to produce the training data. All p-waves detected are passed to the first motion polarity model. The processing produces 13,704,495 phase arrivals containing 5,439,700 p-waves and 8,264,795 s-waves for the 19 stations in the ~5-year period. The arrival times are stored in a data table and additional metrics are collected from the waveforms for post-processing analysis. Additionally, the p-wave and s-wave waveforms are saved using ±1 second around the arrival times.

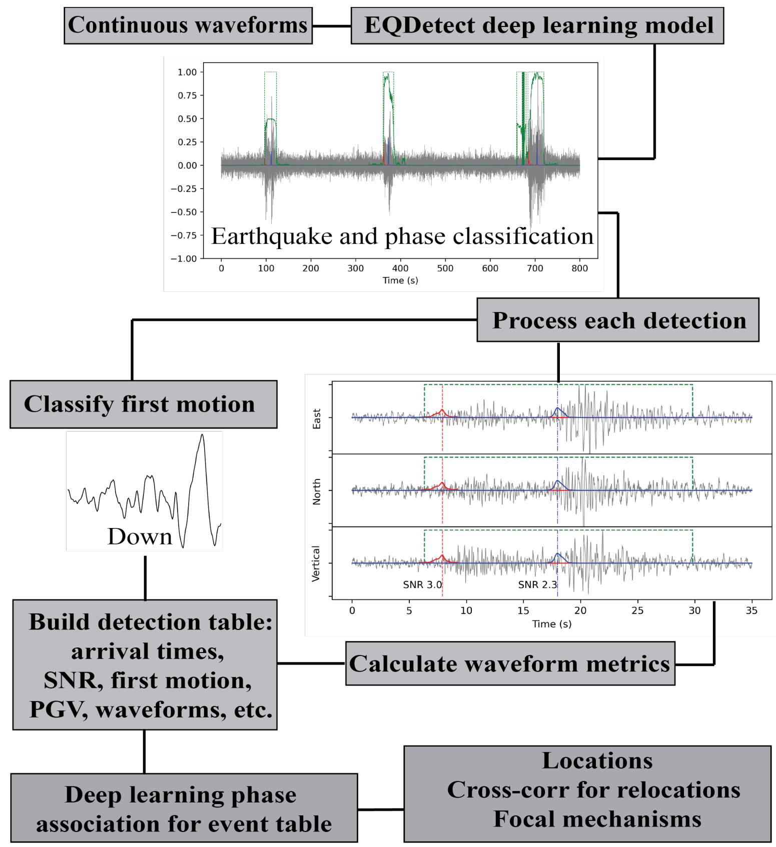

Figure 8. Processing workflow to apply continuous waveforms to deep learning models, build a detection table with waveform metrics, associate arrivals into event table, and develop catalog.

3.3 Earthquake locations and focal mechanisms

The phase detections are associated to event arrivals using a modified implementation of the PhaseLink neural network algorithm (Ross et al., 2019). This is an important processing step to obtain reliable locations and the neural network method shows much improved performance when compared to grid search algorithms (Ross et al., 2019). The model is trained using 10M synthetic phase arrival examples that are calculated for the geometry of the 19 stations using travel times from a 1D velocity model (Rubinstein et al., 2018). The model is trained for 100 epochs and the epoch with the lowest loss function is selected to perform the associations. Event association requires a minimum of 5 arrivals with the criteria of a minimum of 3 p-waves, with 2 stations having a corresponding s-waves at the same station. The arrivals are back projected to a preliminary location, the association procedure is then repeated with a 2 second tolerance for arrival times in the detection table. A minimum of 5 arrivals is required for a final set of associated arrival times, but the majority of events have more than 10 arrivals.

Event locations are calculated using NonLinLoc (Lomax et al., 2000) to implement a probabilistic earthquake location search procedure with a 1D velocity model . The grid search depth is set to a minimum of 2 km to eliminate events locating at the surface and extends to a 200 km square. The region is set between -98.4˚to -97.1 and 36.6˚to 37.45˚ for a catalog of 32,844 events (Figure 9a). These locations are used with the GrowClust relocation algorithm

(Trugman & Shearer, 2017). Waveform cross correlations are performed for p- and s-wave phase arrivals using up to 1,000 events within 10 km that have a correlation coefficient ≥0.5 at a minimum of 4 stations. The waveforms are filtered between 2-8 Hz and a spline interpolation is implemented to subsample the waveforms to 1000 samples per second for increased resolution in the temporal shift to obtain the maximum correlation value (Trugman & Shearer, 2017). The procedure relocates 12,749 (~39%) events in the catalog. Manual inspection of waveforms for events not relocated show p-wave and s-wave arrivals with the appropriate moveout, but the correlations with other events is not sufficient to relocate.

The results for the region of interest between -98.15˚to -97.50˚ and 37.00˚to 37.35˚ shows widespread seismic activity comprising 19,015 events (Figure 9a), with 8,974 relocatable (Figure 9b). The number of relocated events between 2014-2016 is double the number in the catalog produced by Rubinstein et al. (2018). With that catalog, Cochran et al. (2018) applied matched filtering to increase the number of events by an order of magnitude, and applied the template location to the matched event for a detailed temporal analysis of seismicity. The technique applied here detects new earthquakes in areas not previously showing activity and allows a spatiotemporal analysis of the events.

The remarkably widespread seismicity shown in Figure 9a has not previously been detected in this region, and suggests much of the upper crust has been stressed to failure due to fluid injections. The highest concentration of activity occurs in similar zones defined by Rubinstein et al. (2018), but a much more distributed pattern of activity is shown for the entire region. Injection wells documented by the Kansas Geologic Survey are located throughout the region, with the largest number in the southeast, many of which are injecting at much higher volumes than the average. The most active period is throughout 2015 and event numbers steadily decrease in the following years as injection volumes are reduced (Rubinstein et al., 2018). The b-value is between 0.9-1.1 for the catalog duration. The magnitude of completeness for all events in the study area is M_c1.4 and is consistent through the duration of the catalog (Figure S4). However, the local magnitude estimate is obtained by the peak waveform amplitude using the default parameters in the NonLinLoc software package and inconsistencies are expected if compared with other microseismicity catalogs that apply specific station corrections or other magnitude estimation methods (Shelly et al., 2021).

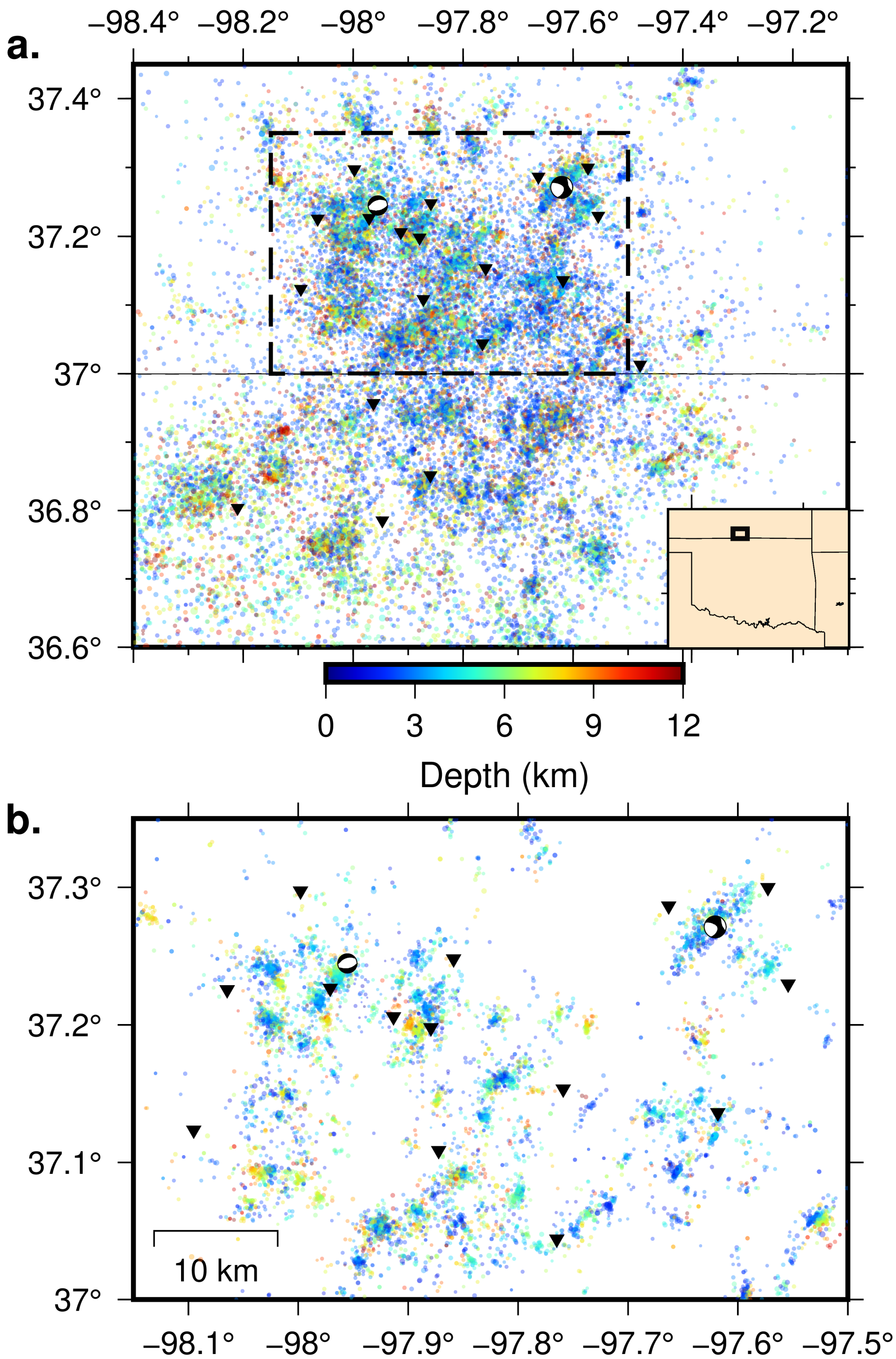

Figure 9. (a) Seismicity is shown by circles colored by depth for the entire area that extends from northern Oklahoma to Harper and Sumner counties in southern Kansas. Inverted black triangles are the seismic sensors. The moment tensors are the U.S. Geological Survey solutions for the M4.3 on 02 October 2014 and the M4.9 on 12 November 2014. The black dashed rectangle shows the region of interest in southern Kansas as shown the in the lower panel. (b) Relocated seismicity in the region of interest with events shown by circles colored by depth.

Focal mechanisms are calculated with the HASH algorithm (Hardebeck & Shearer, 2002) using the first motions determined during the waveform processing. A minimum of 8 polarities are required and the maximum azimuthal gap and takeoff angle gap are set to 120˚ and 70˚, respectively. A Monte Carlo procedure using 500 iterations is implemented to perturb the event depth, and therefore the take-off angle, to obtain a suite of acceptable solutions and report the average focal mechanism strike, dip and rake. The catalog contains 1,980 focal solutions for the region of interest with 202 A quality, 485 B quality, 799 C quality, and 494 D quality.

The increased number of focal mechanisms provide information about the stress orientation throughout the region of interest (Figure 10). Results are shown using quality A, B, and C solutions (N=1,486) and consist results are found when using only A and B quality solutions. The mechanisms are divided using a 0.0333˚ (3.7 km) grid to perform a normalized stress tensor inversion with the SATSI software package (Hardebeck & Michael, 2006). At each location the 30 best focal solutions are selected and a minimum of 12 are required to include in the inversion. The tensor shape is used to describe the expected fault type (Simpson, 1997). The results show a stress environment with oblique normal and strike-slip faulting trending east-northeast and is consistent with previous observations (Rubinstein et al., 2018; Skoumal et al., 2021), but with improved spatial resolution and high concentrations of focal mechanisms in active regions allowing more detailed investigations.

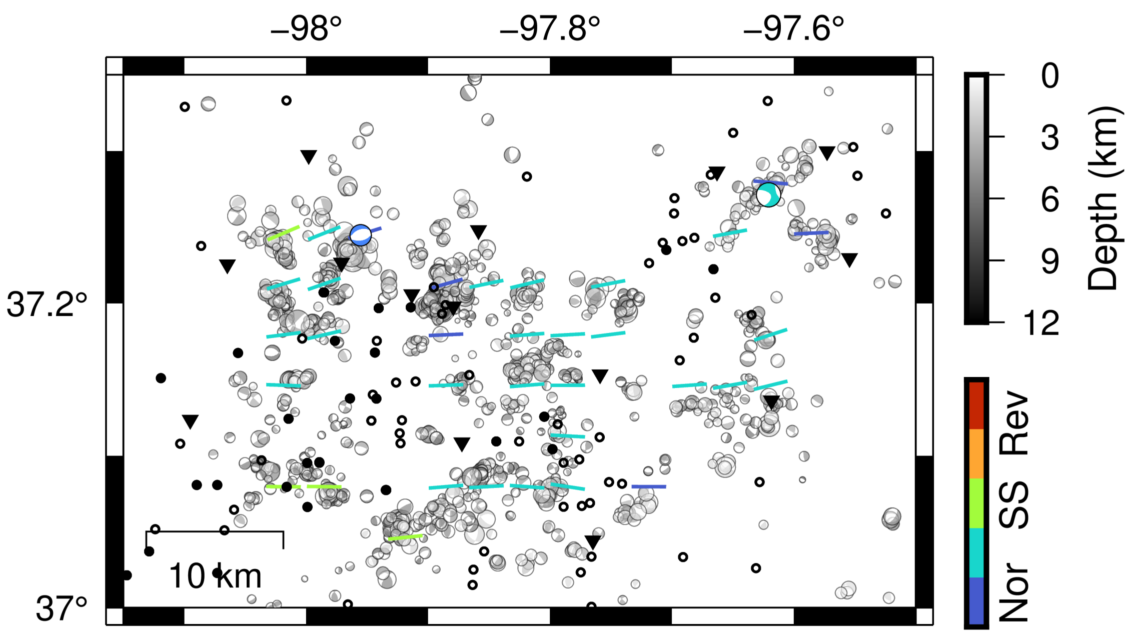

Figure 10. Focal mechanisms (N=1,486; quality A, B, and C) using the relocated catalog are shown in gray scale by depth. The S_{Hmax} orientation is shown with as a bar with the color indicating the expected slip from the shape of the stress tensor. The two moment tensors are the U.S. Geological Survey solutions for the M4.3 on 02 October 2014 and the M4.9 on 12 November 2014 and colored using the same scale for expected slip. The inverted black triangles are the seismic station locations and the black circles are the location of injection wells, with solid black indicating a high-volume well.

4 Discussions

4.1 Deep learning model improvements

The deep learning models developed here are rigorously tested to simplify the design for maximum performance. For the earthquake phase detection, we determined how to increase the information input to the model while reducing the total number of trainable parameters. Similarly, for the first motion polarity we developed a very simple model that performs at the same level as the original design. Two points motivate these efforts to refine models to minimum complexity for maximum performance. The first is to start moving towards explainable machine learning models (Gunning et al., 2019) that provide information to the end-user describing why a model decision was made. Systematic tests are needed to see how different models interpret weak ground motion signals from different training data sets or real-world scenarios. A logical application is real-time networks (Yeck et al., 2020), where having more information provided to the user when making automated decisions would be useful. The second point is related to real-time monitoring using edge computing (Chen & Ran, 2019) to make decisions in the field. This second point is directly applicable to earthquake early warning efforts (Li et al., 2018). These considerations will increase the interpretability of model decisions and allow deployment onto a range of sensors without the loss of performance.

Training data is a critical component for all deep learning models and consensus from the seismology community on a set of benchmark standards is not currently in practice. For example, 3 comprehensive data sets available each use different signal length and sample rate; 60 seconds at 100 samples per second (Mousavi, Sheng, et al., 2019), 27 seconds at 20 samples per second (Magrini et al., 2020), and 60 seconds and 40 samples per second (Yeck et al., 2020). This results from research centers having recording rates appropriate from their network processing. Different network operations have varying performance standards with respect to real-time processing for global detection or local to regional events. These differences are expected for different networks but impactful when designing a training data set. Additionally, the detection model design needs to be crafted for the data-input and model-output most applicable to the task. This will be different for real time processing (Yeck et al., 2020) versus exploratory research as presented in this study. The detection model developed here uses the STEAD data set (Mousavi, Sheng, et al., 2019), which contains about 300,000 examples of waveform noise. Efforts to detect very low SNR signals in non-optimal environments will require novel data sets containing a variety of noise signals. Many natural and anthropogenic processes produce weak ground motions (e.g., De Angelis & Bodin, 2012; Inbal et al., 2018; Johnson, Meng, et al., 2019; Meng et al., 2019; Meng & Ben-Zion, 2018; Qin et al., 2019) that can obscure microseismicity and contains energy in similar spectral bands. Development of a comprehensive collection of non-tectonic noise signals that is specifically designed to mimic earthquake signals will further the effort to produce a generalized phase detection model applicable to any environment.

The first motion polarity training data is reduced to only the high SNR waveforms to ensure correct training labels for an up or down decision (Uchide, 2020). Deep learning models with very deep architectures have been developed using synthetic data for a specific network geometry to predict the focal mechanisms directly from the waveforms (Kuang et al., 2021). This approach is similar to the phase arrival association model here. Our implementation is more generalized and provides the flexibility to determine the appropriate thresholds, e.g., azimuthal coverage or number of misfit polarities, when calculating focal mechanisms for a catalog with precise locations already determined. Community standards, e.g., duration and sample rate, and a comprehensive benchmark data set is still needed for first motion polarities.

4.2 Building the enhanced seismicity catalog

The workflow developed here is designed to facilitate the processing of continuous seismic waveforms at any location, and to study microseismicity and fault structure orientation inferred from focal mechanisms. Efforts to perform detections and event associations simultaneously still require applying a location algorithm and precise relocations (Zhu et al., 2022). The approach implemented here is designed to be flexible and obtain the maximum amount of information from the waveforms during the data processing and build a comprehensive detection table. The processing workflow is designed for multiprocessing and functions with multiple GPU's for rapid processing of large data sets. Building the detection tables is only performed once using the lowest threshold needed in the subsequent processing steps, so the continuous waveforms are only passed through the model one time. The required disk storage for a very large detection table with phase arrival waveforms is negligible when compared to the complete set of network waveforms and allows rapid postprocessing without opening the waveform files. To make the detection table applicable to any location requires training a neural network phase associator with synthetic data for a network geometry of intertest to obtain good event associations from the detections. The model training is straight-forward and allows adding and removing station locations, and adjusting the velocity structure for the area of interest. This workflow allows complete control over the number of detections for an event, the number of collocated p- and s-wave arrivals, and multiple tunable thresholds, e.g., distance, softmax probability, etc., to build an arrival table. The choice of the PhaseLink associator and NonLinLoc location algorithm is applied here, but the detection tables can be applied to any existing or new algorithms.

5 Conclusions

The accuracy of earthquake phase arrival detection and determining the first motion polarity has greatly improved with the application of deep learning as a signal processing tool. Three models are designed and trained to detect microseismicity phase arrivals, predict the first motion polarity for all p-waves, and associate the detections into event arrival tables. Extensive testing is performed to ensure all layers in the deep learning models are contributing to the outcome. Both models implement convolutional filtering to a high-dimensional space and a multi-head attention layer. The models contain many fewer trainable parameters than comparable designs and performs with high accuracy. The detection models have a -0.024 second and -0.011 second average residual for the p- and s-wave arrivals, respectively, for the testing data. The first motion polarity model performs well and the best results are found for the highest SNR signals. The models are implemented in an efficient processing algorithm to utilize multiple CPU's and GPU's for rapid processing of continuous daily waveforms to build a detection table. The workflow is applied to data from continuous waveforms recorded by a temporary deployment of sensors in southern Kansas. The phase association model is trained with synthetic data for the seismic station network geometry for the region of interest. The associated events are located, then a double-difference relocation is performed. The first motion polarity of the arrivals is used to build a focal mechanism catalog. The catalog results show previously undetected widespread activity throughout southern Kansas with more than double the number of events in the first 3 years than previously reported.

Acknowledgments

The authors declare no competing interests. Christopher W. Johnson acknowledges Institutional Support (Laboratory Directed Research and Development) at Los Alamos National Laboratory under grant number 20200681PRD1. Paul A. Johnson acknowledges support by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under grant 89233218CNA000001. The work is available for unlimited release through Los Alamos National Laboratory LA-UR- 22-23463.

Open Research

Seismic waveform data available from Incorporated Research Institutions for Seismology (www.iris.edu/hq/) for stations listed in Supplemental Information. Injection well data is available from the Kansas Geologic Survey (www.kgs.ku.edu/Magellan/Qualified/class2_db). Deep learning modeling was performed using Tensorflow (www.tensorflow.org/). Processing code is under internal review for public release. The catalogs are available in the Supporting Information files.

References

Aguiar, A. C., & Beroza, G. C. (2014). PageRank for Earthquakes. *Seismological Research Letters, 85*(2), 344-350.

Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., et al. (2020). Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking. *Geophysical Research Letters, 47*(16), e2020GL088651.

Chen, J., & Ran, X. (2019). Deep Learning With Edge Computing: A Review. *Proceedings of the IEEE, 107*(8), 1655-1674.

Cochran, E. S., Ross, Z. E., Harrington, R. M., Dougherty, S. L., & Rubinstein, J. L. (2018). Induced Earthquake Families Reveal Distinctive Evolutionary Patterns Near Disposal Wells. *Journal of Geophysical Research: Solid Earth, 123*(9), 8045-8055.

De Angelis, S., & Bodin, P. (2012). Watching the Wind: Seismic Data Contamination at Long Periods due to Atmospheric Pressure-Field-Induced Tilting. *Bulletin of the Seismological Society of America, 102*(3), 1255-1265.

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI—Explainable artificial intelligence. *Science Robotics, 4*(37), eaay7120.

Hammer, C., Beyreuther, M., & Ohrnberger, M. (2012). A Seismic-Event Spotting System for Volcano Fast-Response Systems. *Bulletin of the Seismological Society of America, 102*(3), 948-960.

Hardebeck, J. L., & Michael, A. J. (2006). Damped regional-scale stress inversions: Methodology and examples for southern California and the Coalinga aftershock sequence. *Journal of Geophysical Research, 111*(B11).

Hardebeck, J. L., & Shearer, P. M. (2002). A New Method for Determining First-Motion Focal Mechanisms. *Bulletin of the Seismological Society of America, 92*(6), 2264-2276.

Inbal, A., Cristea-Platon, T., Ampuero, J. P., Hillers, G., Agnew, D., & Hough, S. E. (2018). Sources of Long-Range Anthropogenic Noise in Southern California and Implications for Tectonic Tremor Detection. *Bulletin of the Seismological Society of America*.

Johnson, C. W., Meng, H., Vernon, F., & Ben-Zion, Y. (2019). Characteristics of Ground Motion Generated by Wind Interaction With Trees, Structures, and Other Surface Obstacles. *Journal of Geophysical Research: Solid Earth, 124*.

Johnson, C. W., Vernon, F., Nakata, N., & Ben-Zion, Y. (2019). Atmospheric Processes Modulating Noise in Fairfield Nodal 5 Hz Geophones. *Seismological Research Letters*.

Kuang, W., Yuan, C., & Zhang, J. (2021). Real-time determination of earthquake focal mechanism via deep learning. *Nature Communications, 12*(1), 1432.

Li, Z., Meier, M.-A., Hauksson, E., Zhan, Z., & Andrews, J. (2018). Machine Learning Seismic Wave Discrimination: Application to Earthquake Early Warning. *Geophysical Research Letters, 45*(10), 4773-4779.

Lomax, A., Virieux, J., Volant, P., & Berge-Thierry, C. (2000). Probabilistic Earthquake Location in 3D and Layered Models. In C. H. Thurber & N. Rabinowitz (Eds.), *Advances in Seismic Event Location* (pp. 101-134). Dordrecht: Springer Netherlands.

Magrini, F., Jozinović, D., Cammarano, F., Michelini, A., & Boschi, L. (2020). Local earthquakes detection: A benchmark dataset of 3-component seismograms built on a global scale. *Artificial Intelligence in Geosciences, 1*, 1-10.

Meng, H., Ben-Zion, Y., & Johnson, C. W. (2019). Detection of random noise and anatomy of continuous seismic waveforms in dense array data near Anza California. *Geophysical Journal International*.

Meng, H., & Ben-Zion, Y. (2018). Characteristics of Airplanes and Helicopters Recorded by a Dense Seismic Array Near Anza California. *Journal of Geophysical Research: Solid Earth*.

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*(1), 3952.

Mousavi, S. M., Sheng, Y., Zhu, W., & Beroza, G. C. (2019). STanford EArthquake Dataset (STEAD): A Global Data Set of Seismic Signals for AI. *IEEE Access, 7*, 179464-179476.

Mousavi, S. M., Zhu, W., Sheng, Y., & Beroza, G. C. (2019). CRED: A Deep Residual Network of Convolutional and Recurrent Units for Earthquake Signal Detection. *Scientific Reports, 9*(1), 10267.

Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., et al. (2022). Which Picker Fits My Data? A Quantitative Evaluation of Deep Learning Based Seismic Pickers. *Journal of Geophysical Research: Solid Earth, 127*(1), e2021JB023499.

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances, 4*(2).

Qin, L., Vernon, F. L., Johnson, C. W., & Ben-Zion, Y. (2019). Spectral Characteristics of Daily to Seasonal Ground Motion at the Piñon Flats Observatory from Coherence of Seismic Data. *Bulletin of the Seismological Society of America, 109*(5), 1948-1967.

Reynen, A., & Audet, P. (2017). Supervised machine learning on a network scale: application to seismic event classification and detection. *Geophysical Journal International, 210*(3), 1394-1409.

Ross, Z. E., Cochran, E. S., Trugman, D. T., & Smith, J. D. (2020). 3D fault architecture controls the dynamism of earthquake swarms. *Science, 368*(6497), 1357-1361.

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.

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*(5A), 2894-2901.

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*(1), 856-869.

Rubinstein, J. L., Ellsworth, W. L., & Dougherty, S. L. (2018). The 2013–2016 Induced Earthquakes in Harper and Sumner Counties, Southern Kansas. *Bulletin of the Seismological Society of America, 108*(2), 674-689.

Saad, O. M., Huang, G., Chen, Y., Savvaidis, A., Fomel, S., Pham, N., & Chen, Y. (2021). SCALODEEP: A Highly Generalized Deep Learning Framework for Real-Time Earthquake Detection. *Journal of Geophysical Research: Solid Earth, 126*(4), e2020JB021473.

Shelly, D. R., Mayeda, K., Barno, J., Whidden, K. M., Moschetti, M. P., Llenos, A. L., et al. (2021). A Big Problem for Small Earthquakes: Benchmarking Routine Magnitudes and Conversion Relationships with Coda Envelope-Derived Mw in Southern Kansas and Northern Oklahoma. *Bulletin of the Seismological Society of America, 112*(1), 210-225.

Simpson, R. W. (1997). Quantifying Anderson's fault types. *Journal of Geophysical Research: Solid Earth, 102*(B8), 17909-17919.

Skoumal, R. J., Cochran, E. S., Kroll, K. A., Rubinstein, J. L., & McPhillips, D. (2021). Characterizing Stress Orientations in Southern Kansas. *Bulletin of the Seismological Society of America, 111*(3), 1445-1454.

Trugman, D. T., & Shearer, P. M. (2017). GrowClust: A Hierarchical Clustering Algorithm for Relative Earthquake Relocation, with Application to the Spanish Springs and Sheldon, Nevada, Earthquake Sequences. *Seismological Research Letters, 88*(2A), 379-391.

Uchide, T. (2020). Focal mechanisms of small earthquakes beneath the Japanese islands based on first-motion polarities picked using deep learning. *Geophysical Journal International, 223*(3), 1658-1671.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention Is All You Need. *arXiv 1706.03762*.

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., et al. (2022). SeisBench—A Toolbox for Machine Learning in Seismology. *Seismological Research Letters*.

Woollam, J., Rietbrock, A., Bueno, A., & De Angelis, S. (2019). Convolutional Neural Network for Seismic Phase Classification, Performance Demonstration over a Local Seismic Network. *Seismological Research Letters, 90*(2A), 491-502.

Yeck, W. L., Patton, J. M., Ross, Z. E., Hayes, G. P., Guy, M. R., Ambruz, N. B., et al. (2020). Leveraging Deep Learning in Global 24/7 Real-Time Earthquake Monitoring at the National Earthquake Information Center. *Seismological Research Letters, 92*(1), 469-480.

Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances, 1*(11).

Zhou, Y., Yue, H., Kong, Q., & Zhou, S. (2019). Hybrid Event Detection and Phase-Picking Algorithm Using Convolutional and Recurrent Neural Networks. *Seismological Research Letters, 90*(3), 1079-1087.

Zhu, W., & Beroza, G. C. (2018). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International, 216*(1), 261-273.

Zhu, W., Tai, K. S., Mousavi, S. M., Bailis, P., & Beroza, G. C. (2022). An End-To-End Earthquake Detection Method for Joint Phase Picking and Association Using Deep Learning. *Journal of Geophysical Research: Solid Earth, 127*(3).

Journal of Geophysical Research: Solid Earth

Supporting Information for

EQDetect: Earthquake phase arrivals and first motion polarity applying deep learning

Christopher W. Johnson[1] and Paul A. Johnson[1]

[1]Los Alamos National Laboratory, Geophysics Group, Los Alamos, N.M., USA

Corresponding author: Christopher W Johnson (cwj@lanl.gov)

Contents of this file

Figures S1 to S4
Table S1

Additional Supporting Information (Files uploaded separately)

Captions for Data Set S1 to S2
Captions for Movies S1

Introduction

Supporting figures referenced in the text are provided. The compiled seismicity and focal mechanism catalogs are provided. An animation showing the spatial temporal seismic activity is provided.

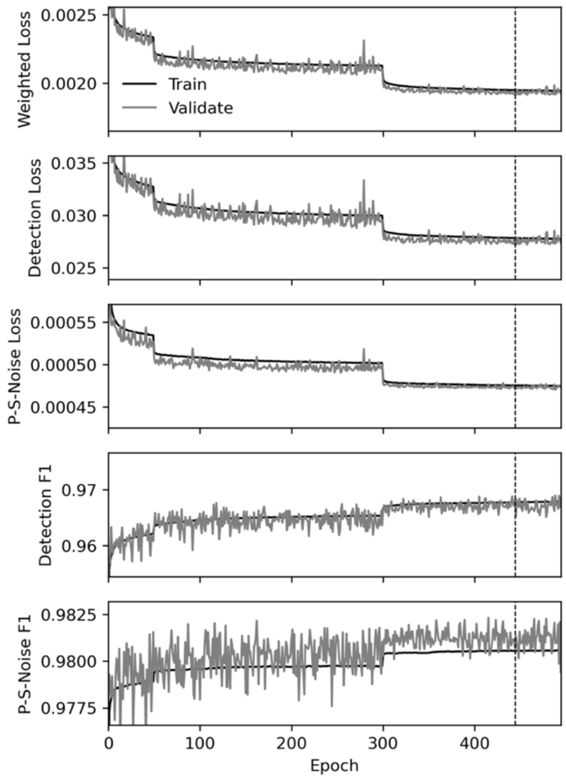

Figure S1. Earthquake detection model training loss function metrics. Black curve is the training data metric and the gray curve is for the validation data. The vertical black dashed line is epoch 445 with the lowest validation loss function value.

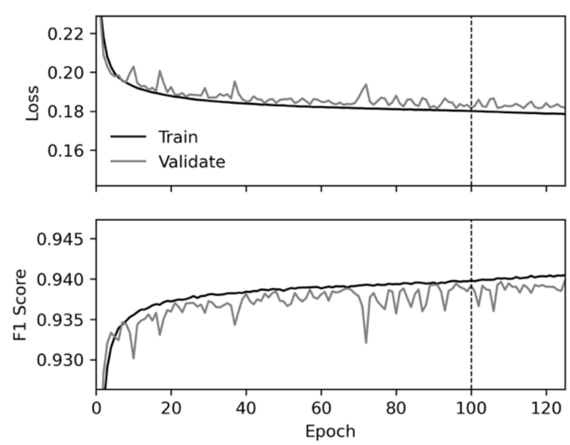

Figure S2. Polarity model training loss function and F1 score. Vertical dashed line is at epoch 100 with the lowest validation data loss value and no improvement is observed for 25 epochs.

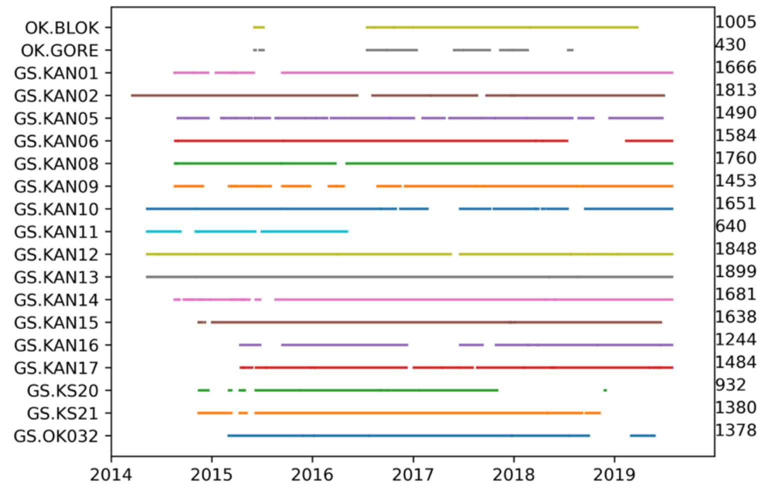

Figure S3. Waveform data available from 2014 to 2020 for the 17 broadband and 2 accelerometer stations used from the GS and OK networks. The left axis shows the network and station with the total number of daily files listed in the right axis. This includes days with missing data but a waveform record does exist for that day.

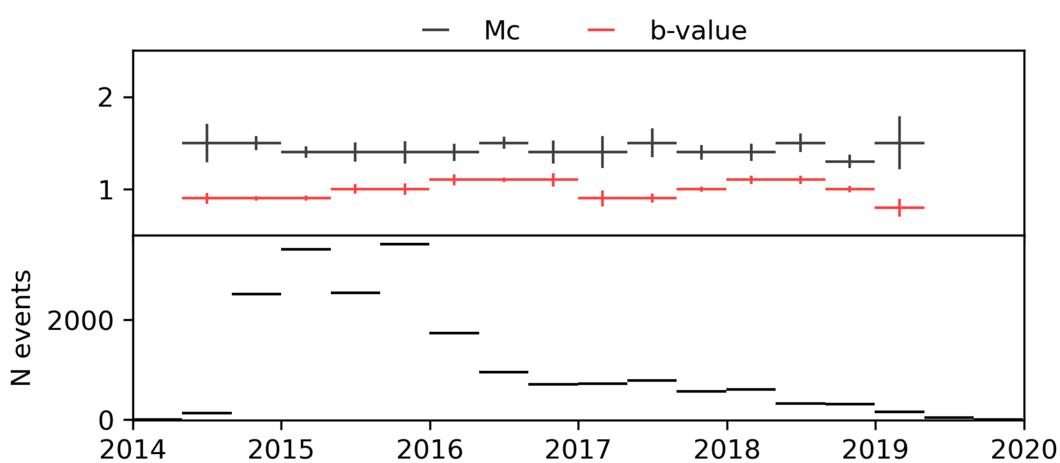

Figure S4. Moving window magnitude of completeness and b-value.

KAN01	37.15342	-97.75897
KAN05	37.10865	-97.87228
KAN06	37.24800	-97.85860
KAN08	37.22672	-97.97094
KAN09	37.13613	-97.61832
KAN10	37.12350	-98.09513
KAN11	37.20596	-97.91330
KAN12	37.29738	-97.99800
KAN13	37.01288	-97.47780
KAN14	36.95682	-97.96302
KAN16	37.22561	-98.06471
KAN17	37.04407	-97.76475
KS20	37.22973	-97.55432
KS21	37.28649	-97.66302
OK32	36.80382	-98.21041
BLOK	36.76061	-97.21502
GC02	36.85150	-97.85959
GORE	36.78563	-97.94706
KAN02	37.19797	-97.87939

Table S1. Station identifier and geographic location.

Data Set S1. Seismicity catalog in Growclust output format for study area shown in Figure 8.

Data Set S2. Focal mechanism catalog for study area shown in Figure 10.

Movie S1. Animation showing the map view, longitude vs. depth, and latitude vs. depth for the study area.