Evaluating Automated Seismic Event Detection Approaches: An Application to Victoria Land, East Antarctica

Long Minh Ho[1], Jacob I. Walter[2], Samantha E Hansen[3], José Luis Sánchez-Roldán[4], and Zhigang Peng[5]

[1]University of Alabama
[2]University of Oklahoma
[3]The University of Alabama, Tuscaloosa
[4]Complutense University of Madrid
[5]Georgia Institute of Technology

March 13, 2024

Abstract

As seismic data collection continues to grow, advanced automated processing techniques for robust phase identification and event detection are becoming increasingly important. However, the performance, benefits, and limitations of different automated detection approaches have not been fully evaluated. Our study examines how the performance of conventional techniques, including the Short-Term Average/Long-Term Average (STA/LTA) method and cross-correlation approaches, compares to that of various deep learning models. We also evaluate the added benefits that transfer learning may provide to machine learning applications. Each detection approach has been applied to three years of seismic data recorded by stations in East Antarctica. Our results emphasize that the most appropriate detection approach depends on the data attributes and the study objectives. STA/LTA is well-suited for applications that require rapid results even if there is a greater likelihood for false positive detections, and correlation-based techniques work well for identifying events with a high degree of waveform similarity. Deep learning models offer the most adaptability if dealing with a range of seismic sources and noise, and their performance can be enhanced with transfer learning, if the detection parameters are fine-tuned to ensure the accuracy and reliability of the generated catalog. Our results in East Antarctic provide new insight into polar seismicity, highlighting both cryospheric and tectonic events, and demonstrate how automated event detection approaches can be optimized to investigate seismic activity in challenging environments.

Evaluating Automated Seismic Event Detection Approaches: An Application to Victoria Land, East Antarctica

L.M. Ho[1], J.I. Walter[2], S.E. Hansen[1], J.L. Sánchez-Roldán[3], and Z. Peng[4]

[1]The University of Alabama, Tuscaloosa, AL 35487

[2]Oklahoma Geological Survey, University of Oklahoma, Norman, OK 73019

[3]Universidad Complutense de Madrid, Madrid, Spain, 28040

[4]Georgia Institute of Technology, Atlanta, GA 30332

Corresponding author: Long Ho (lmho@crimson.ua.edu)

Key Points:

- Deep learning models, enhanced by transfer learning, adapt well to varied seismic sources.
- Automated detection approaches offer insights into both cryospheric and tectonic events in Antarctica.
- Even in regions with limited station coverage, automated detection approaches can help us develop more complete seismicity catalogs.

Abstract

As seismic data collection continues to grow, advanced automated processing techniques for robust phase identification and event detection are becoming increasingly important. However, the performance, benefits, and limitations of different automated detection approaches have not been fully evaluated. Our study examines how the performance of conventional techniques, including the Short-Term Average/Long-Term Average (STA/LTA) method and cross-correlation approaches, compares to that of various deep learning models. We also evaluate the added benefits that transfer learning may provide to machine learning applications. Each detection approach has been applied to three years of seismic data recorded by stations in East Antarctica. Our results emphasize that the most appropriate detection approach depends on the data attributes and the study objectives. STA/LTA is well-suited for applications that require rapid results even if there is a greater likelihood for false positive detections, and correlation-based techniques work well for identifying events with a high degree of waveform similarity. Deep learning models offer the most adaptability if dealing with a range of seismic sources and noise, and their performance can be enhanced with transfer learning, if the detection parameters are fine-tuned to ensure the accuracy and reliability of the generated catalog. Our results in East Antarctic provide new insight into polar seismicity, highlighting both cryospheric and tectonic events, and demonstrate how automated event detection approaches can be optimized to investigate seismic activity in challenging environments.

Plain Language Summary

Given the large quantity of seismic data recorded for geologic investigations, the manual identification of earthquake arrivals is becoming less feasible, and automated detection approaches are becoming increasingly important. However, the benefits and limitations of different detection techniques have not been fully evaluated. We examine a range of automated detection approaches, applied to data recorded by seismic stations in Antarctica, to assess the performance of each method. Additionally, an approach called transfer learning is examined to determine if it can improve the accuracy and reliability of the automated detections. Our results highlight new seismic events in Antarctica, providing insights into both geologic processes and ice-sheet behavior.

1. Introduction

The accurate creation of earthquake catalogs for seismotectonic interpretation requires robust seismic phase identification, event association, and event detection; however, with the ever-increasing availability of seismic data, manual processing by human analysts is becoming less feasible. As such, automated processing techniques are becoming increasingly important. Some event detection techniques, such as the Short-Term Average/Long-Term Average (STA/LTA) method (Allen, 1978; Earle & Shearer, 1994), use relatively simple algorithms and provide rapid results without the need for extensive data pre-processing. Waveform based cross-correlation approaches, such as the matched filter (MF) technique (Gibbons & Ringdal, 2006; Peng & Zhao, 2009; Shelly et al., 2007), can also be applied to STA/LTA generated earthquake catalogs to identify new, closely located events with similar focal mechanisms to those in the initial catalog. However, STA/LTA may not perform well for low signal-to-noise ratio (SNR) data, and cross-correlation based approaches can sometimes generate spatially biased event catalogs (Herrmann & Marzocchi, 2021; Schaff & Beroza, 2004; Yoon et al., 2015). The shortcomings of these methods can also sometimes result in impulsive transient signals or distant regional/teleseismic signals being erroneously identified as local earthquakes (*e.g.*, Meng et al., 2012). In some cases, these challenges can be overcome using phase association algorithms, which analyze triggers from multiple stations to determine whether any combination displays arrival time sequences that align with characteristic seismic event patterns (Myers et al., 2007).

In recent years, advancements in machine learning techniques, coupled with the democratization of open-source software, have provided more sophisticated methods to automatically detect seismic events. In particular, convolutional neural networks (CNN), which perform a sequence of convolution, resampling, and non-linear transformations on raw waveform data, have shown promising results (Perol et al., 2018; Ross et al., 2018; Wu et al., 2018; Zhou et al., 2019; Zhu et al., 2019) when compared to more traditional techniques (Earle & Shearer, 1994; Gibbons & Ringdal, 2006; Peng & Zhao, 2009; Shelly et al., 2007). CNN pickers are designed to provide the added advantage of identifying body wave phases on three-component seismograms, thereby simplifying earthquake association and relocation. However, machine learning algorithms are complex, computationally demanding, and typically require optimization to avoid false-positive event detections.

To date, only a few studies have evaluated the performance of different automated detection approaches with respect to one another or have attempted to combine detection techniques to achieve the best possible outcome (Münchmeyer et al., 2022; Neves et al., 2024; Si et al., 2024; Woollam et al., 2022; Yuan et al., 2023). Further, most of these previous studies have typically only examined select model pairs based on one or a few training datasets (*e.g.*, Han et al., 2023; Jiang et al., 2021; Perol et al., 2018; Vaezi & Van der Baan, 2015), and they largely focus on small magnitude, tectonic-related seismic events. Here, we compare the benefits and limitations of the STA/LTA technique (Earle & Shearer, 1994), the cross-correlation-based MF approach (Peng & Zhao, 2009), and a suite of deep learning models, including EQTransformer (EQT, Mousavi et al., 2020), PhaseNet (Zhu & Beroza, 2019), BasicPhaseAE (Woollam et al., 2019), and the Generalized Phase Detection (GPD) model (Ross et al., 2018). We also update the deep learning models with additional training data derived from this project, a process known as transfer learning. Despite the potential for transfer learning to enhance model adaptability and efficiency (Chai et al., 2020; Lapins et al., 2021), particularly in data-scarce environments, its adoption in seismic studies has not been as rapid or as extensive as in other domains of deep learning research. This gap presents an opportunity to investigate the full capabilities of transfer learning in automatic event detection. We test the performance of the updated versus original deep-learning models using a range of metrics that evaluate each of their abilities to accurately determine the onset time of phase arrivals, to reliably classify phases as P- or S-waves, and to identify events while minimizing the number of false positives. These techniques are applied to a unique set of waveforms that contain a mixture of tectonic earthquake signals and seismic events generated by glacial movement (*e.g.*, icequakes). Collectively, our evaluation allows us to assess the efficacy of each algorithm when applied to a complex dataset.

2. Data and Methods

Broadband seismic deployments across the Antarctic continent have dramatically increased over the past several decades (*e.g.*, Anandakrishnan et al., 2000; Anthony et al., 2015; Hansen et al., 2015; Heeszel et al., 2013; Pyle et al., 2010), providing a valuable and challenging test dataset for automatic event detection. Seismic events in Antarctica are not only associated with tectonic sources (*e.g.*, Lough et al., 2013, 2018; Rowe et al., 2000) but are also caused by other natural phenomena, such as iceberg calving signals (*e.g.*, Chen et al., 2011; Riel et al., 2021; Winberry et al., 2020; Zoet et al., 2012) or ice-stream slip (*e.g.*, Guerin et al., 2021; Hudson et al., 2023; Nettles & Ekström, 2010; Winberry et al., 2014; Walter et al., 2011, 2015), which are collectively classified as icequakes. Our study focuses on a subset of seismic data recorded by 19 stations deployed in the Victoria Land region of East Antarctica (Fig. 1), which provide continuous seismic recordings for several years. Most of these stations (15) were part of the Transantarctic Mountains Northern Network (TAMNNET), which operated between 2012-2015 (Hansen, 2012; Hansen et al., 2015); however, we also incorporated data from two additional networks (ER, GT; Fig. 1; ASL/USGS, 1993). This dataset allows us to provide unique constraints on polar seismic activity and to evaluate automated event detection performance in a region with limited station coverage.

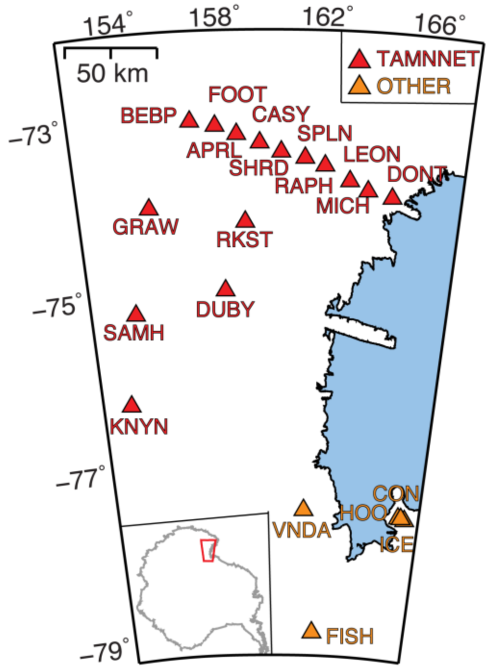

Figure 1. Map highlighting the examined seismic stations in Victoria Land, East Antarctica. Red triangles denote TAMNNET stations (Hansen et al., 2015), and orange triangles denote stations from other networks. Station names are also provided. The location of the main map in relation to the rest of Antarctica is highlighted in the inset on the lower left.

We developed a comprehensive workflow to assess the performance of different automated event detection techniques (Fig. 2). The continuous waveforms recorded by the East Antarctic stations (Fig. 1) were used to develop three starting catalogs, based on the STA/LTA, MF, and EQT machine learning approaches, respectively. Each catalog was then used to fine-tune a series of deep learning models via transfer learning, and their performance was evaluated with various metrics. The fine-tuned detection approach that worked best for our Antarctic dataset was then applied to update the three catalogs, and the events were relocated using a uniform velocity model. Each analysis step is described in detail in the following sections.

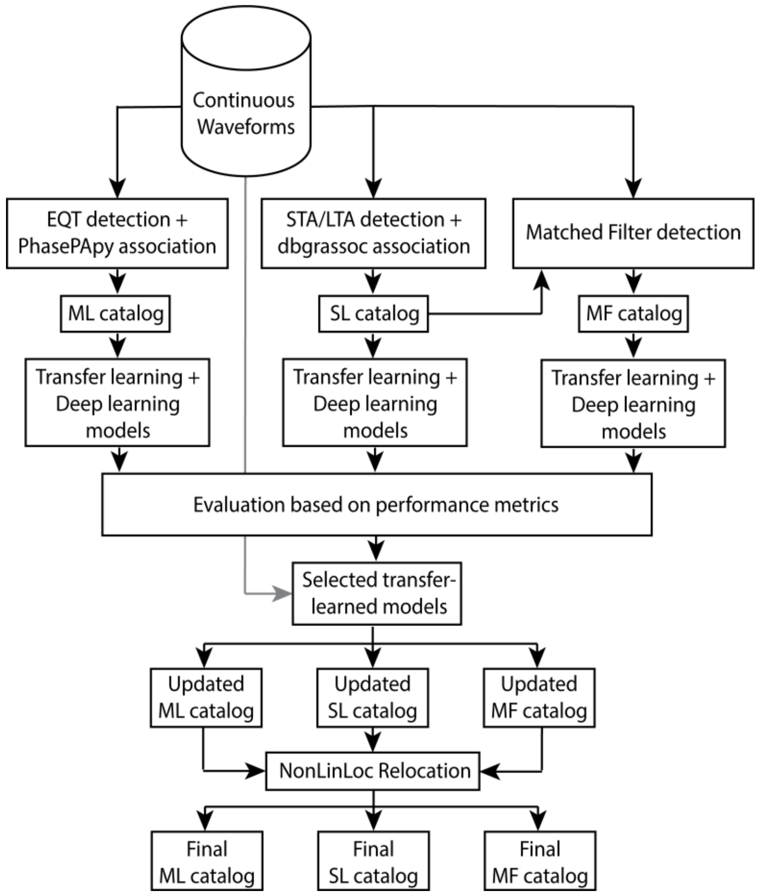

Figure 2. Flowchart summarizing the different automated seismic detection techniques examined in our study and the associated analysis steps.

3. Automated Detection Approaches

As shown in Figure 2, three different automated event detection approaches were initially evaluated by our study, including the STA/LTA method, the MF technique, and a machine learning-based approach using the EQT algorithm. The following subsections highlight the contributions and limitations of each approach as they are applied to our East Antarctic dataset (Fig. 1).

3.1. STA/LTA Method (SL Catalog)

The STA/LTA method (Allen, 1978; Earle & Shearer, 1994) detects high-frequency events in continuous data by identifying signals that have a mean energy ratio above some specified threshold. The STA window contains the dominant frequency of the events the algorithm aims to detect, while the LTA window contains mostly background noise, which should exceed the period of the lowest frequency seismic signal of interest (Trnkoczy, 2009). In continuous data, a trigger is declared when the STA/LTA ratio at any sample point surpasses a pre-defined threshold, indicating that an event is possibly occurring (Allen, 1978; Baer & Kradolfer, 1987). The algorithm remains in this triggered state until the ratio decreases below a specified trigger-release threshold (Fig. 3). One of the strengths of the STA/LTA method is that it does not require any prior knowledge about an event's waveform nor its source (Yoon et al., 2015); however, it does have limitations. For instance, S-waves may not be accurately detected if they arrive within the P-wave coda, and this can be problematic because S-waves are important when trying to determine the depth and origin time for an earthquake. The STA/LTA method is also highly sensitive to the level of noise in the data, and it may not perform well with dense earthquake sequences and/or emergent arrivals (Schaff & Beroza, 2004).

For our study, we designated short-term and long-term window lengths of 0.5 and 8.0 s, respectively. We also set the SNR trigger and trigger-release thresholds to 5 and 2.5, respectively (Fig. 3). Detections were associated with the Antelope dbgrassoc association module (BRTT, 2011), using a pre-computed travel-time grid based on the IASP91 reference velocity model (Kennett & Engdahl, 1991), and events were declared if they were recorded by at least four stations. Between 2012-2014, 560 events were detected using the STA/LTA approach and automatic association, thereby forming our SL catalog (Fig. 2). The data were then bandpass filtered between 2-5 Hz to highlight the signals of interest, and all phase arrivals were manually reviewed and adjusted, as needed. These additional processing steps allowed us to refine our SL catalog of high-quality events with well-determined phase arrivals.

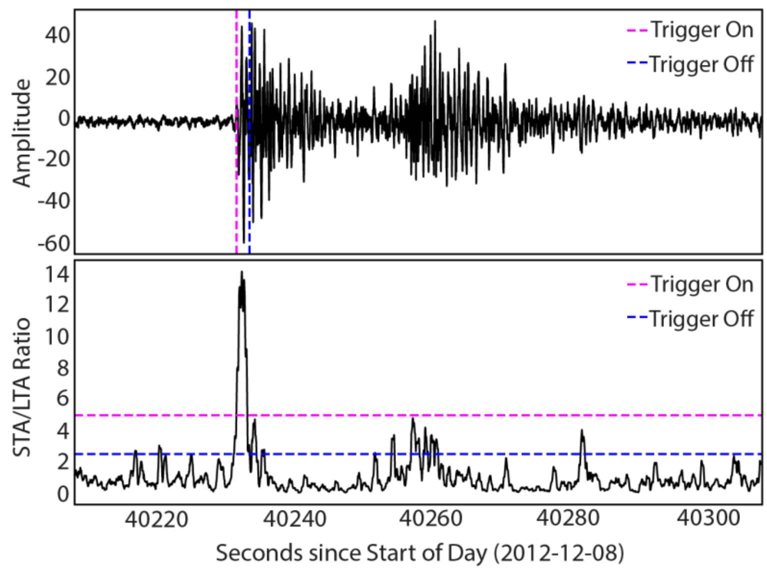

Figure 3. Example illustrating STA/LTA detection thresholds. The upper panel shows an event waveform that was detected by the STA/LTA approach, and the lower panel shows the STA/LTA ratio for the triggered event. Pink lines denote the trigger threshold (5) and trigger time; blue lines denote the trigger release threshold (2.5) and corresponding time.

3.2. Matched Filter Approach (MF Catalog)

The MF technique, also known as template matching or network-based waveform cross-correlation (Gibbons & Ringdal, 2006; Peng & Zhao, 2009; Shelly et al., 2007), provides another approach to automatically detect seismic arrivals, which is based on waveform similarity. Pre-defined template waveforms are cross-correlated with continuous data over successive windows, and signals exceeding a specified correlation threshold are identified as detections (Fig. 4). Generally, the MF approach performs better than the STA/LTA method (Sect. 3.1) when dealing with low SNR data. However, since the template events are often manually determined, the MF method can be time consuming during its initial stages when building the template catalog (if one does not already exist from a regional seismic network or other source). Furthermore, since the approach relies on waveform similarity, seismic signals that differ significantly from the template events may go undetected, leading to an incomplete catalog (Cianetti et al., 2021; Li et al., 2018; Yoon et al., 2015).

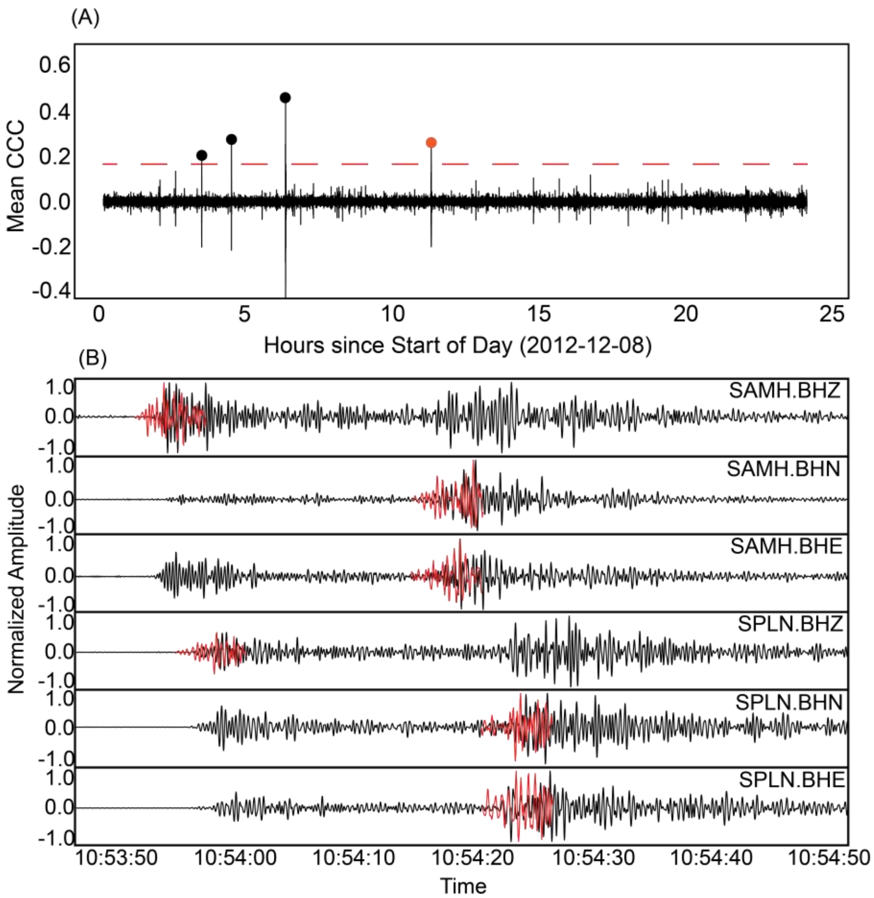

Figure 4. (A) Mean cross-correlation coefficients (CCC) determined by matching a template event, which occurred at 06:13:14 on 2012-12-08, against a full day (2012-12-08) of continuous data. Dots denote detections whose CCC values exceed the detection threshold, which is twelve times the MAD (red dashed line). The orange dot marks the detected event shown in panel (B). (B) Examples illustrating waveform cross-correlation. Template waveforms (red) are plotted on top of the continuous data (black), highlighting detected events from the MF approach. Station names and components are indicated on the right. Amplitudes have been normalized so their absolute maximum values are equal to one. This was done to better illustrate the waveform comparisons.

Using EQcorrscan (Chamberlain et al., 2018), all identified events in the SL catalog were treated as template events (Fig. 2), which were cross-correlated with the bandpass filtered (2-5 Hz) continuous data to identify additional seismic signals (Fig. 4). This bandpass was chosen based on close examination of the template coda, the density of seismic stations in the region, as well as our prior experience working with Antarctic data, where higher frequency information can become scattered by the ice sheet (Bentley & Kohnen, 1976) and thus incoherent when attempting template matching. Each template event was defined by the portion of the waveform 0.5 s before the event's P-wave arrival and 6 s after its S-wave arrival (Peng et al., 2014). The templates were shifted by 0.025 s (1 sample) increments through the continuous waveforms, and correlation coefficients were computed for each increment. Mean correlation coefficients were then determined by stacking the coefficient values across all stations and components (Fig. 4). The relative quality of each cross-correlated, matched waveform was evaluated using the median absolute deviation (MAD; Shelly et al., 2007), which is a measure of dispersion calculated as the median of the absolute difference between each data point for the mean correlation coefficient. The MAD value helps to estimate the variability in data distribution due to uncorrelated noise, thereby providing a robust measure to identify outliers. For a normally distributed dataset, the standard deviation is 1.4826 times the MAD (Hampel, 1974). Due to the noisy nature of real seismic data and the relatively long-period bandpass chosen for this project, a conservative threshold of 12 times the MAD was chosen, and signals that exceed this MAD value are identified as positive event detections (Fig. 4; *e.g.*, Skoumal et al., 2015; Yao et al., 2021).

A time domain, phase-pick SNR threshold was also applied to further ensure robust detections. For a given phase, the SNR was calculated by taking the maximum amplitude of the signal window and dividing it by the root-mean-square of the noise window. The noise windows start 6 s before the phase of interest, and both the signal and noise windows had lengths of 5.5 s (Fig. S1 in Supporting Information). The SNR threshold was subsequently determined by comparing the pick-specific SNR values obtained from all detected picks for each seismic event.

This additional processing step is not only important for robust event detections, but it also helps to remove unwanted signals, such as teleseismic events that originate from distant earthquakes. Sometimes teleseismic signals can be mistakenly detected in MF catalogs for local events, and this can adversely affect the accuracy of local event detections because teleseismic events have unique seismic waves and frequency contents (Waldhauser & Schaff, 2007). We determined that maintaining a SNR greater than 2.0 for both the P and S picks (Fig. S1 in Supporting Information) effectively helps to limit the influence of teleseismic events and reduces the number of false detections. With the MAD and SNR criteria applied, our MF catalog includes 4,577 local events (Fig. 2).

3.3. Machine Learning Approach (ML Catalog)

In addition to the STA/LTA and the MF techniques, we also utilized EQT, a machine learning-based signal detector and phase picker that was trained on a diverse seismic dataset (Mousavi et al., 2020). Further details about EQT and its architecture are provided in Section 4.1. We implemented the EQT picker within the easyQuake Python package (Walter et al., 2021) to identify P- and S-wave picks within the continuous data. The easyQuake associator, which is a modified version of PhasePApy (Chen & Holland, 2016), was used to aggregate pick information and declare event detections. Probability thresholds of 0.1, 0.1, and 0.3 were specified for the P-wave picks, S-wave picks, and event detections, respectively. In total, 1,728 events were detected in the East Antarctic dataset, which compose our initial machine learning (ML) catalog (Fig. 2). It should be noted that this catalog is distinguished from those derived from transfer-learning in later sections because it was generated using phase picks that were based on the original model and parameters specified by Mousavi et al. (2020).

4. Transfer Learning

Each of the catalogs described in Sections 3.1-3.3 were used in a transfer learning process to adapt a series of pre-trained deep learning models. Instead of retraining an entire model from scratch with randomly initialized parameters or different model architecture, a strategy called fine-tuning is employed, where the original model and its architecture serve as the starting point, and training continues with newly added data, thereby refining the model (Pan & Yang, 2010). Transfer learning not only leads to better model performance, but it also overcomes some of the limitations of traditional models that assume training and testing datasets are independent and are identically distributed (Tan et al., 2018).

The effectiveness of transfer learning has been proven in various fields (Long et al., 2013, 2015; Pan et al., 2011), and while its adoption within the field of seismology has been relatively limited so far, the technique demonstrates promising potential. For instance, Zhu et al. (2019) used a CNN-based Phase-Identification Classifier (CPIC), which was initially trained on a dataset with 30,146 labeled phases from the aftershock sequences of the 2008 M_w 7.9 Wenchuan earthquake, to develop a more complete aftershock catalog for the same area. Additionally, when fine-tuned on a smaller dataset from Oklahoma, the CPIC achieved 97% accuracy. This study highlights the potential for transfer learning applications to identify events in regions with no or few labeled phases. In a different study, Chai et al. (2020) enhanced the capabilities of the PhaseNet model (Zhu & Beroza, 2019), which was originally trained on data from regional seismic networks, to efficiently handle microseismic data from South Dakota. About 3,600 three-component seismograms and associated manual picks were used in the transfer learning process, and the performance of the retrained model exceeded that of the original PhaseNet model by over 10% in terms of precision and recall (see Sect. 4.3). Compared to human expert detections, 32% fewer P-wave picks were made, but the fine-tuned model identified 48% more S-wave picks.

We implemented our transfer learning process with Seisbench, a toolbox for machine learning in seismology (Ho, 2024; Münchmeyer et al. 2022; Woollam et al., 2022). Various deep learning model architectures were utilized, including PhaseNet (Zhu & Beroza, 2019), BasicPhaseAE (Woollam et al., 2019), GPD (Ross et al., 2018), and EQT (Mousavi et al., 2020), which are more fully described in Section 4.1. These models were selected given their distinct yet interrelated approaches to seismic signal processing. Additionally, these models share a common approach in terms of pre-processing the seismic data. Regardless of their specific architectures or use cases, they all rely on uniformly sampled data, typically at 100 Hz. If the original data has a different sampling rate, it is resampled to ensure uniformity. The data windows used by these models vary in length, but they all incorporate multiple types of seismic signals, including P-waves, S-waves, and noise, within their respective networks.

4.1 Deep Learning Models

The PhaseNet CNN (Zhu & Beroza, 2019) was developed as a U-Net architecture, which functions as an encoder-decoder mechanism that pulls significant features from input data and subsequently expands them to generate predictions of equivalent size outputs (Ronneberger et al., 2015). While the U-Net was initially created for a broad range of image processing applications, this approach has been adapted for earthquake phase detection. Three-component seismograms are sampled using 30 s windows that include both P- and S-wave arrivals, and these samples serve as the input for PhaseNet. The waveform data are then processed through an iterative down-sampling and up-sampling procedure. During down-sampling, the encoder reduces the dimensionality of the raw seismic data and extracts essential features associated with the seismic phase arrivals. The condensed information provided by the encoder is then increased in dimensionality through up-sampling by the decoder, which converts the information into detailed probability distributions for P-waves, S-waves, and noise at each point in time (Goodfellow et al., 2016; Zhu & Beroza, 2019). For seismic applications, PhaseNet was originally trained and evaluated using 779,514 waveforms containing labeled P- and S-wave arrivals from local earthquakes recorded in northern California (Zhu & Beroza, 2019).

BasicPhaseAE, which is another U-Net-like CNN phase detector, employs three 6 s input windows, with each window sampling an individual component (Woollam et al., 2019). The structure of BasicPhaseAE is similar to PhaseNet, but it differs in a few aspects. BasicPhaseAE uses smaller filter sizes and omits convolutions without stride, which refers to the step size that the filter matrix moves across the input matrix during the convolution process. In addition, BasicPhaseAE lacks residual connections, which are essentially shortcuts or bypass routes that enable the gradient to be back-propagated directly to earlier layers (Woollam et al., 2019; Münchmeyer et al., 2022). The input data, which consists of labels or classes of seismic data (*e.g.*, P-waves, S-waves, noise), undergo several transformations. Convolutional operations first extract the characteristic features for each class. During training, the model uses a designated 6 s window of data that is then divided into sequential sub-windows, each 0.4 s in length. The sub-windows are randomly shuffled to prevent the CNN from learning irrelevant temporal patterns. Extracted features then undergo multiple resampling stages, with a rectified linear unit activation function applied at each stage. The final architecture comprises three convolutional layers and three up-sampling layers. The network ultimately determines the probability of a P-wave, S-wave, or noise for every time sample in the input window. BasicPhaseAE was initially trained and evaluated using 11,000 waveforms from earthquakes located within the Iquique region in northern Chile (Woollam et al., 2019).

The GPD model is a phase identification CNN with six layers, including four convolution layers and two fully connected layers (Ross et al., 2018). Rectified linear units serve as the activation function for each layer, and batch normalization is applied throughout. GPD operates on a short 4 s input window that advances five samples (0.05 s) after each prediction to create a new, slightly overlapped 4 s window for the next prediction (Münchmeyer et al., 2022). Each advanced window is then classified as a P-wave arrival, S-wave arrival, or noise. The GPD model was originally trained and evaluated using 4.5 million three-component seismic records, evenly distributed amongst P- and S-wave seismograms and noise (Ross et al., 2018). Using a multi-class cross-entropy loss for training, the GPD model has been shown to effectively detect and identify seismic phases in various datasets (Münchmeyer et al., 2022; Woollam et al., 2022).

EQT is a model designed for simultaneous seismic event detection, phase identification, and onset timing determination. This model was originally trained on a portion of the STEAD dataset (Mousavi et al., 2019), a global collection of 1.2 million hand-labeled earthquake and noise waveforms. EQT operates on 60 s windows of three-component seismic data. Its architecture comprises a deep encoder and three separate decoders, and it integrates convolution, long short-term memory (LSTM) units, residual connections, and attention mechanisms (Mousavi et al., 2020). The encoder processes the seismic data into high-level contextual representations, while the decoders convert these representations into probability sequences for events as well as for P- and S-wave detections. LSTM, which resembles human auditory memory processing, and attention mechanisms, which simulate selective focusing in high-resolution areas, work in tandem to enhance the model's performance (Gers et al., 1999). The attention mechanisms function on two levels: global for earthquake events and local for phases within those events. During training, EQT employs data augmentation techniques, such as adding Gaussian noise, introducing gaps, and removing channels, which are implemented to enhance the model's robustness, teaching it how to handle various real-world data imperfections and irregularities. This helps to improve its overall performance and generalization ability (Mousavi et al., 2020).

Each of the above models has a different level of complexity, adaptability, and suitability for seismic datasets. For example, since BasicPhaseAE lacks residual connections, which are shortcuts that skip one or more layers to help train deep neural networks, its learning efficiency may be lower compared to PhaseNet (Münchmeyer et al., 2022). Compared to EQT, GPD is much slower, but it requires less memory. Further, the sophisticated EQT architecture and its comprehensive functionality may require more computational resources for complex analyses. We evaluate the performance of each model in relation to one another using our East Antarctic catalogs described in Sections 3.1-3.3, but it should be emphasized that the most suitable model for a given investigation depends on the type of data, the available processing time, and the computational resources available. We did not evaluate the relative computational performance of the specific algorithms in this study.

4.2 Applying Transfer Learning to the East Antarctic Catalogs

Each of the pre-trained models described in the previous section were fine-tuned via transfer learning using each of the event catalogs (Sects. 3.1-3.3). The SL, MF, and ML catalogs contain a total of 1,536, 13,731, and 5,388 waveform segments, respectively. The metadata for each catalog were assembled into a QuakeML-formatted file, and we also developed HDF5-formatted files by combining the event metadata with the waveforms, similar to the STEAD dataset format (Mousavi et al., 2019), for inclusion into Seisbench (Ho, 2024; Woollam et al., 2022). Each catalog was divided into a training subset, which is composed of 70% of the data, a validation subset, which contains 15% of the data, and a testing subset, which includes the remaining 15% of the data. The training subset was used to adjust the model's weights and biases during the transfer learning process, while the validation subset was used to fine-tune the model's hyperparameters. The validation subset was also essential in determining which model iteration performed the best, using the parameters described in Section 4.3. Once the optimal model configuration was identified based on the validation subset's results, the updated model was then evaluated on the testing subset. The final, reported results (Section 5) are based on this evaluation of the testing subset, thereby ensuring an unbiased assessment of each models' performance on unseen data.

Using the Münchmeyer et al. (2022) data augmentation techniques within SeisBench (Woollam et al., 2022), we built training pipelines, which are a series of steps that prepare and transform the waveform data for model training. Since our waveforms are long compared to each aforementioned model input length, a two-step approach was employed for window selection. First, for two-thirds of the training subset, windows were selected to ensure that they contained at least one labeled pick. For the remaining one-third, the windows were randomly selected from the entire waveform, and they may or may not include labeled picks. This approach guarantees that the training subsets are not overwhelmed by noise samples, which is particularly important for models with short input windows (*e.g.*, PhaseNet, BasicPhaseAE, GPD). The same approach was also applied to the validation subset.

Additionally, as part of the transfer learning process for each catalog, we employed the Adam optimizer (Kingma & Ba, 2014), which efficiently updates the model parameters to minimize the error between predicted and actual values. A corresponding learning rate of 0.001 was selected, which controls the magnitude of changes made to the model parameters during updates and ensures a steady convergence without overshooting (*i.e.*, where the model might skip over the optimal parameters). Further, a batch size of 256 was used in the optimizer, which means that 256 training samples were processed together during each iteration. This helps to balance computational efficiency and the quality of the model's gradient estimation (Coleman et al., 2017; Smith, 2018). Early stopping was also employed to obtain an optimal model. This strategy halts the training when the validation loss (a measure of prediction error) throughout the entire training subset fails to improve after ten successive cycles (epochs).

4.3. Evaluating Model Performance

To evaluate each fine-tuned, deep learning model’s ability to differentiate between seismic events and noise, we adopted the approach of Münchmeyer et al. (2022). First, a 30 s window of a random seismic waveform from either the validation or testing subset is analyzed to determine if it contains an event onset (*i.e.*, a first arriving seismic wave). Noise samples are also extracted from the window using labeled noise traces, if present. Otherwise, the noise sample is defined based on the presence or absence of P-wave and S-wave arrivals. That is, windows containing neither P- nor S-wave arrivals are labeled as noise, while those with either or both are labeled as an event. The event and noise labels were used as “ground truth” to compare with our models’ predictions.

A variety of metrics are used to evaluate the performance of each model. First, to assess a model’s ability to accurately identify event onsets while minimizing false positives, we examined the receiver operating characteristics (ROC), the area under the curve (AUC), and the F1 score. The ROC describes the true and false positive rates across all possible detection thresholds, allowing for different trade-offs between these rates, depending on the application scenario (Fawcett, 2006). For example, in early earthquake warning systems, a high true positive rate is important to ensure timely alerts, even if it means getting some false alarms (Meier et al., 2020). Alternatively, in a tomography research setting, where detection precision might be prioritized, reducing false positives could be more important, even if it means potentially missing some weaker seismic events. The AUC is a single value that defines the area under the ROC curve. It quantifies the overall ability of the model to distinguish between positive and negative classes. An AUC of one indicates a perfect model, meaning the model can identify all events correctly without any false positives. Conversely, an AUC of 0.5 represents a random model (Hanley & McNeil, 1982). The F1 score is the harmonic mean of the precision (*i.e.*, the number of correct detections among all detections) and recall (*i.e.*, the number of detections among all possible detections). It serves as a combined measure of the model's sensitivity and specificity. As part of the transfer learning process, the AUC value is selected to optimize the F1 score, thereby fine-tuning the model to achieve an optimal trade-off between the false positive rate and the true positive rate.

In order to measure each model's binary classification performance, we used the Matthews Correlation Coefficient (MCC). It is ambiguous to assign P and S phases as positive and negative classes, and the MCC is insensitive to class assignment (Chicco & Jurman, 2020; Matthews, 1975; Münchmeyer et al., 2022). We analyzed 10 s windows containing exactly one phase arrival to determine if that arrival is a P- or an S-wave. The MCC is calculated as the correlation coefficient of the confusion matrix, and its value ranges from -1 (total disagreement) to 1 (full agreement). Even in cases of class imbalance, the MCC provides an appropriate measure for binary classification performance (Münchmeyer et al., 2022; Powers, 2011). Further, the MCC value was selected to optimize the phase threshold, which is used to calibrate the P- and S-wave pick probability thresholds. The pick probability indicates the likelihood of a specific data point corresponding to a seismic phase arrival (*i.e.*, a P- or an S-wave signal), where a higher probability directly correlates with a heightened level of confidence from the model regarding the presence of an arrival at the identified data point. For the P pick threshold, we multiplied the detection threshold by the square root of the phase threshold. This adjustment enhances the P-wave detection sensitivity and improves identification of these arrivals. For the S pick threshold, we adopted a more conservative approach, dividing the detection threshold by the square root of the phase threshold. This approach was taken to minimize the risk of false positives.

Finally, we evaluated each model's ability to accurately determine the onset time of phase arrivals within a given catalog. Using the same 10 s window used for the MCC assessment, we calculated the pick residuals, which are the differences between the transfer-learning-based pick times and the labeled pick times from the validation subset. The residual distribution is analyzed using both the root-mean-square error (RMSE) and the mean absolute error (MAE). Lower values of RMSE and MAE indicate greater accuracy in predicting the phase arrival onset times. Together, these provide a comprehensive evaluation given their different performance, with RMSE being sensitive to outliers and MAE being less sensitive to them (Willmott & Matsuura, 2005).

5. Results of Transfer Learning

The performance metrics (Sect. 4.3) used to evaluate the four deep learning models (Sect. 4.1) applied to each catalog (Sects. 3.1-3.3) elucidate the effects of transfer learning, and these metrics are summarized in Tables 1-3. Generally, transfer learning has a positive effect on all models, as is evident from the AUC metrics, for example. The most dramatic change was observed for the ML catalog and the BasicPhaseAE model, where the AUC increased from 0.45 to 0.81. That said, even models like GPD that already had a high AUC value (0.87) saw an increase (0.90). These results highlight the benefits of transfer learning. However, it is important to consider how each model defines an event detection. For instance, EQT needs both P- and S-wave labels to declare a detection within the seismogram time series (data from other stations is commonly aggregated during event association, discussed later), while GPD and PhaseNet do not. For scenarios where datasets might lack certain labels, such as in our SL and MF catalogs, this could lead to reduced performance, as reflected in the metric results. It is worth noting that our results are qualitatively comparable to those made by Münchmeyer et al. (2022) for the ETHZ dataset (Woollam et al., 2022), where some P- or S-wave labels were missing.

The RMSE and MAE metrics were reduced for both P and S picks across all catalogs, again indicating improved performance from the fine-tuning and transfer learning. Among all the models, EQT had the lowest of these metrics, indicating it had the highest pick accuracy. However, GPD also displayed significant improvements in RMSE and MAE and closely followed EQT across all catalogs (Tables 1-3). As for the MCC metrics, where higher values indicate better classification performance, every model exhibited a MCC rise following transfer learning. Comparing the three catalogs (Tables 1-3), the P and S picks are notably better classified in the ML catalog for all models, followed by the SL and then the MF catalog. These variations might be due to discrepancies in P- and S-wave labeling consistency across the catalogs. For example, the starting ML catalog was exclusively generated using EQT, perhaps leading to higher pick consistency and, as a result, lower RMSE and MAE values. As a result, variations in performance across the three catalogs reveal that the efficiency of transfer learning also depends on the consistency and quality of the training subset.

Figure 5 shows an example of the pick probabilities for different deep learning models when applied to continuous data. EQT, GPD, and PhaseNet all have improved pick probabilities after transfer learning. The BasicPhaseAE pick probabilities did not increase post-transfer learning, and this could be due to the shorter input windows used by this model, together with its shorter filters and missing residual connections (Münchmeyer et al., 2022).

Table 1. Fine-tuned metric results before (left columns) and after (right columns) transfer learning was applied to the ML catalog. AUC: Area under the Curve; RMSE: root-mean-square error; MAE: mean absolute error; MCC: Matthews Correlation Coefficient.

Model	AUC		P picks RMSE		S picks RMSE		P picks MAE		S picks MAE		MCC	
PhaseNet	0.7	0.8	3.0	2.1	3.0	2.3	2.2	1.4	2.2	1.5	0.3	0.6
BasicPhaseAE	0.4	0.7	3.2	2.3	3.0	2.5	2.5	1.6	2.3	1.7	0.3	0.5
GPD	0.8	0.8	2.2	1.8	2.3	2.1	1.5	1.2	1.6	1.4	0.6	0.8
EQTransformer	0.7	0.7	3.4	1.8	3.0	1.8	2.4	1.1	2.1	1.1	0.6	0.9

Table 2. Fine-tuned metric results before (left columns) and after (right columns) transfer learning was applied to the MF catalog. Columns are the same as in Table 1.

Model	AUC		P picks RMSE		S picks RMSE		P picks MAE		S picks MAE		MCC	
PhaseNet	0.7	0.9	2.8	1.1	2.4	1.2	1.8	0.5	1.6	0.6	0.3	0.7
BasicPhaseAE	0.4	0.8	3.2	1.1	2.8	1.3	2.5	0.6	2.0	0.7	0.3	0.7
GPD	0.8	0.9	1.2	0.6	1.3	0.8	0.6	0.3	0.7	0.4	0.7	0.9
EQTransformer	0.8	0.9	2.7	0.6	2.2	0.5	1.6	0.3	1.2	0.2	0.7	1.0

Table 3. Fine-tuned metric results before (left columns) and after (right columns) transfer learning was applied to the SL catalog. Columns are the same as in Tables 1 and 2.

Model	AUC		P picks RMSE		S picks RMSE		P picks MAE		S picks MAE		MCC	
PhaseNet	0.7	0.8	2.0	1.4	2.4	2.0	1.2	0.8	1.6	1.2	0.4	0.8
BasicPhaseAE	0.4	0.7	2.8	1.7	2.7	2.2	2.0	1.0	1.9	1.4	0.4	0.7
GPD	0.8	0.9	1.4	0.9	2.0	2.0	0.8	0.6	1.3	1.2	0.8	0.9
EQTransformer	0.8	0.8	2.7	0.9	2.3	1.9	1.6	0.5	1.4	1.1	0.7	1.0

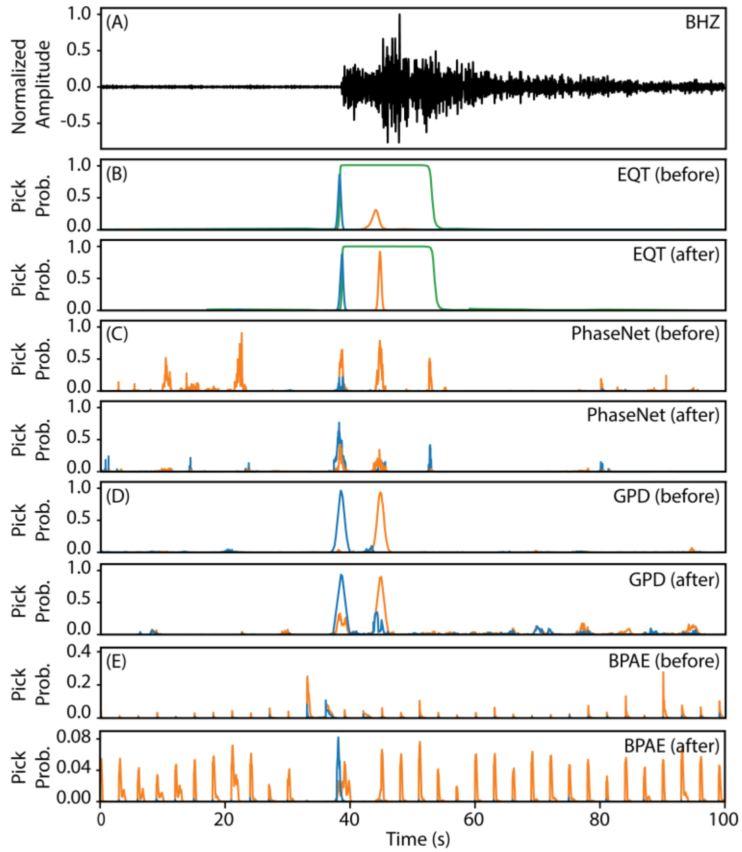

Figure 5. (A) Sample of the continuous Antarctic data recorded by station LEON (Fig. 1), and corresponding pick probabilities for (B) EQT, (C) PhaseNet, (D) GPD, and (E) BasicPhaseAE (BPAE). For each model, the top and bottom panels show the pick probabilities before and after transfer learning, respectively (note that the vertical scales can vary by panel). Blue lines correspond to P-waves, and orange lines correspond to S-waves. For EQT, the green lines show the detection probability.

6. Model Assessment

6.1 Benefits and Limitations of Each Automated Event Detection Approach

Each automated event detection approach has its benefits and limitations, and the choice of which approach to use depends on the objective of the study and the characteristics of the dataset. The STA/LTA method stands out given its minimal pre-processing requirements, straightforward algorithm, and low computational demands, making this technique efficient and readily applicable. Notably, the approach can also identify low magnitude earthquakes if the data has sufficiently high quality (Fig. 6). However, as noted in Section 3.1, STA/LTA can struggle to identify emergent or low SNR arrivals (Schaff & Beroza, 2004; Yoon et al., 2015), which can make this technique more prone to errors, including an increased risks of false positive detections and/or detection failures (Kato et al., 2012). This limitation is partly due to the nature of the STA and LTA window lengths, which are not adjusted during the detection process (Trnkoczy, 2009) and hence restrict the method's ability to adapt to varying seismic signal characteristics. Figure S2 in the Supporting Information shows several examples of missed detections that resulted from the STA/LTA inflexibility. Given its performance, STA/LTA is likely suitable for real-time seismic event detection applications, particularly in situations where an existing, trained model is not available. This method is applicable for systems such as earthquake early warning and volcanic monitoring, which require rapid results. It is important to note that in these scenarios, the immediate availability of results may be prioritized, even if it means accepting a higher likelihood of false positive detections for lower magnitude events (*e.g.*, Kumar et al., 2018; Li et al., 2016; Meier et al., 2020; Tepp, 2018).

The MF approach detects events with high precision, particularly if the events have a high degree of waveform similarity. However, developing a comprehensive set of template events can be time consuming, and the need to compare each of those templates to the continuous data can be computationally demanding (Liu et al., 2020; Meng et al., 2012). Further, since the MF technique is heavily dependent on the pre-defined templates, it is susceptible to missing events that diverge from recognized patterns (Gardonio et al., 2019; Kato & Nakagawa, 2014; Peng & Zhao, 2009; Ross et al., 2018). Several examples of such missed events are shown in Figure S3 in the Supporting Information. Automatic event detection with this method is best-suited to environments where the seismic events are self-similar, such as volcanic-related seismic swarms (*e.g.*, Tan et al., 2023; Whidden et al., 2023; Wimez & Frank, 2022) and repeating stick-slip activity beneath glaciers (*e.g.*, Helmstetter, 2022; Lucas et al., 2023; Ma et al., 2020).

Deep learning event detection techniques can help to address some of the problems faced by the STA/LTA and the MF approaches. Since deep learning models can be trained to recognize intricate seismic patterns, this approach has a greater degree of adaptability across a range of seismic signals and noise. Our analysis also illustrates how deep learning model performance can be further enhanced via transfer learning, where pre-trained models are adapted to recognize the characteristics of unique seismic sources (Chai et al., 2020; Liao et al., 2021). That said, deep learning approaches, with or without transfer learning, have their own set of challenges. ML methods are generally computationally intensive and do not provide rapid results (García et al., 2022; Zhu et al., 2022). Their performance is strongly linked to the quality and volume of their training subsets, and the oft-cited 'black box' nature of ML makes its decision-making processes ambiguous (Gonzalez Garibay et al., 2023). The effectiveness of transfer learning depends on whether the pre-trained model is relevant to the target dataset. If there is a mismatch between the source and target architecture, there is a risk of negative transfer, where the pre-trained model may fail to effectively adapt to the new task (Civilini et al., 2021; Zhou et al., 2021). Careful fine-tuning of the pre-trained model is needed to ensure its applicability to the specific seismic context, and this requires a certain level of understanding regarding the model's architecture. All that said, seismic event catalogs based on ML models typically have a greater magnitude of completeness (*i.e.*, the minimum magnitude above which all events have been detected) compared to those generated by other approaches (Fig. 6; *e.g.*, Ma & Chen, 2022; Reynen & Audet, 2017; Ross et al., 2018), Therefore, if a given study requires robust, extensive seismic constraints, the additional computational resources and complexity of ML algorithms are worth the investment.

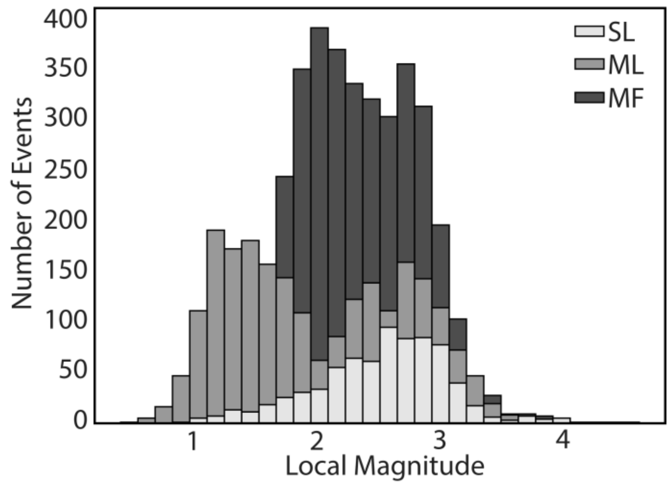

Figure 6. Histogram summarizing the number of events in each catalog after transfer learning was applied, along with their corresponding local magnitude estimates. Light grey bars represent the SL catalog, medium grey bars denote the ML catalog, and dark grey bars correspond to the MF catalog.

6.2 Preferred ML model for East Antarctica

The metrics discussed in Sections 4.3 and 5 provide important information regarding the most applicable model for a given seismic study. For our East Antarctic investigation, we prioritized thorough seismic event detection. While it is important to identify events accurately and precisely, the limited seismic station coverage in our study region (Fig. 1) emphasizes the need to develop an event catalog that is as complete as possible. As such, our ideal model is one that strikes a balance between sensitivity and accuracy, and our extensive analyses indicate that the fine-tuned GPD model is an optimal choice. While EQT displays somewhat better pick accuracy, as indicated by its RMSE and MAE values, its ability to distinguish between positive and negative classes (AUC score) lags behind GPD (Tables 1-3). The trained GPD model's high AUC score emphasizes that this model robustly distinguishes true events from noise. That is, events with low SNR, potentially overlooked by other models and methods, are identified by

GPD. Furthermore, the inherent variability of seismic data demands a model that performs consistently, and the GPD model displays consistent performance across all three examined catalogs, both before and after transfer learning is applied (Tables 1-3). This indicates that the GPD model is highly adaptable, regardless of the data's origin.

7. Application

7.1 GPD Results for Each Catalog

We applied the fine-tuned (transfer-learned) GPD model to the full suite of East Antarctic data (2012-2015; Fig. 1), running three versions of the GPD detection algorithm concurrently, corresponding to our SL, MF, and ML catalogs. As noted in Section 5, each model generates pick probabilities for the designated P- and S-wave arrivals (Fig. 5). Picks with probabilities below a specified threshold (Table 4) are discarded. These thresholds are essential for reducing the number of spurious picks, thereby enhancing the accuracy and reliability of the detected seismic events, and the thresholds ultimately control the number of event identifications. Table 4 summarizes the corresponding pick probability thresholds used to determine qualifying P- and S-waves. These thresholds have led to the identification of new seismic events post-transfer learning. Specifically, after transfer learning, the number of new events in the SL, ML, and MF catalogs is 618, 372, and 201, respectively.

Table 4. P- and S-wave pick probability thresholds for the three transfer-learned catalogs. A P- or S-wave pick is declared if the probability exceeds the specified threshold.

Catalog	P Threshold	S Threshold
ML	0.68	0.81
MF	0.42	0.51
SL	0.51	0.60

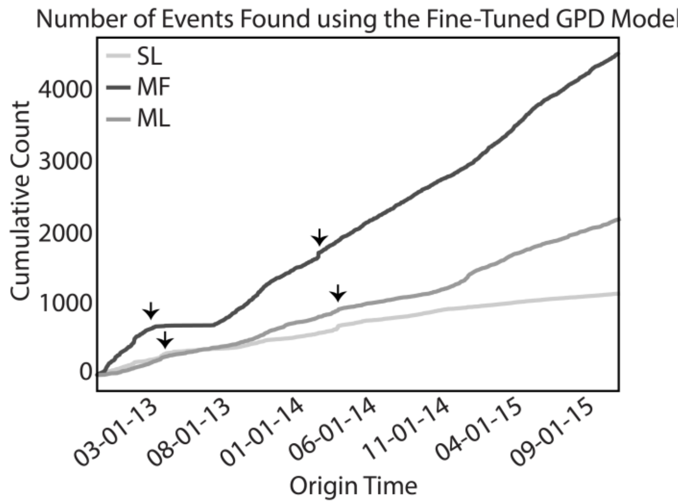

Figure 7. Cumulative number of events included in each catalog after transfer learning was applied. The light grey line corresponds to the SL catalog, the dark grey line corresponds to the MF catalog, and the medium grey line corresponds to the ML catalog. Arrows denote time periods where an increased number of events are observed.

All three catalogs display an increase in the number of events around May 2013 and May 2014 (Fig. 7). These time periods correlate with seasonal changes in Antarctica as the austral winter sets in. Tensile stresses in the ice sheet can be influenced by temperature, and this can impact the formation of crevasses (Harper et al., 1998; Holdsworth, 1969). Specifically, when temperatures drop, the surface layers of the ice sheet can become substantially colder than the underlying firn, and this temperature gradient subjects the colder, more brittle surface layers to an increase in tensile stress. Consequently, new crevasses may form and propagate along the ice sheet surface (Nath & Vaughan, 2003), thereby leading to an increased number of icequakes. This may explain the increase in detected events at these particular time intervals (Fig. 7). Local magnitudes (M_L) were also computed for the SL, ML, and MF catalog events (Fig. 6), though we note that the magnitudes were determined using amplitude attenuation parameters developed for southern California (Hutton & Boore, 1987). While not specific to our study region, these parameters do not impact our assessment since our goal was to simply determine relative event magnitudes rather than to make any interpretations of absolute magnitude. As shown in Figure 6, all three techniques effectively detect low magnitude ($M_L \leq 3$) seismic events, though the ML technique detects a higher number of signals with magnitudes below two.

7.2 Event Relocations

After the fine-tuned GPD model was applied to the full East Antarctic dataset, as described in Section 7.1, the events from each of the updated catalogs were relocated using the NonLinLoc software package (Lomax et al., 2000). An equal differential-time likelihood function and the Oct-Tree sampling approach were used to compute the maximum likelihood hypocenters, based on the corresponding probability density functions (PDFs; Lomax et al., 2000; Zhou, 1994). We also utilized a modified version of the crustal velocity model (Fig. S4 in Supporting Information) from Pyle et al. (2010), which was developed for a nearby region in East Antarctica. Only earthquakes with at least four P- and S-wave arrival times were relocated. Additionally, to account for any possible bias in the procedure, we performed a second inversion using the average arrival-time residuals at each station (Lomax et al., 2009), thereby leading to better constrained event locations.

For each event relocation, the average horizontal and vertical uncertainties of the confidence ellipsoid, which are estimated by the PDFs, were used to determine the volume of the 68% confidence ellipse. This, in turn, was used to determine the average uncertainty (R_e) of each event location (Lomax et al., 2000). The relocated events in each catalog were then grouped based on their uncertainty thresholds. The best constrained event locations (Group A) had $R_e \leq 5$ km. Groups B, C, and D had progressively higher R_e values (Group B: $5 < R_e \leq 10$ km; Group C: $10 < R_e \leq 20$ km; Group D: $R_e \geq 20$ km), indicating less well-constrained locations. The number of events in each quality group is provided in Table S1 in the Supporting Information. Figure 8 highlights event locations that had $R_e \leq 10$ km (*i.e.*, Groups A and B) within each catalog, and events from all groups are shown in Figure S5 in the Supporting Information.

Many of the detected events in all three catalogs are situated near David Glacier (Fig. 8). Shallow events (< 5 km) in this region are consistent with those identified in previous studies (*e.g.*, Bannister & Kennett, 2002; Danesi et al., 2007, 2022; Zoet et al., 2012; 2013), which have been attributed to stick-slip behavior at the base of the ice sheet. However, all three catalogs also show deeper events (> 10 km) beneath the David Glacier region as well, which could be associated with solid Earth processes. For example, movement and mass redistribution within the East Antarctic ice sheet may induce stress changes in the underlying lithosphere, creating the deep-seated events highlighted in our catalogs (Lund, 2015; Steffen, 2013; Steffen et al., 2020).

All three event catalogs also show notable seismicity beneath Victoria Land, in the northeastern portion of the study region (Fig. 8). The prevalence of event detections in this area may reflect some degree of spatial bias given the locations of the stations available for this study (Fig. 1). The TAMNNET stations, in particular, provide somewhat better coverage in this region; therefore, nearby events may more likely meet the enforced minimum number of P- and S-wave arrivals needed for relocation. That said, the Victoria Land event cluster (Fig. 8) is concentrated near several other glaciers that move across the Transantarctic Mountains and towards the Ross Sea, including the Campbell, Priestley, and Aviator Glaciers. The best located events in this cluster are relatively shallow and therefore may reflect ice-bed processes, similar to those suggested for David Glacier further to the south. Deeper events are also seen beneath this region, down to about 25-50 km, which are more likely associated with tectonic processes, such as faulting (*e.g.*, Pisarska-Jamrozy et al., 2018), or with crustal deformation driven by cryospheric fluctuations (*e.g.*, Stewart et al., 2000). Further investigations would be needed to evaluate the sources of the seismic events beneath David Glacier and Victoria Land, but the automatically identified events from our analyses provide some insight into the complex relationship between the solid Earth structure and the Antarctic ice sheet.

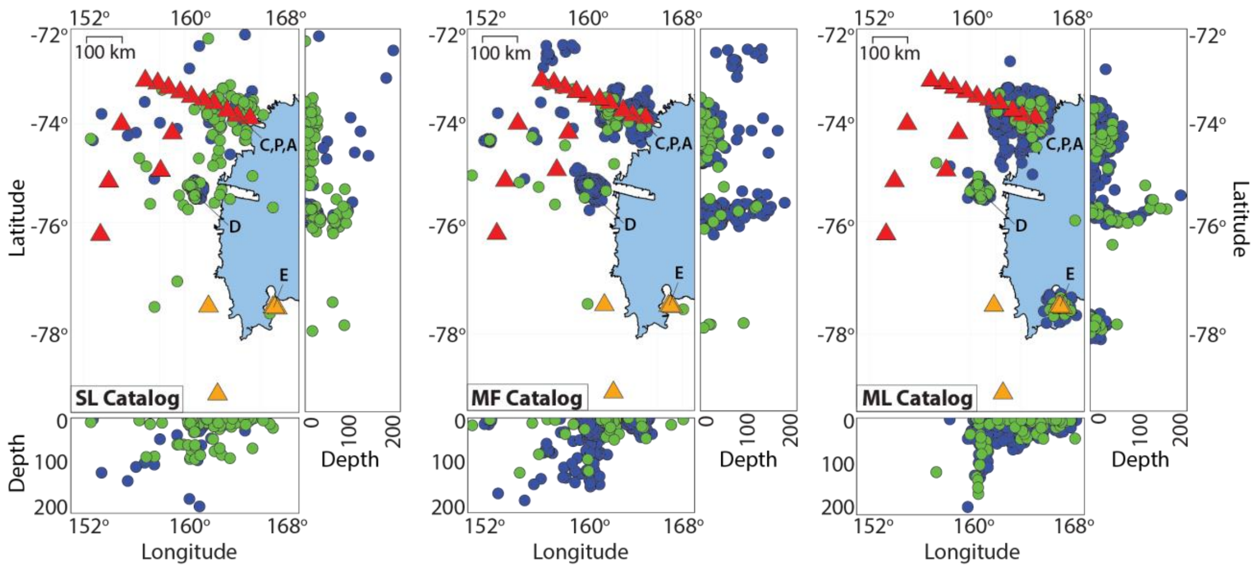

Figure 8. Seismic event relocations from NonLinLoc for quality Group A and B events. From left to right: SL catalog, MF catalog, and ML catalog. Blue circles denote events that were detected by the corresponding original technique (*i.e.*, STA/LTA, template matching, EQT machine learning). Green circles denote new events detected by transfer learning. Red triangles indicate TAMNNET stations, and orange triangles denote other stations. Abbreviations denote key locations including David Glacier (D), Campbell, Priestly, and Aviator Glaciers (C,P,A), and Mount Erebus (E).

It is also worth noting the cluster of seismic events near Mount Erebus, which was uniquely identified by the ML catalog (Fig. 8). Some prior studies that have also recognized seismicity in this region attribute the events to small magnitude icequake sources near the volcano's summit (Chaput et al., 2015; Li et al., 2021; Podolskiy & Walter, 2016). Other investigations have attributed the Mount Erebus seismicity to volcanic activity within its shallow magmatic system (*e.g.*, Aster et al., 2008; Hansen & Schmandt, 2015; Kaminuma, 1987; Rowe et al., 1998, 2000). The absence of the Mount Erebus event cluster in the SL and MF catalogs underscores the effectiveness of deep learning techniques in seismic detection, particularly in elucidating events with a range of sources.

8. Conclusions

Our study has evaluated the benefits and limitations of different automated seismic event detection methods, and our results emphasize that the most appropriate approach depends on the specific attributes of the examined data as well as the objectives of a given study. The STA/LTA method is well-suited for real-time event detection applications that require rapid results, even if there is a higher likelihood for false detections. The MF technique works well for environments that generate seismic events with a high degree of waveform similarity. Deep learning models offer the most adaptability if dealing with a range of seismic sources and noise, and their performance can be enhanced with transfer learning, which provides an effective approach to adapt pre-trained models for unique datasets.

For our East Antarctic investigation, the fine-tuned GPD model, characterized by its high AUC score, reliable picking accuracy, and consistent performance across the examined catalogs, emerged as the most robust, providing new insights into seismic sources in the region. Event relocations based on the fine-tuned catalogs offer new insights into potential seismic sources, including both shallow cryospheric and deeper tectonic processes. Arguably, the most comprehensive seismic event catalog would be one created by integrating the results from each of the applied detection methods; however, the separate results highlight the performance of each approach. Our findings have expanded seismic event detections in East Antarctica, including the identification of previously unrecognized seismicity, and these results underscore the potential for automated event detection approaches to enhance our understanding of seismic activity even in areas with limited station coverage.

Acknowledgments

Funding for this research was provided by the National Science Foundation (grants OPP-1914698 (SEH), EAR-1745135 (JIW), and EAR-1925965 (ZP)), the Ministerio de Ciencia e Innovación / Agencia Estatal de Investigació (MCIN/AEI/10.13039/501100011033), and the European Union Next Generation EU “Plan de Recuperación, Transformación y Resiliencia” (PRTR; grant Model_SHaKER, PID2021-124155NB-C31 (JLSR)).

Open Research

The facilities of SAGE Data Services, particularly the SAGE Data Management Center, were used to access the waveforms and related metadata used in this project (http://ds.iris.edu/mda/). Specifically, we employed datasets from ASL/USGS (1993) and Hansen (2012). SAGE Data Services are funded through the Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048. Some figures were generated with the Generic Mapping Tools software (Wessel et al., 2013). Our ML catalog was developed with the easyQuake software package (Walter et al., 2021; Walter, 2022), and transfer learning was implemented with Seisbench (Ho, 2024; Woollam et al., 2022).

References

Albuquerque Seismological Laboratory (ASL)/USGS (1993). Global Telemetered Seismograph Network [Dataset]. International Federation of Digital Seismograph Networks, doi: 10.7914/SN/GT.

Allen, R. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, *68*(5), 1521–1532, doi: 10.1785/BSSA0680051521

Anandakrishnan, S., Voigt, D. E., Burkett, P. G., Long, B., & Henry, R. (2000). Deployment of a broadband seismic network in West Antarctica. *Geophysical Research Letters*, *27*(14), 2053-2056, doi: 10.1029/1999GL011189

Anthony, R. E., Aster, R. C., Wiens, D., Nyblade, A., Anandakrishnan, S., Huerta, A., Winberry, J. P., Wilson, T. & Rowe, C. (2015). The seismic noise environment of Antarctica. *Seismological Research Letters*, *86*(1), 89-100, doi: 10.1785/0220140109

Aster, R., Zandomeneghi, D., Mah, S., McNamara, S., Henderson, D. B., Knox, H., & Jones, K. (2008). Moment tensor inversion of very long period seismic signals from Strombolian eruptions of Erebus Volcano. *Journal of Volcanology and Geothermal Research*, 177(3), 635-647, doi: 10.1016/j.jvolgeores.2008.08.013

Baer, M., & Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, *77*(4), 1437–1445, doi: 10.1785/BSSA0770041437

Bannister, S., & Kennett, B. L. N. (2002). Seismic Activity in the Transantarctic Mountains - Results from a Broadband Array Deployment. *Terra Antarctica*, *9*, 41-46.

Bentley, C. R., & Kohnen, H. (1976). Seismic refraction measurements of internal friction in Antarctic ice. *Journal of Geophysical Research: Solid Earth*, *81*(8), 1519-1526, doi: 10.1029/JB081i008p01519

Boulder Real Time Technologies (BRTT; 2011). Evolution of the Commercial ANTELOPE Software.

Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., Beroza, G. C., Thurber, C., & EGS Collaboration Team (2020). Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical Research Letters*, *47*(16), doi: 10.1029/2020GL088651

Chamberlain, C. J., Hopp, C. J., Boese, C. M., Warren-Smith, E., Chambers, D., Chu, S. X., Michailos, K., & Townend, J. (2018). EQcorrscan: Repeating and near-repeating earthquake detection and analysis in Python. *Seismological Research Letters*, *89*(1), 173-181, doi: 10.1785/0220170151

Chaput, J., Campillo, M., Aster, R. C., Roux, P., Kyle, P. R., Knox, H., & Czoski, P. (2015). Multiple scattering from icequakes at Erebus volcano, Antarctica: Implications for imaging at glaciated volcanoes. *Journal of Geophysical Research: Solid Earth*, 120(2), 1129-1141, doi: 10.1002/2014JB011278

Chen, X., Shearer, P. M., Walter, F., & Fricker, H. A. (2011). Seventeen Antarctic seismic events detected by global surface waves and a possible link to calving events from satellite images. *Journal of Geophysical Research: Solid Earth*, *116*(B6), doi: 10.1029/2011JB008262

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

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), doi: 10.1186/s12864-019-6413-7

Cianetti, S., Bruni, R., Gaviano, S., Keir, D., Piccinini, D., Saccorotti, G., & Giunchi, C. (2021). Comparison of deep learning techniques for the investigation of a seismic sequence: An application to the 2019, Mw 4.5 Mugello (Italy) earthquake. *Journal of Geophysical Research: Solid Earth*, *126*(12), doi: 10.1029/2021JB023405

Civilini, F., Weber, R. C., Jiang, Z., Phillips, D., & Pan, W. D. (2021). Detecting moonquakes using convolutional neural networks, a non-local training set, and transfer learning. *Geophysical Journal International*, *225*(3), 2120-2134, doi: 10.1093/gji/ggab083

Coleman, C., Narayanan, D., Kang, D., Zhao, T., Zhang, J., Nardi, L., et al. (2017). Dawnbench: An end-to-end deep learning benchmark and competition. *Training*, *100*(101), 102.

Danesi, S., Bannister, S., & Morelli, A. (2007). Repeating earthquakes from rupture of an asperity under an Antarctic outlet glacier. *Earth and Planetary Science Letters*, *253*, 151-158, doi: 10.1016/j.epsl.2006.10.023

Danesi, S., Salimbeni, S., Borghi, A., Urbini, S., & Frezzotti, M. (2022). Cryo-seismicity triggered by ice mass discharge through the Antarctic subglacial hydrographic network. *The Cryosphere Discussions*, doi: 10.5194/egusphere-2022-29.

Earle, P. S., & Shearer, P. M. (1994). Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the Seismological Society of America*, 84(2), 366-376, doi: 10.1785/BSSA0840020366

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874, doi: 10.1016/j.patrec.2005.10.010

García, J. E., Fernández-Prieto, L. M., Villaseñor, A., Sanz, V., Ammirati, J. B., Díaz Suárez, E. A., & García, C. (2022). Performance of deep learning pickers in routine network processing applications. *Seismological Society of America*, *93*(5), 2529-2542, doi: 10.1785/0220210323

Gardonio, B., Campillo, M., Marsan, D., Lecointre, A., Bouchon, M., & Letort, J. (2019). Seismic activity preceding the 2011 M w 9.0 Tohoku earthquake, Japan, analyzed with multidimensional template matching. *Journal of Geophysical Research: Solid Earth*, 124(7), 6815-6831, doi: 10.1029/2018JB016751

Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: continual prediction with LSTM, *Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, Edinburgh, UK, 850-855, doi: 10.1049/cp:19991218.

Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophysical Journal International*, *165*(1), 149–166, doi: 10.1111/j.1365-246X.2006.02865.x

González Garibay, M., Primc, K., & Slabe-Erker, R. (2023). Insights into advanced models for energy poverty forecasting. *Nature Energy*, *8*, 903-905, doi: 10.1038/s41560-023-01311-x

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, ISBN: 9780262035613.

Guerin, G., Mordret, A., Rivet, D., Lipovsky, B. P., & Minchew, B. M. (2021). Frictional origin of slip events of the Whillans Ice Stream, Antarctica. *Geophysical Research Letters*, *48*(11), doi: 10.1029/2021GL092950.

Hampel, F. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*, 383-393, doi: 10.1080/01621459.1974.10482962

Han, J., Kim, S., Sheen, D. H., Lee, D., Lee, S. J., Yoo, S. H., & Park, D. (2023). Seismic event and phase detection using deep learning for the 2016 Gyeongju earthquake sequence. *Geosciences Journal*, *27*(3), 285-295, doi: 10.1007/s12303-023-0004-y

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36, doi: 10.1148/radiology.143.1.7063747

Hansen, S. (2012). Transantarctic Mountains Northern Network [Dataset]. International Federation of Digital Seismograph Networks, doi: 10.7914/SN/ZJ_2012.

Hansen, S. E., Reusch, A., Parker, T., Bloomquist, D., Carpenter, P., Graw, J. H., & Brenn, G. R. (2015). The Transantarctic Mountains Northern Network (TAMNNET): Deployment and Performance of a Seismic Array in Antarctica. *Seismological Research Letters*, *86*, 1636-1644, doi: 10.1785/0220150117

Hansen, S. M., & Schmandt, B. (2015). Automated detection and location of microseismicity at Mount St. Helens with a large-N geophone array. *Geophysical Research Letters*, *42*(18), 7390-7397, doi: 10.1002/2015GL064848

Harper, J. T., Humphrey, N.F., & Pfeffer, W. T. (1998). Crevasse patterns and the strain-rate tensor: A high resolution comparison, *Journal of Glaciology*, *44*, 68-77, doi: 10.3189/S0022143000002367

Heeszel, D. S., Wiens, D. A., Nyblade, A. A., Hansen, S. E., Kanao, M., An, M., & Zhao, Y. (2013). Rayleigh wave constraints on the structure and tectonic history of the Gamburtsev Subglacial Mountains, East Antarctica. *Journal of Geophysical Research: Solid Earth*, *118*(5), 2138-2153, doi: https://doi.org/10.1002/jgrb.50171

Helmstetter, A. (2022). Repeating Low Frequency Icequakes in the Mont-Blanc Massif Triggered by Snowfalls. *Journal of Geophysical Research: Earth Surface*, *127*(12), doi: 10.1029/2022JF006837

Herrmann, M., & Marzocchi, W. (2021). Inconsistencies and lurking pitfalls in the magnitude–frequency distribution of high-resolution earthquake catalogs. *Seismological Research Letters*, 92(2A), 909-922, doi: 10.1785/0220200337

Ho, L. M. (2024). Transfer_Learning_and_Seisbench [Software], GitHub, https://github.com/longmho/Transfer_Learning_and_Seisbench.

Holdsworth, G. (1969). Primary transverse crevasses, *Journal of Glaciology*, *8*, 107-129.

Hudson, T. S., Kufner, S. K., Brisbourne, A. M., Kendall, J. M., Smith, A. M., Alley, R. B., et al. (2023). Highly variable friction and slip observed at Antarctic ice stream bed. *Nature Geoscience*, *16*, 612-618, doi: 10.1038/s41561-023-01204-4

Hutton, L. K., & Boore, D. M. (1987). The ML scale in south California, *Bulletin of the Seismological Society of America*, *77*(6), 2074-2094, doi: 10.1785/BSSA0740051827

Jiang, C., Fang, L., Fan, L., & Li, B. (2021). Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. *Earthquake Science, 34*(5), 425-435, doi: 10.29382/eqs-2021-0038

Kaminuma, K. (1987). Seismic activity of Erebus volcano, Antarctica. *Pure and Applied Geophysics*, *125*, 993-1008, doi: 10.1007/BF00879364

Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., & Hirata, N. (2012). Propagation of slow slip leading up to the 2011 M w 9.0 Tohoku-Oki earthquake. *Science*, 335(6069), 705-708, doi: 10.1126/science.1215141

Kato, A., & Nakagawa, S. (2014). Multiple slow-slip events during a foreshock sequence of the 2014 Iquique, Chile Mw 8.1 earthquake. *Geophysical Research Letters*, 41(15), 5420-5427, doi: 10.1002/2014GL061138

Kennett, B. L. N., & Engdahl, E. R. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105(2), 429-465, doi: 10.1111/j.1365-246X.1991.tb06724.x

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

Kumar, S., Vig, R., & Kapur, P. (2018). Development of Earthquake Event Detection Technique Based on STA/LTA Algorithm for Seismic Alert System. *Journal of the Geological Society of India*, 92, 679-686, doi: 10.1007/s12594-018-1087-3

Lapins, S., Goitom, B., Kendall, J. M., Werner, M. J., Cashman, K. V., & Hammond, J. O. (2021). A little data goes a long way: Automating seismic phase arrival picking at Nabro volcano with transfer learning. *Journal of Geophysical Research: Solid Earth, 126*(7), doi: 10.1029/2021JB021910.

Li, X., Shang, X., Wang, Z., Dong, L., & Weng, L. (2016). Identifying P-phase arrivals with noise: An improved Kurtosis method based on DWT and STA/LTA. *Journal of Applied Geophysics*, *133*, 50-61, doi: 10.1016/j.jappgeo.2016.07.022

Li, Z., Peng, Z., Hollis, D., Zhu, L., & McClellan, J. (2018). High-resolution seismic event detection using local similarity for Large-N arrays. *Scientific Reports*, *8*(1), doi: 10.1038/s41598-018-19728-w.

Li, C., Peng, Z., Chaput, J. A., Walter, J. I., & Aster, R. C. (2021). Remote triggering of Icequakes at Mt. Erebus, Antarctica by large teleseismic earthquakes. *Seismological Research Letters*, *92*(5), 2866-2875, doi: 10.1785/0220210027

Liao, W., Chen, X., Lu, X., Huang, Y., & Tian, Y. (2021). Deep transfer learning and time-frequency characteristics-based identification method for structural seismic response. *Frontiers in Built Environment*, *7*, doi: 10.3389/fbuil.2021.627058.

Liu, M., Li, H., Zhang, M., & Wang, T. (2020). Graphics processing unit-based match and locate (GPU-M&L): An improved match and locate method and its application. *Seismological Research Letters*, *91*(2A), 1019-1029, doi: 10.1785/0220190241

Lomax, A., Virieux, J., Volant, P., & Berge-Thierry, C. (2000). Probabilistic earthquake location in 3D and layered models. In *Advances in Seismic Event Location*, eds. C.H. Thurber & N. Rabinowitz, 101-134, doi: 10.1007/978094-015-9536-0_5.

Lomax, A., Michelini, A., Curtis, A., & Meyers, R. A. (2009). Earthquake location, direct, global-search methods. *Encyclopedia of Complexity and Systems Science*, *5*, 2449-2473, doi:10.1007/978-0-387-30440-3

Long, M., Wang, J., Ding, G., Sun, J., & Yu, P. S. (2013). Transfer Feature Learning with Joint Distribution Adaptation. *Proceedings of the IEEE International Conference on Computer Vision*, 2200–2207.

Long, M., Cao, Z., Wang, J., & Jordan, M. I. (2015). Learning transferable features with deep adaptation networks. *Proceedings of the 32nd International conference on machine learning*, 97-105.

Lough, A. C., Wiens, D. A., Barcheck, G.C., Anandakrishnan, S., Aster, R. C., Blankenship, D. D., et al. (2013). Seismic detection of an active subglacial magmatic complex in Marie Byrd Land, Antarctica. *Nature Geoscience*, *6*(12), 1031-1035, doi: 10.1038/ngeo1992

Lough, A. C., Wiens, D. A., & Nyblade, A. (2018). Reactivation of ancient Antarctic rift zones by intraplate seismicity. *Nature Geoscience*, *11*(7), 515-519, doi: 10.1038/s41561-018-0140-6

Lucas, E. M., Nyblade, A. A., Aster, R. C., Wiens, D. A., Wilson, T. J., Winberry, J. P., & Huerta, A. D. (2023). Tidally Modulated Glacial Seismicity at the Foundation Ice Stream, West Antarctica. *Journal of Geophysical Research: Earth Surface*, *128*(7), doi: 10.1029/2023JF007172.

Lund, B. (2015). Palaeoseismology of Glaciated Terrain. Encyclopedia of Earthquake Engineering, Springer, Berlin, doi: 10.1007/978-3-642-36197-5_25-1.

Ma, H., Chu, R., Sheng, M., & Yu, Z. (2020). Template matching for simple waveforms with low signal-to-noise ratio and its application to icequake detection. *Earthquake Science*, *33*, 256-263, doi: 10.29382/eqs-2020-0256-01

Ma, X., & Chen, T. (2022). Small Seismic Events in Oklahoma Detected and Located by Machine Learning–Based Models. *Bulletin of the Seismological Society of America*, *112*(6), 2859-2869.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *405*(2), 442-451, doi: 10.1016/0005-2795(75)90109-9

Meier, M. A., Kodera, Y., Böse, M., Chung, A., Hoshiba, M., Cochran, E., et al. (2020). How often can earthquake early warning systems alert sites with high-intensity ground motion?. *Journal of Geophysical Research: Solid Earth*, *125*(2), doi: 10.1029/2019JB017718

Meng, X., Yu, X., Peng, Z., & Hong, B. (2012). Detecting earthquakes around Salton Sea following the 2010 Mw7. 2 El Mayor-Cucapah earthquake using GPU parallel computing. *Procedia Computer Science*, *9*, 937-946, doi: 10.1016/j.procs.2012.04.100

Meng, X., Peng, Z., & Hardebeck, J. (2013). Seismicity around Parkfield correlates with static shear stress changes following the 2003 Mw6.5 San Simeon earthquake, *Journal of Geophysical Research*, 118(7), 3576-3591, doi:10.1002/jgrb.50271

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*, 179,464-179,476, doi: 10.1109/ACCESS.2019.2947848

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), doi: 10.1038/s41467-020-17591-w

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), doi: 10.1029/2021JB023499

Myers, S. C., Johannesson, G., & Hanley, W. (2007). A Bayesian hierarchical method for multiple-event seismic location. *Geophysical Journal International*, *171*(3), 1049–1063, doi: 0.1111/j.1365-246X.2007.03555.x

Nath, P. C., & Vaughan, D. G. (2003). Subsurface crevasse formation in glaciers and ice sheets. *Journal of Geophysical Research: Solid Earth*, *108*(B1), doi: 10.1029/2001JB000453

Nettles, M., & Ekström, G. (2010). Glacial earthquakes in Greenland and Antarctica. *Annual Review of Earth and Planetary Sciences*, *38*, 467-491, doi: 10.1146/annurev-earth-040809-152414

Neves, M., Chuang, L.Y., Li. W., Peng, Z., Figueiredo, P. M., & Ni. S. (2024). A dense microearthquake catalog reveals the complex rupture of the extremely shallow M5.1 Sparta, North Carolina Earthquake. *Communications Earth & Environment.*

Pan, S. J. & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345-1359, doi: 10.1109/TKDE.2009.191

Pan, S. J., Tsang, I. W., Kwok, J. T., & Yang, Q. (2011). Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks*, *22*(2), 199–210, doi: 0.1109/TNN.2010.2091281

Peng, Z., & Zhao, P. (2009). Migration of early aftershocks following the 2004 Parkfield earthquake. *Nature Geoscience*, *2*(12), 877-881, doi: 10.1038/ngeo697

Peng, Z., Walter, J. I., Aster, R. C., Nyblade, A., Wiens, D. A., & Anandakrishnan, S. (2014). Antarctic icequakes triggered by the 2010 Maule earthquake in Chile. *Nature Geoscience*, *7*(9), 677-681, doi: 10.1038/ngeo2212

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), doi: 10.1126/sciadv.170057

Pisarska-Jamroży, M., Belzyt, S., Börner, A., Hoffmann, G., Hüneke, H., Kenzler, M., et al. (2018). Evidence from seismites for glacio-isostatically induced crustal faulting in front of an advancing land-ice mass (Rügen Island, SW Baltic Sea). *Tectonophysics*, *745*, 338-348, doi: 10.1016/j.tecto.2018.08.004

Podolskiy, E. A., & Walter, F. (2016). Cryoseismology. *Reviews of Geophysics*, *54*(4), 708-758.

Powers, D.M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, *2*(1), 37-63, doi: 10.48550/arXiv.2010.16061

Pyle, M. L., Wiens, D. A., Nyblade, A. A., & Anandakrishnan, S. (2010). Crustal structure of the Transantarctic Mountains near the Ross Sea from ambient seismic noise tomography. *Journal of Geophysical Research: Solid Earth*, *115*(B11), doi: 10.1029/2009JB007081

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, doi: 10.1093/gji/ggx238

Riel, B., Minchew, B., & Bischoff, T. (2021). Data-Driven Inference of the Mechanics of Slip Along Glacier Beds Using Physics-Informed Neural Networks: Case Study on Rutford Ice Stream, Antarctica. *Journal of Advances in Modeling Earth Systems*, *13*(11), doi: 10.1029/2021MS002621

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention*, *9351*, 234-241, doi: 10.1007/978-3-319-24574-4_28

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, doi: 10.1029/2017JB015251

Rowe, C. A., Aster, R. C., Kyle, P. R., Schlue, J. W., & Dibble, R. R. (1998). Broadband recording of Strombolian explosions and associated very-long-period seismic signals on Mount Erebus Volcano, Ross Island, Antarctica. *Geophysical Research Letters*, *25*(13), 2297-2300, doi: 10.1029/98GL01622

Rowe, C. A., Aster, R. C., Kyle, P. R., Dibble, R. R., & Schlue, J. W. (2000). Seismic and acoustic observations at Mount Erebus volcano, Ross Island, Antarctica, 1994–1998. *Journal of Volcanology and Geothermal Research*, *101*(1-2), 105-128, doi: 10.1016/S0377-0273(00)00170-0

Schaff, D. P., & Beroza, G. C. (2004). Coseismic and postseismic velocity changes measured by repeating earthquakes. *Journal of Geophysical Research: Solid Earth*, *109*(B10), doi: 10.1029/2004JB003011

Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, *446*(7133), 305-307, doi: 10.1038/nature05666

Si, X., Wu, X., Li, Z., Wang, S., & Zhu, J. (2024). An all-in-one seismic phase picking, location, and association network for multi-task multi-station earthquake monitoring. *Communications Earth & Environment*, 5(1), 22, doi: 10.1038/s43247-023-01188-4

Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2015). Earthquakes induced by hydraulic fracturing in Poland Township, Ohio. *Bulletin of the Seismological Society of America*, *105*(1), 189-197, doi: 10.1785/0120140168

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. *US Naval Research Laboratory Technical Report*, 5510-026, doi: 10.48550/arXiv.1803.09820

Steffen, R. (2013). The influence of glacial isostatic adjustment on intraplate seismicity in northeastern Canada. *Doctoral Thesis*, University of Calgary, Canada, doi: 10.11575/PRISM/28209

Steffen, R., Steffen, H., Weiss, R., Lecavalier, B. S., Milne, G. A., Woodroffe, S. A., & Bennike, O. (2020). Early Holocene Greenland-ice mass loss likely triggered earthquakes and tsunami. *Earth and Planetary Science Letters*, *546*, doi: 10.1016/j.epsl.2020.116443.

Stewart, I. S., Sauber, J., & Rose, J. (2000). Glacio-seismotectonics: ice sheets, crustal deformation and seismicity. *Quaternary Science Reviews*, *19*(14-15), 1367-1389, doi: 10.1016/S0277-3791(00)00094-9

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27* (pp. 270-279). Springer International Publishing.

Tan, D., Fee, D., Hotovec-Ellis, A. J., Pesicek, J.D., Haney, M.M., Power, J. A., & Girona, T. (2023). Volcanic earthquake catalog enhancement using integrated detection, matched-filtering, and relocation tools. *Frontiers in Earth Science*, *11*, doi: 10.3389/feart.2023.1158442

Tepp, G. (2018). A Repeating Event Sequence Alarm for Monitoring Volcanoes. *Seismological Research Letters*, *89*(5), 1863-1876, doi: 10.1785/0220170263

Trnkoczy, A. (2009). Understanding and parameter setting of STA/LTA trigger algorithm. In *New manual of seismological observatory practice,* Deutsches GeoForschungsZentrum GFZ, 1-20.

Vaezi, Y., & Van der Baan, M. (2015). Comparison of the STA/LTA and power spectral density methods for microseismic event detection. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, *203*(3), 1896-1908, doi: 10.1093/gji/ggv419

Waldhauser, F., & Schaff, D. (2007). Regional and teleseismic double-difference earthquake relocation using waveform cross-correlation and global bulletin data. *Journal of Geophysical Research: Solid Earth*, *112*(B12), doi: 10.1029/2007JB004938

Walter, J. I., E. E. Brodsky, S. Tulaczyk, S. Y. Schwartz, & R. Pettersson (2011). Transient slip events from near-field seismic and geodetic data on a glacier fault, Whillans Ice Plain, West Antarctica, *Journal of Geophysical Research: Solid Earth*, *116*, doi:10.1029/2010JF001754

Walter, J. I., I. Svelitsky, J. Fineberg, S. Tulaczyk, E. E. Brodsky, C. G. Barcheck, & S. P. Carter (2015). Rupture speed dependence on initial stress profiles: Insights from glacier and laboratory stick-slip, *Earth and Planetary Science Letters*, *411*, 112-120, doi: 10.1016/j.epsl.214.11.025

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., & Woelfel, I. (2021). easyQuake: Putting machine learning to work for your regional seismic network or local earthquake study, *Seismological Research Letters*, *92*(1): 555–563, doi: 10.1785/0220200226

Walter, J. I. (2022). easyQuake [Software], GitHub, https://github.com/jakewalter/easyQuake.

Wessel, P., Smith, W. H., Scharroo, R., Luis, J., & Wobbe, F. (2013). Generic mapping tools: improved version released. *Eos, Transactions American Geophysical Union*, *94*(45), 409-410, doi: 10.1002/2013EO450001

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79-82, doi: 10.3354/cr030079

Wimez, M., & Frank, W. B. (2022). A recursive matched-filter to systematically explore volcanic long-period earthquake swarms. *Geophysical Journal International*, *231*(2), 912-920, doi: 10.1093/gji/ggac221

Winberry, J. P., Anandakrishnan, S., Alley, R. B., Wiens, D. A., & Pratt, M. J. (2014). Tidal pacing, skipped slips and the slowdown of Whillans Ice Stream, Antarctica. *Journal of Glaciology*, *60*(222), 795-807, doi: 10.3189/2014JoG14J038

Winberry, J. P., Huerta, A. D., Anandakrishnan, S., Aster, R.C., Nyblade, A.A., & Wiens, D. A. (2020). Glacial earthquakes and precursory seismicity associated with Thwaites Glacier calving. *Geophysical Research Letters*, *47*(3), doi: 10.1029/2019GL086178

Whidden, K. M., Petersen, G. M., Mesimeri, M., & Pankow, K. L. (2023). Analysis of the 2021 Milford, Utah earthquake swarm: Enhanced earthquake catalog and migration patterns. *Frontiers in Earth Science*, *11*, doi: 10.3389/feart.2023.1057982

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, doi: 10.1785/0220180312

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., & Soto, H. (2022). Seisbench-A toolbox for machine learning in seismology. *Seismological Research Letters*, *81*(3), 1695–1709, doi: 10.1785/0220210324

Wu, Y., Lin, Y., Zhou, Z., Bolton, D. C., Liu, J., & Johnson, P. (2018). DeepDetect: A Cascaded Region-Based Densely Connected Network for Seismic Event Detection, *IEEE Transactions on Geoscience and Remote Sensing*, *57*, 62-75, doi: 0.1109/TGRS.2018.2852302

Yao, D., Peng, Z., Kaneko, Y., Fry, B., & Meng, X. (2021). Dynamic triggering of earthquakes in the North Island of New Zealand following the 2016 Mw 7.8 Kaikōura earthquake. *Earth and Planetary Science Letters*, *557*, doi: 10.1016/j.epsl.2020.116723

Yuan, C., Ni, Y., Lin, Y., & Denolle, M. (2023). Better Together: Ensemble Learning for Earthquake Detection and Phase Picking. *IEEE Transactions on Geoscience and Remote Sensing*, doi: 10.1109/TGRS.2023.3320148

Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances*, *1*(11), doi: 10.1126/sciadv.1501057

Zhou, H. (1994). Rapid three-dimensional hypocentral determination using a master station method. *Journal of Geophysical Research*, *99*(B8), doi: 10.1029/94JB00934

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, doi: 10.1785/0220180319

Zhou, R., Yao, X., Hu, G., & Yu, F. (2021). Learning from unlabelled real seismic data: Fault detection based on transfer learning. *Geophysical Prospecting*, *69*(6), 1218-1234.

Zhu, W. & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Research Letters*, *45*(13), 6644-6653, doi: 10.1093/gji/ggy423

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 Mw 7.9 Wenchuan Earthquake. *Physics of the Earth and Planetary Interiors*, *293*, doi: 10.1016/j.pepi.2019.05.004

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), doi: 10.1029/2021JB023283

Zoet, L. K., Anandakrishnan, S., Alley, R. B., Nyblade, A. A., & Wiens, D. A. (2012). Motion of an Antarctic glacier by repeated tidally modulated earthquakes. *Nature Geoscience*, *5*(9), 623-626, doi: 10.1038/ngeo1555

Zoet, L. K., Alley, R. B., Anandakrishnan, S., & Christianson, K. (2013). Accelerated subglacial erosion in response to stick-slip motion. *Geology*, *41*(2), 159-162 doi: 10.1130/G33624.1

Evaluating Automated Seismic Event Detection Approaches: An Application to Victoria Land, East Antarctica

L.M. Ho[1], J.I. Walter[2], S.E. Hansen[1], J.L. Sánchez-Roldán[3], and Z. Peng[4]

[1]The University of Alabama, Tuscaloosa, AL 35487

[2]Oklahoma Geological Survey, University of Oklahoma, Norman, OK 73019

[3]Universidad Complutense de Madrid, Madrid, Spain, 28040

[4]Georgia Institute of Technology, Atlanta, GA 30332

Corresponding author: Long Ho (lmho@crimson.ua.edu)

Key Points:

- Deep learning models, enhanced by transfer learning, adapt well to varied seismic sources.
- Automated detection approaches offer insights into both cryospheric and tectonic events in Antarctica.
- Even in regions with limited station coverage, automated detection approaches can help us develop more complete seismicity catalogs.

Abstract

As seismic data collection continues to grow, advanced automated processing techniques for robust phase identification and event detection are becoming increasingly important. However, the performance, benefits, and limitations of different automated detection approaches have not been fully evaluated. Our study examines how the performance of conventional techniques, including the Short-Term Average/Long-Term Average (STA/LTA) method and cross-correlation approaches, compares to that of various deep learning models. We also evaluate the added benefits that transfer learning may provide to machine learning applications. Each detection approach has been applied to three years of seismic data recorded by stations in East Antarctica. Our results emphasize that the most appropriate detection approach depends on the data attributes and the study objectives. STA/LTA is well-suited for applications that require rapid results even if there is a greater likelihood for false positive detections, and correlation-based techniques work well for identifying events with a high degree of waveform similarity. Deep learning models offer the most adaptability if dealing with a range of seismic sources and noise, and their performance can be enhanced with transfer learning, if the detection parameters are fine-tuned to ensure the accuracy and reliability of the generated catalog. Our results in East Antarctic provide new insight into polar seismicity, highlighting both cryospheric and tectonic events, and demonstrate how automated event detection approaches can be optimized to investigate seismic activity in challenging environments.

Plain Language Summary

Given the large quantity of seismic data recorded for geologic investigations, the manual identification of earthquake arrivals is becoming less feasible, and automated detection approaches are becoming increasingly important. However, the benefits and limitations of different detection techniques have not been fully evaluated. We examine a range of automated detection approaches, applied to data recorded by seismic stations in Antarctica, to assess the performance of each method. Additionally, an approach called transfer learning is examined to determine if it can improve the accuracy and reliability of the automated detections. Our results highlight new seismic events in Antarctica, providing insights into both geologic processes and ice-sheet behavior.

1. Introduction

The accurate creation of earthquake catalogs for seismotectonic interpretation requires robust seismic phase identification, event association, and event detection; however, with the ever-increasing availability of seismic data, manual processing by human analysts is becoming less feasible. As such, automated processing techniques are becoming increasingly important. Some event detection techniques, such as the Short-Term Average/Long-Term Average (STA/LTA) method (Allen, 1978; Earle & Shearer, 1994), use relatively simple algorithms and provide rapid results without the need for extensive data pre-processing. Waveform based cross-correlation approaches, such as the matched filter (MF) technique (Gibbons & Ringdal, 2006; Peng & Zhao, 2009; Shelly et al., 2007), can also be applied to STA/LTA generated earthquake catalogs to identify new, closely located events with similar focal mechanisms to those in the initial catalog. However, STA/LTA may not perform well for low signal-to-noise ratio (SNR) data, and cross-correlation based approaches can sometimes generate spatially biased event catalogs (Herrmann & Marzocchi, 2021; Schaff & Beroza, 2004; Yoon et al., 2015). The shortcomings of these methods can also sometimes result in impulsive transient signals or distant regional/teleseismic signals being erroneously identified as local earthquakes (*e.g.*, Meng et al., 2012). In some cases, these challenges can be overcome using phase association algorithms, which analyze triggers from multiple stations to determine whether any combination displays arrival time sequences that align with characteristic seismic event patterns (Myers et al., 2007).

In recent years, advancements in machine learning techniques, coupled with the democratization of open-source software, have provided more sophisticated methods to automatically detect seismic events. In particular, convolutional neural networks (CNN), which perform a sequence of convolution, resampling, and non-linear transformations on raw waveform data, have shown promising results (Perol et al., 2018; Ross et al., 2018; Wu et al., 2018; Zhou et al., 2019; Zhu et al., 2019) when compared to more traditional techniques (Earle & Shearer, 1994; Gibbons & Ringdal, 2006; Peng & Zhao, 2009; Shelly et al., 2007). CNN pickers are designed to provide the added advantage of identifying body wave phases on three-component seismograms, thereby simplifying earthquake association and relocation. However, machine learning algorithms are complex, computationally demanding, and typically require optimization to avoid false-positive event detections.

To date, only a few studies have evaluated the performance of different automated detection approaches with respect to one another or have attempted to combine detection techniques to achieve the best possible outcome (Münchmeyer et al., 2022; Neves et al., 2024; Si et al., 2024; Woollam et al., 2022; Yuan et al., 2023). Further, most of these previous studies have typically only examined select model pairs based on one or a few training datasets (*e.g.*, Han et al., 2023; Jiang et al., 2021; Perol et al., 2018; Vaezi & Van der Baan, 2015), and they largely focus on small magnitude, tectonic-related seismic events. Here, we compare the benefits and limitations of the STA/LTA technique (Earle & Shearer, 1994), the cross-correlation-based MF approach (Peng & Zhao, 2009), and a suite of deep learning models, including EQTransformer (EQT, Mousavi et al., 2020), PhaseNet (Zhu & Beroza, 2019), BasicPhaseAE (Woollam et al., 2019), and the Generalized Phase Detection (GPD) model (Ross et al., 2018). We also update the deep learning models with additional training data derived from this project, a process known as transfer learning. Despite the potential for transfer learning to enhance model adaptability and efficiency (Chai et al., 2020; Lapins et al., 2021), particularly in data-scarce environments, its adoption in seismic studies has not been as rapid or as extensive as in other domains of deep learning research. This gap presents an opportunity to investigate the full capabilities of transfer learning in automatic event detection. We test the performance of the updated versus original deep-learning models using a range of metrics that evaluate each of their abilities to accurately determine the onset time of phase arrivals, to reliably classify phases as P- or S-waves, and to identify events while minimizing the number of false positives. These techniques are applied to a unique set of waveforms that contain a mixture of tectonic earthquake signals and seismic events generated by glacial movement (*e.g.*, icequakes). Collectively, our evaluation allows us to assess the efficacy of each algorithm when applied to a complex dataset.

2. Data and Methods

Broadband seismic deployments across the Antarctic continent have dramatically increased over the past several decades (*e.g.*, Anandakrishnan et al., 2000; Anthony et al., 2015; Hansen et al., 2015; Heeszel et al., 2013; Pyle et al., 2010), providing a valuable and challenging test dataset for automatic event detection. Seismic events in Antarctica are not only associated with tectonic sources (*e.g.*, Lough et al., 2013, 2018; Rowe et al., 2000) but are also caused by other natural phenomena, such as iceberg calving signals (*e.g.*, Chen et al., 2011; Riel et al., 2021; Winberry et al., 2020; Zoet et al., 2012) or ice-stream slip (*e.g.*, Guerin et al., 2021; Hudson et al., 2023; Nettles & Ekström, 2010; Winberry et al., 2014; Walter et al., 2011, 2015), which are collectively classified as icequakes. Our study focuses on a subset of seismic data recorded by 19 stations deployed in the Victoria Land region of East Antarctica (Fig. 1), which provide continuous seismic recordings for several years. Most of these stations (15) were part of the Transantarctic Mountains Northern Network (TAMNNET), which operated between 2012-2015 (Hansen, 2012; Hansen et al., 2015); however, we also incorporated data from two additional networks (ER, GT; Fig. 1; ASL/USGS, 1993). This dataset allows us to provide unique constraints on polar seismic activity and to evaluate automated event detection performance in a region with limited station coverage.

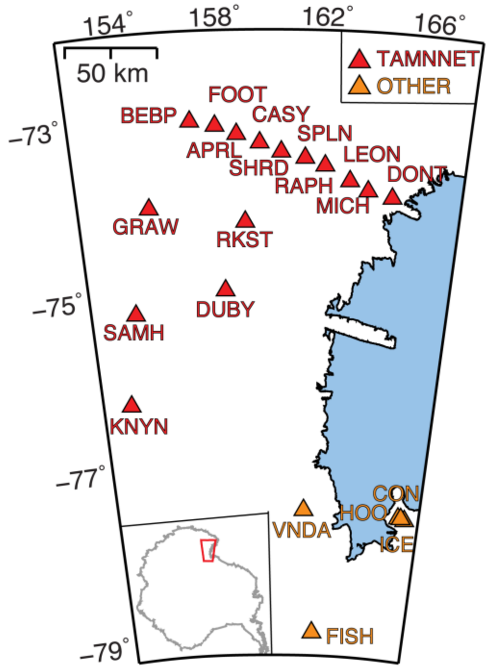

Figure 1. Map highlighting the examined seismic stations in Victoria Land, East Antarctica. Red triangles denote TAMNNET stations (Hansen et al., 2015), and orange triangles denote stations from other networks. Station names are also provided. The location of the main map in relation to the rest of Antarctica is highlighted in the inset on the lower left.

We developed a comprehensive workflow to assess the performance of different automated event detection techniques (Fig. 2). The continuous waveforms recorded by the East Antarctic stations (Fig. 1) were used to develop three starting catalogs, based on the STA/LTA, MF, and EQT machine learning approaches, respectively. Each catalog was then used to fine-tune a series of deep learning models via transfer learning, and their performance was evaluated with various metrics. The fine-tuned detection approach that worked best for our Antarctic dataset was then applied to update the three catalogs, and the events were relocated using a uniform velocity model. Each analysis step is described in detail in the following sections.

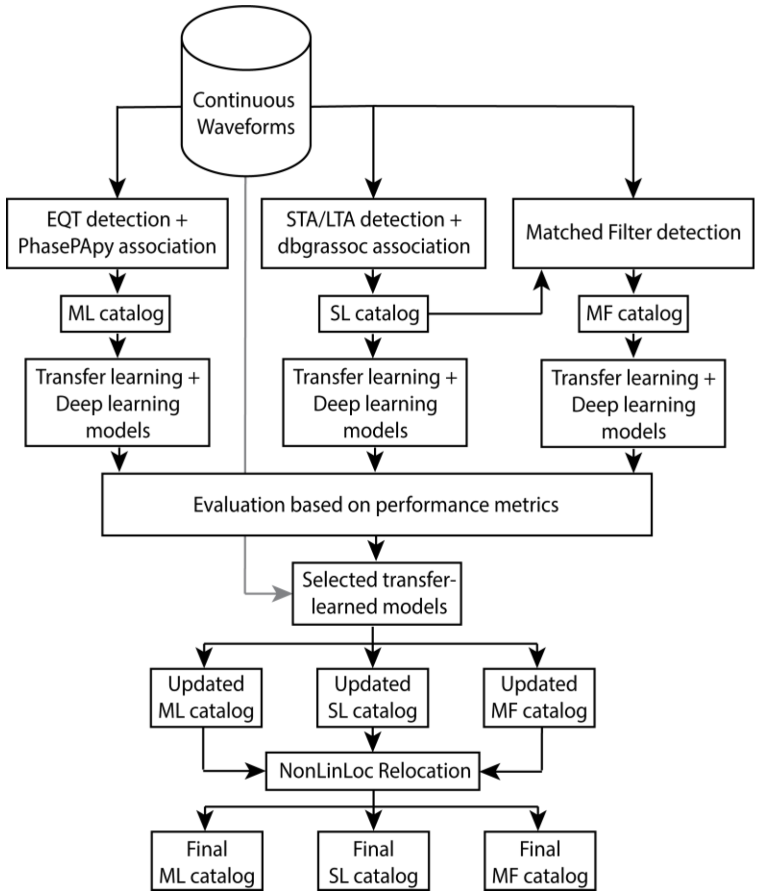

Figure 2. Flowchart summarizing the different automated seismic detection techniques examined in our study and the associated analysis steps.

3. Automated Detection Approaches

As shown in Figure 2, three different automated event detection approaches were initially evaluated by our study, including the STA/LTA method, the MF technique, and a machine learning-based approach using the EQT algorithm. The following subsections highlight the contributions and limitations of each approach as they are applied to our East Antarctic dataset (Fig. 1).

3.1. STA/LTA Method (SL Catalog)

The STA/LTA method (Allen, 1978; Earle & Shearer, 1994) detects high-frequency events in continuous data by identifying signals that have a mean energy ratio above some specified threshold. The STA window contains the dominant frequency of the events the algorithm aims to detect, while the LTA window contains mostly background noise, which should exceed the period of the lowest frequency seismic signal of interest (Trnkoczy, 2009). In continuous data, a trigger is declared when the STA/LTA ratio at any sample point surpasses a pre-defined threshold, indicating that an event is possibly occurring (Allen, 1978; Baer & Kradolfer, 1987). The algorithm remains in this triggered state until the ratio decreases below a specified trigger-release threshold (Fig. 3). One of the strengths of the STA/LTA method is that it does not require any prior knowledge about an event's waveform nor its source (Yoon et al., 2015); however, it does have limitations. For instance, S-waves may not be accurately detected if they arrive within the P-wave coda, and this can be problematic because S-waves are important when trying to determine the depth and origin time for an earthquake. The STA/LTA method is also highly sensitive to the level of noise in the data, and it may not perform well with dense earthquake sequences and/or emergent arrivals (Schaff & Beroza, 2004).

For our study, we designated short-term and long-term window lengths of 0.5 and 8.0 s, respectively. We also set the SNR trigger and trigger-release thresholds to 5 and 2.5, respectively (Fig. 3). Detections were associated with the Antelope dbgrassoc association module (BRTT, 2011), using a pre-computed travel-time grid based on the IASP91 reference velocity model (Kennett & Engdahl, 1991), and events were declared if they were recorded by at least four stations. Between 2012-2014, 560 events were detected using the STA/LTA approach and automatic association, thereby forming our SL catalog (Fig. 2). The data were then bandpass filtered between 2-5 Hz to highlight the signals of interest, and all phase arrivals were manually reviewed and adjusted, as needed. These additional processing steps allowed us to refine our SL catalog of high-quality events with well-determined phase arrivals.

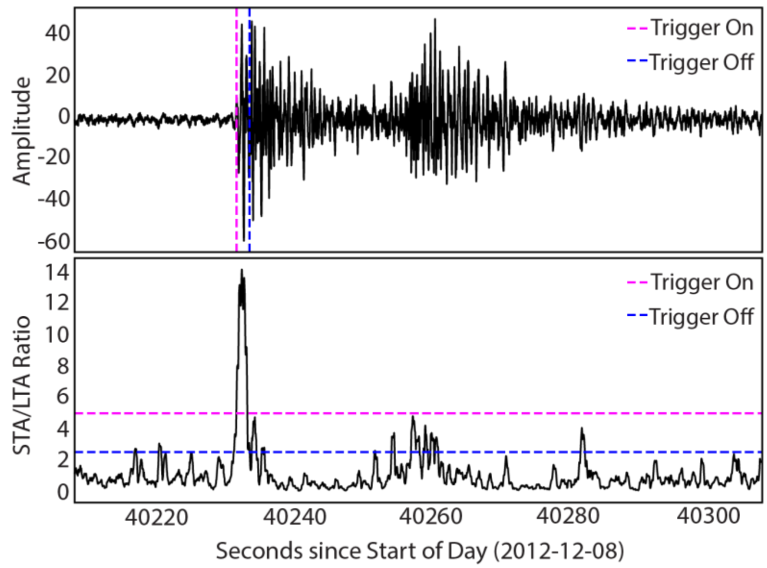

Figure 3. Example illustrating STA/LTA detection thresholds. The upper panel shows an event waveform that was detected by the STA/LTA approach, and the lower panel shows the STA/LTA ratio for the triggered event. Pink lines denote the trigger threshold (5) and trigger time; blue lines denote the trigger release threshold (2.5) and corresponding time.

3.2. Matched Filter Approach (MF Catalog)

The MF technique, also known as template matching or network-based waveform cross-correlation (Gibbons & Ringdal, 2006; Peng & Zhao, 2009; Shelly et al., 2007), provides another approach to automatically detect seismic arrivals, which is based on waveform similarity. Pre-defined template waveforms are cross-correlated with continuous data over successive windows, and signals exceeding a specified correlation threshold are identified as detections (Fig. 4). Generally, the MF approach performs better than the STA/LTA method (Sect. 3.1) when dealing with low SNR data. However, since the template events are often manually determined, the MF method can be time consuming during its initial stages when building the template catalog (if one does not already exist from a regional seismic network or other source). Furthermore, since the approach relies on waveform similarity, seismic signals that differ significantly from the template events may go undetected, leading to an incomplete catalog (Cianetti et al., 2021; Li et al., 2018; Yoon et al., 2015).

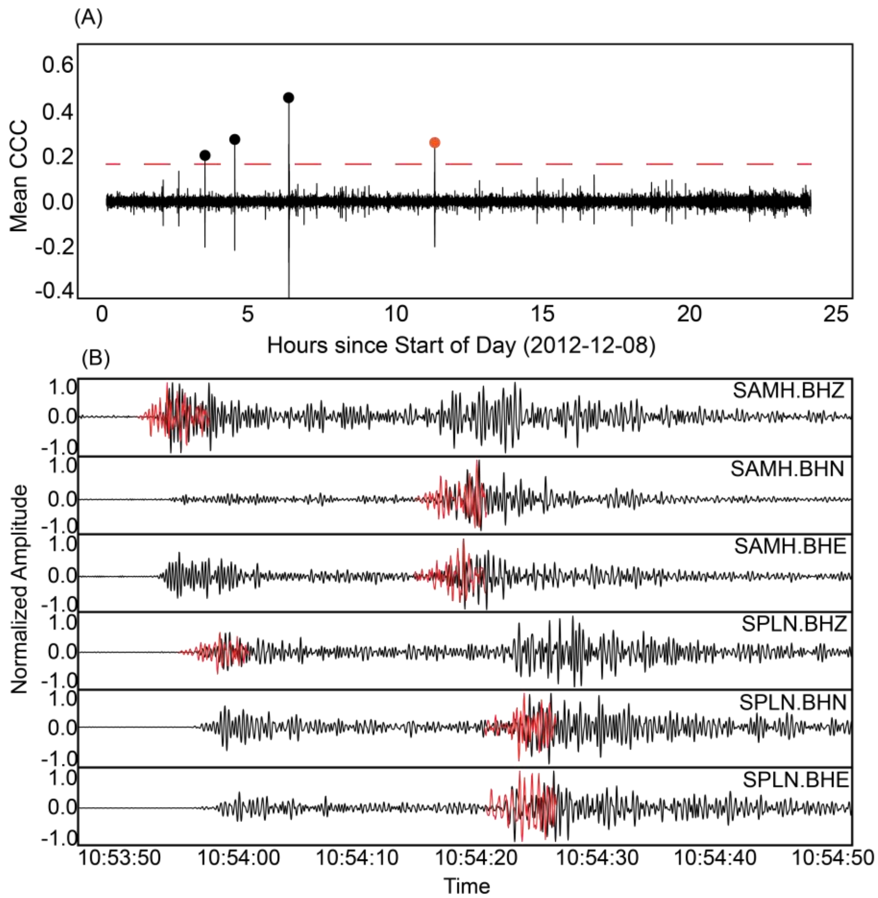

Figure 4. (A) Mean cross-correlation coefficients (CCC) determined by matching a template event, which occurred at 06:13:14 on 2012-12-08, against a full day (2012-12-08) of continuous data. Dots denote detections whose CCC values exceed the detection threshold, which is twelve times the MAD (red dashed line). The orange dot marks the detected event shown in panel (B). (B) Examples illustrating waveform cross-correlation. Template waveforms (red) are plotted on top of the continuous data (black), highlighting detected events from the MF approach. Station names and components are indicated on the right. Amplitudes have been normalized so their absolute maximum values are equal to one. This was done to better illustrate the waveform comparisons.

Using EQcorrscan (Chamberlain et al., 2018), all identified events in the SL catalog were treated as template events (Fig. 2), which were cross-correlated with the bandpass filtered (2-5 Hz) continuous data to identify additional seismic signals (Fig. 4). This bandpass was chosen based on close examination of the template coda, the density of seismic stations in the region, as well as our prior experience working with Antarctic data, where higher frequency information can become scattered by the ice sheet (Bentley & Kohnen, 1976) and thus incoherent when attempting template matching. Each template event was defined by the portion of the waveform 0.5 s before the event's P-wave arrival and 6 s after its S-wave arrival (Peng et al., 2014). The templates were shifted by 0.025 s (1 sample) increments through the continuous waveforms, and correlation coefficients were computed for each increment. Mean correlation coefficients were then determined by stacking the coefficient values across all stations and components (Fig. 4). The relative quality of each cross-correlated, matched waveform was evaluated using the median absolute deviation (MAD; Shelly et al., 2007), which is a measure of dispersion calculated as the median of the absolute difference between each data point for the mean correlation coefficient. The MAD value helps to estimate the variability in data distribution due to uncorrelated noise, thereby providing a robust measure to identify outliers. For a normally distributed dataset, the standard deviation is 1.4826 times the MAD (Hampel, 1974). Due to the noisy nature of real seismic data and the relatively long-period bandpass chosen for this project, a conservative threshold of 12 times the MAD was chosen, and signals that exceed this MAD value are identified as positive event detections (Fig. 4; *e.g.*, Skoumal et al., 2015; Yao et al., 2021).

A time domain, phase-pick SNR threshold was also applied to further ensure robust detections. For a given phase, the SNR was calculated by taking the maximum amplitude of the signal window and dividing it by the root-mean-square of the noise window. The noise windows start 6 s before the phase of interest, and both the signal and noise windows had lengths of 5.5 s (Fig. S1 in Supporting Information). The SNR threshold was subsequently determined by comparing the pick-specific SNR values obtained from all detected picks for each seismic event.

This additional processing step is not only important for robust event detections, but it also helps to remove unwanted signals, such as teleseismic events that originate from distant earthquakes. Sometimes teleseismic signals can be mistakenly detected in MF catalogs for local events, and this can adversely affect the accuracy of local event detections because teleseismic events have unique seismic waves and frequency contents (Waldhauser & Schaff, 2007). We determined that maintaining a SNR greater than 2.0 for both the P and S picks (Fig. S1 in Supporting Information) effectively helps to limit the influence of teleseismic events and reduces the number of false detections. With the MAD and SNR criteria applied, our MF catalog includes 4,577 local events (Fig. 2).

3.3. Machine Learning Approach (ML Catalog)

In addition to the STA/LTA and the MF techniques, we also utilized EQT, a machine learning-based signal detector and phase picker that was trained on a diverse seismic dataset (Mousavi et al., 2020). Further details about EQT and its architecture are provided in Section 4.1. We implemented the EQT picker within the easyQuake Python package (Walter et al., 2021) to identify P- and S-wave picks within the continuous data. The easyQuake associator, which is a modified version of PhasePApy (Chen & Holland, 2016), was used to aggregate pick information and declare event detections. Probability thresholds of 0.1, 0.1, and 0.3 were specified for the P-wave picks, S-wave picks, and event detections, respectively. In total, 1,728 events were detected in the East Antarctic dataset, which compose our initial machine learning (ML) catalog (Fig. 2). It should be noted that this catalog is distinguished from those derived from transfer-learning in later sections because it was generated using phase picks that were based on the original model and parameters specified by Mousavi et al. (2020).

4. Transfer Learning

Each of the catalogs described in Sections 3.1-3.3 were used in a transfer learning process to adapt a series of pre-trained deep learning models. Instead of retraining an entire model from scratch with randomly initialized parameters or different model architecture, a strategy called fine-tuning is employed, where the original model and its architecture serve as the starting point, and training continues with newly added data, thereby refining the model (Pan & Yang, 2010). Transfer learning not only leads to better model performance, but it also overcomes some of the limitations of traditional models that assume training and testing datasets are independent and are identically distributed (Tan et al., 2018).

The effectiveness of transfer learning has been proven in various fields (Long et al., 2013, 2015; Pan et al., 2011), and while its adoption within the field of seismology has been relatively limited so far, the technique demonstrates promising potential. For instance, Zhu et al. (2019) used a CNN-based Phase-Identification Classifier (CPIC), which was initially trained on a dataset with 30,146 labeled phases from the aftershock sequences of the 2008 M_w 7.9 Wenchuan earthquake, to develop a more complete aftershock catalog for the same area. Additionally, when fine-tuned on a smaller dataset from Oklahoma, the CPIC achieved 97% accuracy. This study highlights the potential for transfer learning applications to identify events in regions with no or few labeled phases. In a different study, Chai et al. (2020) enhanced the capabilities of the PhaseNet model (Zhu & Beroza, 2019), which was originally trained on data from regional seismic networks, to efficiently handle microseismic data from South Dakota. About 3,600 three-component seismograms and associated manual picks were used in the transfer learning process, and the performance of the retrained model exceeded that of the original PhaseNet model by over 10% in terms of precision and recall (see Sect. 4.3). Compared to human expert detections, 32% fewer P-wave picks were made, but the fine-tuned model identified 48% more S-wave picks.

We implemented our transfer learning process with Seisbench, a toolbox for machine learning in seismology (Ho, 2024; Münchmeyer et al. 2022; Woollam et al., 2022). Various deep learning model architectures were utilized, including PhaseNet (Zhu & Beroza, 2019), BasicPhaseAE (Woollam et al., 2019), GPD (Ross et al., 2018), and EQT (Mousavi et al., 2020), which are more fully described in Section 4.1. These models were selected given their distinct yet interrelated approaches to seismic signal processing. Additionally, these models share a common approach in terms of pre-processing the seismic data. Regardless of their specific architectures or use cases, they all rely on uniformly sampled data, typically at 100 Hz. If the original data has a different sampling rate, it is resampled to ensure uniformity. The data windows used by these models vary in length, but they all incorporate multiple types of seismic signals, including P-waves, S-waves, and noise, within their respective networks.

4.1 Deep Learning Models

The PhaseNet CNN (Zhu & Beroza, 2019) was developed as a U-Net architecture, which functions as an encoder-decoder mechanism that pulls significant features from input data and subsequently expands them to generate predictions of equivalent size outputs (Ronneberger et al., 2015). While the U-Net was initially created for a broad range of image processing applications, this approach has been adapted for earthquake phase detection. Three-component seismograms are sampled using 30 s windows that include both P- and S-wave arrivals, and these samples serve as the input for PhaseNet. The waveform data are then processed through an iterative down-sampling and up-sampling procedure. During down-sampling, the encoder reduces the dimensionality of the raw seismic data and extracts essential features associated with the seismic phase arrivals. The condensed information provided by the encoder is then increased in dimensionality through up-sampling by the decoder, which converts the information into detailed probability distributions for P-waves, S-waves, and noise at each point in time (Goodfellow et al., 2016; Zhu & Beroza, 2019). For seismic applications, PhaseNet was originally trained and evaluated using 779,514 waveforms containing labeled P- and S-wave arrivals from local earthquakes recorded in northern California (Zhu & Beroza, 2019).

BasicPhaseAE, which is another U-Net-like CNN phase detector, employs three 6 s input windows, with each window sampling an individual component (Woollam et al., 2019). The structure of BasicPhaseAE is similar to PhaseNet, but it differs in a few aspects. BasicPhaseAE uses smaller filter sizes and omits convolutions without stride, which refers to the step size that the filter matrix moves across the input matrix during the convolution process. In addition, BasicPhaseAE lacks residual connections, which are essentially shortcuts or bypass routes that enable the gradient to be back-propagated directly to earlier layers (Woollam et al., 2019; Münchmeyer et al., 2022). The input data, which consists of labels or classes of seismic data (*e.g.*, P-waves, S-waves, noise), undergo several transformations. Convolutional operations first extract the characteristic features for each class. During training, the model uses a designated 6 s window of data that is then divided into sequential sub-windows, each 0.4 s in length. The sub-windows are randomly shuffled to prevent the CNN from learning irrelevant temporal patterns. Extracted features then undergo multiple resampling stages, with a rectified linear unit activation function applied at each stage. The final architecture comprises three convolutional layers and three up-sampling layers. The network ultimately determines the probability of a P-wave, S-wave, or noise for every time sample in the input window. BasicPhaseAE was initially trained and evaluated using 11,000 waveforms from earthquakes located within the Iquique region in northern Chile (Woollam et al., 2019).

The GPD model is a phase identification CNN with six layers, including four convolution layers and two fully connected layers (Ross et al., 2018). Rectified linear units serve as the activation function for each layer, and batch normalization is applied throughout. GPD operates on a short 4 s input window that advances five samples (0.05 s) after each prediction to create a new, slightly overlapped 4 s window for the next prediction (Münchmeyer et al., 2022). Each advanced window is then classified as a P-wave arrival, S-wave arrival, or noise. The GPD model was originally trained and evaluated using 4.5 million three-component seismic records, evenly distributed amongst P- and S-wave seismograms and noise (Ross et al., 2018). Using a multi-class cross-entropy loss for training, the GPD model has been shown to effectively detect and identify seismic phases in various datasets (Münchmeyer et al., 2022; Woollam et al., 2022).

EQT is a model designed for simultaneous seismic event detection, phase identification, and onset timing determination. This model was originally trained on a portion of the STEAD dataset (Mousavi et al., 2019), a global collection of 1.2 million hand-labeled earthquake and noise waveforms. EQT operates on 60 s windows of three-component seismic data. Its architecture comprises a deep encoder and three separate decoders, and it integrates convolution, long short-term memory (LSTM) units, residual connections, and attention mechanisms (Mousavi et al., 2020). The encoder processes the seismic data into high-level contextual representations, while the decoders convert these representations into probability sequences for events as well as for P- and S-wave detections. LSTM, which resembles human auditory memory processing, and attention mechanisms, which simulate selective focusing in high-resolution areas, work in tandem to enhance the model's performance (Gers et al., 1999). The attention mechanisms function on two levels: global for earthquake events and local for phases within those events. During training, EQT employs data augmentation techniques, such as adding Gaussian noise, introducing gaps, and removing channels, which are implemented to enhance the model's robustness, teaching it how to handle various real-world data imperfections and irregularities. This helps to improve its overall performance and generalization ability (Mousavi et al., 2020).

Each of the above models has a different level of complexity, adaptability, and suitability for seismic datasets. For example, since BasicPhaseAE lacks residual connections, which are shortcuts that skip one or more layers to help train deep neural networks, its learning efficiency may be lower compared to PhaseNet (Münchmeyer et al., 2022). Compared to EQT, GPD is much slower, but it requires less memory. Further, the sophisticated EQT architecture and its comprehensive functionality may require more computational resources for complex analyses. We evaluate the performance of each model in relation to one another using our East Antarctic catalogs described in Sections 3.1-3.3, but it should be emphasized that the most suitable model for a given investigation depends on the type of data, the available processing time, and the computational resources available. We did not evaluate the relative computational performance of the specific algorithms in this study.

4.2 Applying Transfer Learning to the East Antarctic Catalogs

Each of the pre-trained models described in the previous section were fine-tuned via transfer learning using each of the event catalogs (Sects. 3.1-3.3). The SL, MF, and ML catalogs contain a total of 1,536, 13,731, and 5,388 waveform segments, respectively. The metadata for each catalog were assembled into a QuakeML-formatted file, and we also developed HDF5-formatted files by combining the event metadata with the waveforms, similar to the STEAD dataset format (Mousavi et al., 2019), for inclusion into Seisbench (Ho, 2024; Woollam et al., 2022). Each catalog was divided into a training subset, which is composed of 70% of the data, a validation subset, which contains 15% of the data, and a testing subset, which includes the remaining 15% of the data. The training subset was used to adjust the model's weights and biases during the transfer learning process, while the validation subset was used to fine-tune the model's hyperparameters. The validation subset was also essential in determining which model iteration performed the best, using the parameters described in Section 4.3. Once the optimal model configuration was identified based on the validation subset's results, the updated model was then evaluated on the testing subset. The final, reported results (Section 5) are based on this evaluation of the testing subset, thereby ensuring an unbiased assessment of each models' performance on unseen data.

Using the Münchmeyer et al. (2022) data augmentation techniques within SeisBench (Woollam et al., 2022), we built training pipelines, which are a series of steps that prepare and transform the waveform data for model training. Since our waveforms are long compared to each aforementioned model input length, a two-step approach was employed for window selection. First, for two-thirds of the training subset, windows were selected to ensure that they contained at least one labeled pick. For the remaining one-third, the windows were randomly selected from the entire waveform, and they may or may not include labeled picks. This approach guarantees that the training subsets are not overwhelmed by noise samples, which is particularly important for models with short input windows (*e.g.*, PhaseNet, BasicPhaseAE, GPD). The same approach was also applied to the validation subset.

Additionally, as part of the transfer learning process for each catalog, we employed the Adam optimizer (Kingma & Ba, 2014), which efficiently updates the model parameters to minimize the error between predicted and actual values. A corresponding learning rate of 0.001 was selected, which controls the magnitude of changes made to the model parameters during updates and ensures a steady convergence without overshooting (*i.e.*, where the model might skip over the optimal parameters). Further, a batch size of 256 was used in the optimizer, which means that 256 training samples were processed together during each iteration. This helps to balance computational efficiency and the quality of the model's gradient estimation (Coleman et al., 2017; Smith, 2018). Early stopping was also employed to obtain an optimal model. This strategy halts the training when the validation loss (a measure of prediction error) throughout the entire training subset fails to improve after ten successive cycles (epochs).

4.3. Evaluating Model Performance

To evaluate each fine-tuned, deep learning model’s ability to differentiate between seismic events and noise, we adopted the approach of Münchmeyer et al. (2022). First, a 30 s window of a random seismic waveform from either the validation or testing subset is analyzed to determine if it contains an event onset (*i.e.*, a first arriving seismic wave). Noise samples are also extracted from the window using labeled noise traces, if present. Otherwise, the noise sample is defined based on the presence or absence of P-wave and S-wave arrivals. That is, windows containing neither P- nor S-wave arrivals are labeled as noise, while those with either or both are labeled as an event. The event and noise labels were used as “ground truth” to compare with our models’ predictions.

A variety of metrics are used to evaluate the performance of each model. First, to assess a model’s ability to accurately identify event onsets while minimizing false positives, we examined the receiver operating characteristics (ROC), the area under the curve (AUC), and the F1 score. The ROC describes the true and false positive rates across all possible detection thresholds, allowing for different trade-offs between these rates, depending on the application scenario (Fawcett, 2006). For example, in early earthquake warning systems, a high true positive rate is important to ensure timely alerts, even if it means getting some false alarms (Meier et al., 2020). Alternatively, in a tomography research setting, where detection precision might be prioritized, reducing false positives could be more important, even if it means potentially missing some weaker seismic events. The AUC is a single value that defines the area under the ROC curve. It quantifies the overall ability of the model to distinguish between positive and negative classes. An AUC of one indicates a perfect model, meaning the model can identify all events correctly without any false positives. Conversely, an AUC of 0.5 represents a random model (Hanley & McNeil, 1982). The F1 score is the harmonic mean of the precision (*i.e.*, the number of correct detections among all detections) and recall (*i.e*., the number of detections among all possible detections). It serves as a combined measure of the model's sensitivity and specificity. As part of the transfer learning process, the AUC value is selected to optimize the F1 score, thereby fine-tuning the model to achieve an optimal trade-off between the false positive rate and the true positive rate.

In order to measure each model's binary classification performance, we used the Matthews Correlation Coefficient (MCC). It is ambiguous to assign P and S phases as positive and negative classes, and the MCC is insensitive to class assignment (Chicco & Jurman, 2020; Matthews, 1975; Münchmeyer et al., 2022). We analyzed 10 s windows containing exactly one phase arrival to determine if that arrival is a P- or an S-wave. The MCC is calculated as the correlation coefficient of the confusion matrix, and its value ranges from -1 (total disagreement) to 1 (full agreement). Even in cases of class imbalance, the MCC provides an appropriate measure for binary classification performance (Münchmeyer et al., 2022; Powers, 2011). Further, the MCC value was selected to optimize the phase threshold, which is used to calibrate the P- and S-wave pick probability thresholds. The pick probability indicates the likelihood of a specific data point corresponding to a seismic phase arrival (*i.e.*, a P- or an S-wave signal), where a higher probability directly correlates with a heightened level of confidence from the model regarding the presence of an arrival at the identified data point. For the P pick threshold, we multiplied the detection threshold by the square root of the phase threshold. This adjustment enhances the P-wave detection sensitivity and improves identification of these arrivals. For the S pick threshold, we adopted a more conservative approach, dividing the detection threshold by the square root of the phase threshold. This approach was taken to minimize the risk of false positives.

Finally, we evaluated each model's ability to accurately determine the onset time of phase arrivals within a given catalog. Using the same 10 s window used for the MCC assessment, we calculated the pick residuals, which are the differences between the transfer-learning-based pick times and the labeled pick times from the validation subset. The residual distribution is analyzed using both the root-mean-square error (RMSE) and the mean absolute error (MAE). Lower values of RMSE and MAE indicate greater accuracy in predicting the phase arrival onset times. Together, these provide a comprehensive evaluation given their different performance, with RMSE being sensitive to outliers and MAE being less sensitive to them (Willmott & Matsuura, 2005).

5. Results of Transfer Learning

The performance metrics (Sect. 4.3) used to evaluate the four deep learning models (Sect. 4.1) applied to each catalog (Sects. 3.1-3.3) elucidate the effects of transfer learning, and these metrics are summarized in Tables 1-3. Generally, transfer learning has a positive effect on all models, as is evident from the AUC metrics, for example. The most dramatic change was observed for the ML catalog and the BasicPhaseAE model, where the AUC increased from 0.45 to 0.81. That said, even models like GPD that already had a high AUC value (0.87) saw an increase (0.90). These results highlight the benefits of transfer learning. However, it is important to consider how each model defines an event detection. For instance, EQT needs both P- and S-wave labels to declare a detection within the seismogram time series (data from other stations is commonly aggregated during event association, discussed later), while GPD and PhaseNet do not. For scenarios where datasets might lack certain labels, such as in our SL and MF catalogs, this could lead to reduced performance, as reflected in the metric results. It is worth noting that our results are qualitatively comparable to those made by Münchmeyer et al. (2022) for the ETHZ dataset (Woollam et al., 2022), where some P- or S-wave labels were missing.

The RMSE and MAE metrics were reduced for both P and S picks across all catalogs, again indicating improved performance from the fine-tuning and transfer learning. Among all the models, EQT had the lowest of these metrics, indicating it had the highest pick accuracy. However, GPD also displayed significant improvements in RMSE and MAE and closely followed EQT across all catalogs (Tables 1-3). As for the MCC metrics, where higher values indicate better classification performance, every model exhibited a MCC rise following transfer learning. Comparing the three catalogs (Tables 1-3), the P and S picks are notably better classified in the ML catalog for all models, followed by the SL and then the MF catalog. These variations might be due to discrepancies in P- and S-wave labeling consistency across the catalogs. For example, the starting ML catalog was exclusively generated using EQT, perhaps leading to higher pick consistency and, as a result, lower RMSE and MAE values. As a result, variations in performance across the three catalogs reveal that the efficiency of transfer learning also depends on the consistency and quality of the training subset.

Figure 5 shows an example of the pick probabilities for different deep learning models when applied to continuous data. EQT, GPD, and PhaseNet all have improved pick probabilities after transfer learning. The BasicPhaseAE pick probabilities did not increase post-transfer learning, and this could be due to the shorter input windows used by this model, together with its shorter filters and missing residual connections (Münchmeyer et al., 2022).

Table 1. Fine-tuned metric results before (left columns) and after (right columns) transfer learning was applied to the ML catalog. AUC: Area under the Curve; RMSE: root-mean-square error; MAE: mean absolute error; MCC: Matthews Correlation Coefficient.

Model	AUC		P picks RMSE		S picks RMSE		P picks MAE		S picks MAE		MCC	
PhaseNet	0.7	0.8	3.0	2.1	3.0	2.3	2.2	1.4	2.2	1.5	0.3	0.6
BasicPhaseAE	0.4	0.7	3.2	2.3	3.0	2.5	2.5	1.6	2.3	1.7	0.3	0.5
GPD	0.8	0.8	2.2	1.8	2.3	2.1	1.5	1.2	1.6	1.4	0.6	0.8
EQTransformer	0.7	0.7	3.4	1.8	3.0	1.8	2.4	1.1	2.1	1.1	0.6	0.9

Table 2. Fine-tuned metric results before (left columns) and after (right columns) transfer learning was applied to the MF catalog. Columns are the same as in Table 1.

Model	AUC		P picks RMSE		S picks RMSE		P picks MAE		S picks MAE		MCC	
PhaseNet	0.7	0.9	2.8	1.1	2.4	1.2	1.8	0.5	1.6	0.6	0.3	0.7
BasicPhaseAE	0.4	0.8	3.2	1.1	2.8	1.3	2.5	0.6	2.0	0.7	0.3	0.7
GPD	0.8	0.9	1.2	0.6	1.3	0.8	0.6	0.3	0.7	0.4	0.7	0.9
EQTransformer	0.8	0.9	2.7	0.6	2.2	0.5	1.6	0.3	1.2	0.2	0.7	1.0

Table 3. Fine-tuned metric results before (left columns) and after (right columns) transfer learning was applied to the SL catalog. Columns are the same as in Tables 1 and 2.

Model	AUC		P picks RMSE		S picks RMSE		P picks MAE		S picks MAE		MCC	
PhaseNet	0.7	0.8	2.0	1.4	2.4	2.0	1.2	0.8	1.6	1.2	0.4	0.8
BasicPhaseAE	0.4	0.7	2.8	1.7	2.7	2.2	2.0	1.0	1.9	1.4	0.4	0.7
GPD	0.8	0.9	1.4	0.9	2.0	2.0	0.8	0.6	1.3	1.2	0.8	0.9
EQTransformer	0.8	0.8	2.7	0.9	2.3	1.9	1.6	0.5	1.4	1.1	0.7	1.0

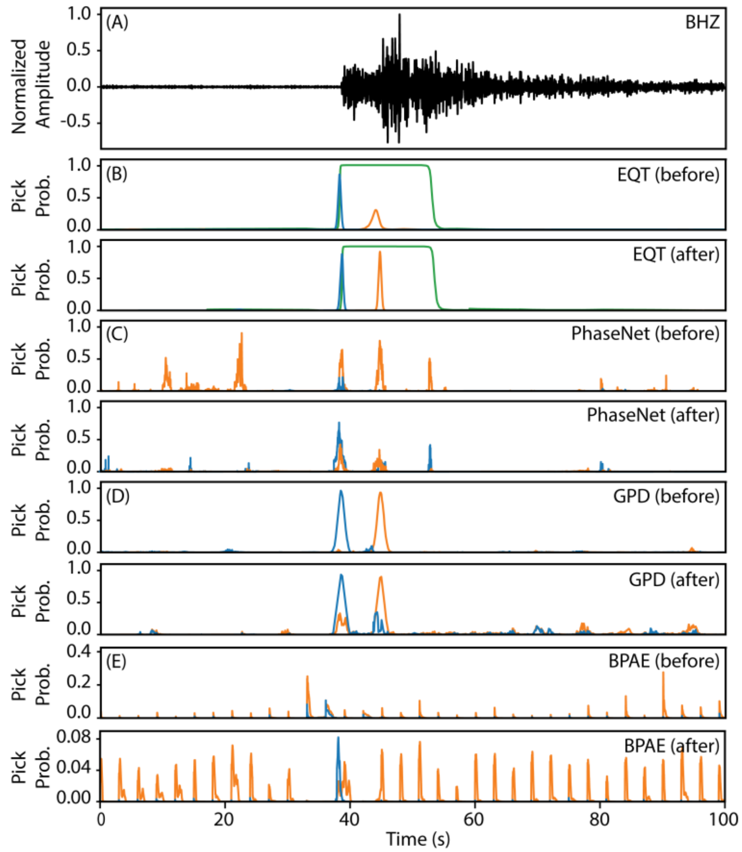

Figure 5. (A) Sample of the continuous Antarctic data recorded by station LEON (Fig. 1), and corresponding pick probabilities for (B) EQT, (C) PhaseNet, (D) GPD, and (E) BasicPhaseAE (BPAE). For each model, the top and bottom panels show the pick probabilities before and after transfer learning, respectively (note that the vertical scales can vary by panel). Blue lines correspond to P-waves, and orange lines correspond to S-waves. For EQT, the green lines show the detection probability.

6. Model Assessment

6.1 Benefits and Limitations of Each Automated Event Detection Approach

Each automated event detection approach has its benefits and limitations, and the choice of which approach to use depends on the objective of the study and the characteristics of the dataset. The STA/LTA method stands out given its minimal pre-processing requirements, straightforward algorithm, and low computational demands, making this technique efficient and readily applicable. Notably, the approach can also identify low magnitude earthquakes if the data has sufficiently high quality (Fig. 6). However, as noted in Section 3.1, STA/LTA can struggle to identify emergent or low SNR arrivals (Schaff & Beroza, 2004; Yoon et al., 2015), which can make this technique more prone to errors, including an increased risks of false positive detections and/or detection failures (Kato et al., 2012). This limitation is partly due to the nature of the STA and LTA window lengths, which are not adjusted during the detection process (Trnkoczy, 2009) and hence restrict the method's ability to adapt to varying seismic signal characteristics. Figure S2 in the Supporting Information shows several examples of missed detections that resulted from the STA/LTA inflexibility. Given its performance, STA/LTA is likely suitable for real-time seismic event detection applications, particularly in situations where an existing, trained model is not available. This method is applicable for systems such as earthquake early warning and volcanic monitoring, which require rapid results. It is important to note that in these scenarios, the immediate availability of results may be prioritized, even if it means accepting a higher likelihood of false positive detections for lower magnitude events (*e.g.*, Kumar et al., 2018; Li et al., 2016; Meier et al., 2020; Tepp, 2018).

The MF approach detects events with high precision, particularly if the events have a high degree of waveform similarity. However, developing a comprehensive set of template events can be time consuming, and the need to compare each of those templates to the continuous data can be computationally demanding (Liu et al., 2020; Meng et al., 2012). Further, since the MF technique is heavily dependent on the pre-defined templates, it is susceptible to missing events that diverge from recognized patterns (Gardonio et al., 2019; Kato & Nakagawa, 2014; Peng & Zhao, 2009; Ross et al., 2018). Several examples of such missed events are shown in Figure S3 in the Supporting Information. Automatic event detection with this method is best-suited to environments where the seismic events are self-similar, such as volcanic-related seismic swarms (*e.g.*, Tan et al., 2023; Whidden et al., 2023; Wimez & Frank, 2022) and repeating stick-slip activity beneath glaciers (*e.g.*, Helmstetter, 2022; Lucas et al., 2023; Ma et al., 2020).

Deep learning event detection techniques can help to address some of the problems faced by the STA/LTA and the MF approaches. Since deep learning models can be trained to recognize intricate seismic patterns, this approach has a greater degree of adaptability across a range of seismic signals and noise. Our analysis also illustrates how deep learning model performance can be further enhanced via transfer learning, where pre-trained models are adapted to recognize the characteristics of unique seismic sources (Chai et al., 2020; Liao et al., 2021). That said, deep learning approaches, with or without transfer learning, have their own set of challenges. ML methods are generally computationally intensive and do not provide rapid results (García et al., 2022; Zhu et al., 2022). Their performance is strongly linked to the quality and volume of their training subsets, and the oft-cited 'black box' nature of ML makes its decision-making processes ambiguous (Gonzalez Garibay et al., 2023). The effectiveness of transfer learning depends on whether the pre-trained model is relevant to the target dataset. If there is a mismatch between the source and target architecture, there is a risk of negative transfer, where the pre-trained model may fail to effectively adapt to the new task (Civilini et al., 2021; Zhou et al., 2021). Careful fine-tuning of the pre-trained model is needed to ensure its applicability to the specific seismic context, and this requires a certain level of understanding regarding the model's architecture. All that said, seismic event catalogs based on ML models typically have a greater magnitude of completeness (*i.e.*, the minimum magnitude above which all events have been detected) compared to those generated by other approaches (Fig. 6; *e.g.*, Ma & Chen, 2022; Reynen & Audet, 2017; Ross et al., 2018), Therefore, if a given study requires robust, extensive seismic constraints, the additional computational resources and complexity of ML algorithms are worth the investment.

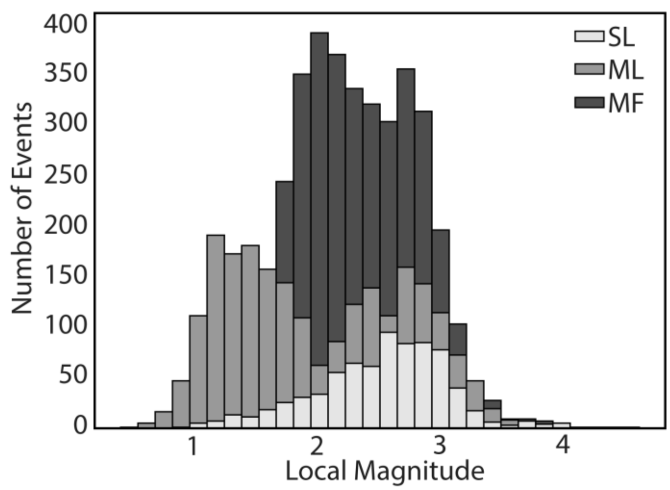

Figure 6. Histogram summarizing the number of events in each catalog after transfer learning was applied, along with their corresponding local magnitude estimates. Light grey bars represent the SL catalog, medium grey bars denote the ML catalog, and dark grey bars correspond to the MF catalog.

6.2 Preferred ML model for East Antarctica

The metrics discussed in Sections 4.3 and 5 provide important information regarding the most applicable model for a given seismic study. For our East Antarctic investigation, we prioritized thorough seismic event detection. While it is important to identify events accurately and precisely, the limited seismic station coverage in our study region (Fig. 1) emphasizes the need to develop an event catalog that is as complete as possible. As such, our ideal model is one that strikes a balance between sensitivity and accuracy, and our extensive analyses indicate that the fine-tuned GPD model is an optimal choice. While EQT displays somewhat better pick accuracy, as indicated by its RMSE and MAE values, its ability to distinguish between positive and negative classes (AUC score) lags behind GPD (Tables 1-3). The trained GPD model's high AUC score emphasizes that this model robustly distinguishes true events from noise. That is, events with low SNR, potentially overlooked by other models and methods, are identified by

GPD. Furthermore, the inherent variability of seismic data demands a model that performs consistently, and the GPD model displays consistent performance across all three examined catalogs, both before and after transfer learning is applied (Tables 1-3). This indicates that the GPD model is highly adaptable, regardless of the data's origin.

7. Application

7.1 GPD Results for Each Catalog

We applied the fine-tuned (transfer-learned) GPD model to the full suite of East Antarctic data (2012-2015; Fig. 1), running three versions of the GPD detection algorithm concurrently, corresponding to our SL, MF, and ML catalogs. As noted in Section 5, each model generates pick probabilities for the designated P- and S-wave arrivals (Fig. 5). Picks with probabilities below a specified threshold (Table 4) are discarded. These thresholds are essential for reducing the number of spurious picks, thereby enhancing the accuracy and reliability of the detected seismic events, and the thresholds ultimately control the number of event identifications. Table 4 summarizes the corresponding pick probability thresholds used to determine qualifying P- and S-waves. These thresholds have led to the identification of new seismic events post-transfer learning. Specifically, after transfer learning, the number of new events in the SL, ML, and MF catalogs is 618, 372, and 201, respectively.

Table 4. P- and S-wave pick probability thresholds for the three transfer-learned catalogs. A P- or S-wave pick is declared if the probability exceeds the specified threshold.

Catalog	P Threshold	S Threshold
ML	0.68	0.81
MF	0.42	0.51
SL	0.51	0.60

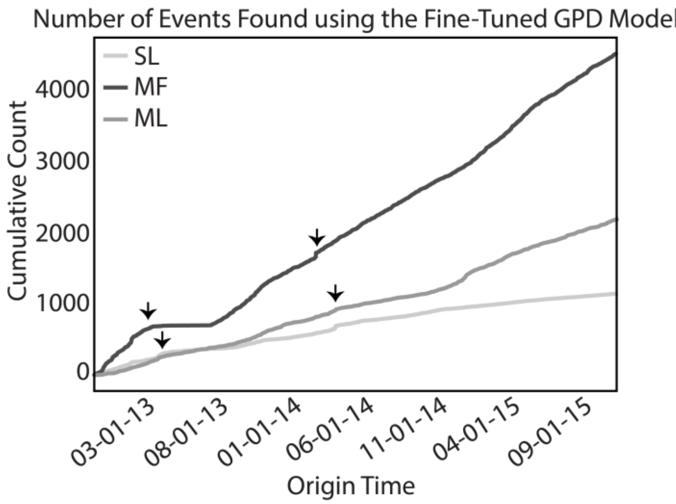

Figure 7. Cumulative number of events included in each catalog after transfer learning was applied. The light grey line corresponds to the SL catalog, the dark grey line corresponds to the MF catalog, and the medium grey line corresponds to the ML catalog. Arrows denote time periods where an increased number of events are observed.

All three catalogs display an increase in the number of events around May 2013 and May 2014 (Fig. 7). These time periods correlate with seasonal changes in Antarctica as the austral winter sets in. Tensile stresses in the ice sheet can be influenced by temperature, and this can impact the formation of crevasses (Harper et al., 1998; Holdsworth, 1969). Specifically, when temperatures drop, the surface layers of the ice sheet can become substantially colder than the underlying firn, and this temperature gradient subjects the colder, more brittle surface layers to an increase in tensile stress. Consequently, new crevasses may form and propagate along the ice sheet surface (Nath & Vaughan, 2003), thereby leading to an increased number of icequakes. This may explain the increase in detected events at these particular time intervals (Fig. 7). Local magnitudes (M_L) were also computed for the SL, ML, and MF catalog events (Fig. 6), though we note that the magnitudes were determined using amplitude attenuation parameters developed for southern California (Hutton & Boore, 1987). While not specific to our study region, these parameters do not impact our assessment since our goal was to simply determine relative event magnitudes rather than to make any interpretations of absolute magnitude. As shown in Figure 6, all three techniques effectively detect low magnitude ($M_L \leq 3$) seismic events, though the ML technique detects a higher number of signals with magnitudes below two.

7.2 Event Relocations

After the fine-tuned GPD model was applied to the full East Antarctic dataset, as described in Section 7.1, the events from each of the updated catalogs were relocated using the NonLinLoc software package (Lomax et al., 2000). An equal differential-time likelihood function and the Oct-Tree sampling approach were used to compute the maximum likelihood hypocenters, based on the corresponding probability density functions (PDFs; Lomax et al., 2000; Zhou, 1994). We also utilized a modified version of the crustal velocity model (Fig. S4 in Supporting Information) from Pyle et al. (2010), which was developed for a nearby region in East Antarctica. Only earthquakes with at least four P- and S-wave arrival times were relocated. Additionally, to account for any possible bias in the procedure, we performed a second inversion using the average arrival-time residuals at each station (Lomax et al., 2009), thereby leading to better constrained event locations.

For each event relocation, the average horizontal and vertical uncertainties of the confidence ellipsoid, which are estimated by the PDFs, were used to determine the volume of the 68% confidence ellipse. This, in turn, was used to determine the average uncertainty (R_e) of each event location (Lomax et al., 2000). The relocated events in each catalog were then grouped based on their uncertainty thresholds. The best constrained event locations (Group A) had $R_e \leq 5$ km. Groups B, C, and D had progressively higher R_e values (Group B: $5 < R_e \leq 10$ km; Group C: $10 < R_e \leq 20$ km; Group D: $R_e \geq 20$ km), indicating less well-constrained locations. The number of events in each quality group is provided in Table S1 in the Supporting Information. Figure 8 highlights event locations that had $R_e \leq 10$ km (*i.e.*, Groups A and B) within each catalog, and events from all groups are shown in Figure S5 in the Supporting Information.

Many of the detected events in all three catalogs are situated near David Glacier (Fig. 8). Shallow events (< 5 km) in this region are consistent with those identified in previous studies (*e.g.*, Bannister & Kennett, 2002; Danesi et al., 2007, 2022; Zoet et al., 2012; 2013), which have been attributed to stick-slip behavior at the base of the ice sheet. However, all three catalogs also show deeper events (> 10 km) beneath the David Glacier region as well, which could be associated with solid Earth processes. For example, movement and mass redistribution within the East Antarctic ice sheet may induce stress changes in the underlying lithosphere, creating the deep-seated events highlighted in our catalogs (Lund, 2015; Steffen, 2013; Steffen et al., 2020).

All three event catalogs also show notable seismicity beneath Victoria Land, in the northeastern portion of the study region (Fig. 8). The prevalence of event detections in this area may reflect some degree of spatial bias given the locations of the stations available for this study (Fig. 1). The TAMNNET stations, in particular, provide somewhat better coverage in this region; therefore, nearby events may more likely meet the enforced minimum number of P- and S-wave arrivals needed for relocation. That said, the Victoria Land event cluster (Fig. 8) is concentrated near several other glaciers that move across the Transantarctic Mountains and towards the Ross Sea, including the Campbell, Priestley, and Aviator Glaciers. The best located events in this cluster are relatively shallow and therefore may reflect ice-bed processes, similar to those suggested for David Glacier further to the south. Deeper events are also seen beneath this region, down to about 25-50 km, which are more likely associated with tectonic processes, such as faulting (*e.g.*, Pisarska-Jamrozy et al., 2018), or with crustal deformation driven by cryospheric fluctuations (*e.g.*, Stewart et al., 2000). Further investigations would be needed to evaluate the sources of the seismic events beneath David Glacier and Victoria Land, but the automatically identified events from our analyses provide some insight into the complex relationship between the solid Earth structure and the Antarctic ice sheet.

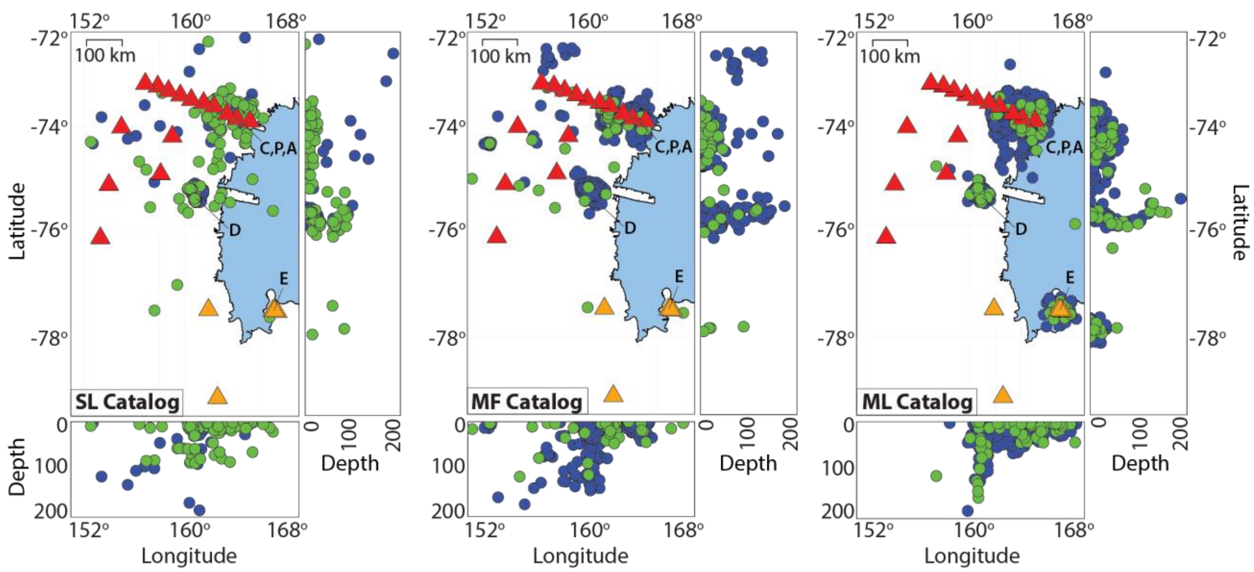

Figure 8. Seismic event relocations from NonLinLoc for quality Group A and B events. From left to right: SL catalog, MF catalog, and ML catalog. Blue circles denote events that were detected by the corresponding original technique (*i.e.*, STA/LTA, template matching, EQT machine learning). Green circles denote new events detected by transfer learning. Red triangles indicate TAMNNET stations, and orange triangles denote other stations. Abbreviations denote key locations including David Glacier (D), Campbell, Priestly, and Aviator Glaciers (C,P,A), and Mount Erebus (E).

It is also worth noting the cluster of seismic events near Mount Erebus, which was uniquely identified by the ML catalog (Fig. 8). Some prior studies that have also recognized seismicity in this region attribute the events to small magnitude icequake sources near the volcano's summit (Chaput et al., 2015; Li et al., 2021; Podolskiy & Walter, 2016). Other investigations have attributed the Mount Erebus seismicity to volcanic activity within its shallow magmatic system (*e.g.*, Aster et al., 2008; Hansen & Schmandt, 2015; Kaminuma, 1987; Rowe et al., 1998, 2000). The absence of the Mount Erebus event cluster in the SL and MF catalogs underscores the effectiveness of deep learning techniques in seismic detection, particularly in elucidating events with a range of sources.

8. Conclusions

Our study has evaluated the benefits and limitations of different automated seismic event detection methods, and our results emphasize that the most appropriate approach depends on the specific attributes of the examined data as well as the objectives of a given study. The STA/LTA method is well-suited for real-time event detection applications that require rapid results, even if there is a higher likelihood for false detections. The MF technique works well for environments that generate seismic events with a high degree of waveform similarity. Deep learning models offer the most adaptability if dealing with a range of seismic sources and noise, and their performance can be enhanced with transfer learning, which provides an effective approach to adapt pre-trained models for unique datasets.

For our East Antarctic investigation, the fine-tuned GPD model, characterized by its high AUC score, reliable picking accuracy, and consistent performance across the examined catalogs, emerged as the most robust, providing new insights into seismic sources in the region. Event relocations based on the fine-tuned catalogs offer new insights into potential seismic sources, including both shallow cryospheric and deeper tectonic processes. Arguably, the most comprehensive seismic event catalog would be one created by integrating the results from each of the applied detection methods; however, the separate results highlight the performance of each approach. Our findings have expanded seismic event detections in East Antarctica, including the identification of previously unrecognized seismicity, and these results underscore the potential for automated event detection approaches to enhance our understanding of seismic activity even in areas with limited station coverage.

Acknowledgments

Funding for this research was provided by the National Science Foundation (grants OPP-1914698 (SEH), EAR-1745135 (JIW), and EAR-1925965 (ZP)), the Ministerio de Ciencia e Innovación / Agencia Estatal de Investigació (MCIN/AEI/10.13039/501100011033), and the European Union Next Generation EU "Plan de Recuperación, Transformación y Resiliencia" (PRTR; grant Model_SHaKER, PID2021-124155NB-C31 (JLSR)).

Open Research

The facilities of SAGE Data Services, particularly the SAGE Data Management Center, were used to access the waveforms and related metadata used in this project (http://ds.iris.edu/mda/). Specifically, we employed datasets from ASL/USGS (1993) and Hansen (2012). SAGE Data Services are funded through the Seismological Facilities for the Advancement of Geoscience (SAGE) Award of the National Science Foundation under Cooperative Support Agreement EAR-1851048. Some figures were generated with the Generic Mapping Tools software (Wessel et al., 2013). Our ML catalog was developed with the easyQuake software package (Walter et al., 2021; Walter, 2022), and transfer learning was implemented with Seisbench (Ho, 2024; Woollam et al., 2022).

References

Albuquerque Seismological Laboratory (ASL)/USGS (1993). Global Telemetered Seismograph Network [Dataset]. International Federation of Digital Seismograph Networks, doi: 10.7914/SN/GT.

Allen, R. (1978). Automatic earthquake recognition and timing from single traces. *Bulletin of the Seismological Society of America*, *68*(5), 1521–1532, doi: 10.1785/BSSA0680051521

Anandakrishnan, S., Voigt, D. E., Burkett, P. G., Long, B., & Henry, R. (2000). Deployment of a broadband seismic network in West Antarctica. *Geophysical Research Letters*, *27*(14), 2053-2056, doi: 10.1029/1999GL011189

Anthony, R. E., Aster, R. C., Wiens, D., Nyblade, A., Anandakrishnan, S., Huerta, A., Winberry, J. P., Wilson, T. & Rowe, C. (2015). The seismic noise environment of Antarctica. *Seismological Research Letters*, *86*(1), 89-100, doi: 10.1785/0220140109

Aster, R., Zandomeneghi, D., Mah, S., McNamara, S., Henderson, D. B., Knox, H., & Jones, K. (2008). Moment tensor inversion of very long period seismic signals from Strombolian eruptions of Erebus Volcano. *Journal of Volcanology and Geothermal Research*, 177(3), 635-647, doi: 10.1016/j.jvolgeores.2008.08.013

Baer, M., & Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. *Bulletin of the Seismological Society of America*, *77*(4), 1437–1445, doi: 10.1785/BSSA0770041437

Bannister, S., & Kennett, B. L. N. (2002). Seismic Activity in the Transantarctic Mountains - Results from a Broadband Array Deployment. *Terra Antarctica*, *9*, 41-46.

Bentley, C. R., & Kohnen, H. (1976). Seismic refraction measurements of internal friction in Antarctic ice. *Journal of Geophysical Research: Solid Earth*, *81*(8), 1519-1526, doi: 10.1029/JB081i008p01519

Boulder Real Time Technologies (BRTT; 2011). Evolution of the Commercial ANTELOPE Software.

Chai, C., Maceira, M., Santos-Villalobos, H. J., Venkatakrishnan, S. V., Schoenball, M., Zhu, W., Beroza, G. C., Thurber, C., & EGS Collaboration Team (2020). Using a deep neural network and transfer learning to bridge scales for seismic phase picking. *Geophysical Research Letters*, *47*(16), doi: 10.1029/2020GL088651

Chamberlain, C. J., Hopp, C. J., Boese, C. M., Warren-Smith, E., Chambers, D., Chu, S. X., Michailos, K., & Townend, J. (2018). EQcorrscan: Repeating and near-repeating earthquake detection and analysis in Python. *Seismological Research Letters*, *89*(1), 173-181, doi: 10.1785/0220170151

Chaput, J., Campillo, M., Aster, R. C., Roux, P., Kyle, P. R., Knox, H., & Czoski, P. (2015). Multiple scattering from icequakes at Erebus volcano, Antarctica: Implications for imaging at glaciated volcanoes. *Journal of Geophysical Research: Solid Earth*, 120(2), 1129-1141, doi: 10.1002/2014JB011278

Chen, X., Shearer, P. M., Walter, F., & Fricker, H. A. (2011). Seventeen Antarctic seismic events detected by global surface waves and a possible link to calving events from satellite images. *Journal of Geophysical Research: Solid Earth*, *116*(B6), doi: 10.1029/2011JB008262

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

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1), doi: 10.1186/s12864-019-6413-7

Cianetti, S., Bruni, R., Gaviano, S., Keir, D., Piccinini, D., Saccorotti, G., & Giunchi, C. (2021). Comparison of deep learning techniques for the investigation of a seismic sequence: An application to the 2019, Mw 4.5 Mugello (Italy) earthquake. *Journal of Geophysical Research: Solid Earth*, *126*(12), doi: 10.1029/2021JB023405

Civilini, F., Weber, R. C., Jiang, Z., Phillips, D., & Pan, W. D. (2021). Detecting moonquakes using convolutional neural networks, a non-local training set, and transfer learning. *Geophysical Journal International*, *225*(3), 2120-2134, doi: 10.1093/gji/ggab083

Coleman, C., Narayanan, D., Kang, D., Zhao, T., Zhang, J., Nardi, L., et al. (2017). Dawnbench: An end-to-end deep learning benchmark and competition. *Training*, *100*(101), 102.

Danesi, S., Bannister, S., & Morelli, A. (2007). Repeating earthquakes from rupture of an asperity under an Antarctic outlet glacier. *Earth and Planetary Science Letters*, *253*, 151-158, doi: 10.1016/j.epsl.2006.10.023

Danesi, S., Salimbeni, S., Borghi, A., Urbini, S., & Frezzotti, M. (2022). Cryo-seismicity triggered by ice mass discharge through the Antarctic subglacial hydrographic network. *The Cryosphere Discussions*, doi: 10.5194/egusphere-2022-29.

Earle, P. S., & Shearer, P. M. (1994). Characterization of global seismograms using an automatic-picking algorithm. *Bulletin of the Seismological Society of America*, 84(2), 366-376, doi: 10.1785/BSSA0840020366

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861-874, doi: 10.1016/j.patrec.2005.10.010

García, J. E., Fernández-Prieto, L. M., Villaseñor, A., Sanz, V., Ammirati, J. B., Díaz Suárez, E. A., & García, C. (2022). Performance of deep learning pickers in routine network processing applications. *Seismological Society of America*, *93*(5), 2529-2542, doi: 10.1785/0220210323

Gardonio, B., Campillo, M., Marsan, D., Lecointre, A., Bouchon, M., & Letort, J. (2019). Seismic activity preceding the 2011 M w 9.0 Tohoku earthquake, Japan, analyzed with multidimensional template matching. *Journal of Geophysical Research: Solid Earth*, 124(7), 6815-6831, doi: 10.1029/2018JB016751

Gers, F. A., Schmidhuber, J., & Cummins, F. (1999). Learning to forget: continual prediction with LSTM, *Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, Edinburgh, UK, 850-855, doi: 10.1049/cp:19991218.

Gibbons, S. J., & Ringdal, F. (2006). The detection of low magnitude seismic events using array-based waveform correlation, *Geophysical Journal International*, *165*(1), 149–166, doi: 10.1111/j.1365-246X.2006.02865.x

González Garibay, M., Primc, K., & Slabe-Erker, R. (2023). Insights into advanced models for energy poverty forecasting. *Nature Energy*, *8*, 903-905, doi: 10.1038/s41560-023-01311-x

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press, ISBN: 9780262035613.

Guerin, G., Mordret, A., Rivet, D., Lipovsky, B. P., & Minchew, B. M. (2021). Frictional origin of slip events of the Whillans Ice Stream, Antarctica. *Geophysical Research Letters*, *48*(11), doi: 10.1029/2021GL092950.

Hampel, F. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, *69*, 383-393, doi: 10.1080/01621459.1974.10482962

Han, J., Kim, S., Sheen, D. H., Lee, D., Lee, S. J., Yoo, S. H., & Park, D. (2023). Seismic event and phase detection using deep learning for the 2016 Gyeongju earthquake sequence. *Geosciences Journal*, *27*(3), 285-295, doi: 10.1007/s12303-023-0004-y

Hanley, J. A. & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36, doi: 10.1148/radiology.143.1.7063747

Hansen, S. (2012). Transantarctic Mountains Northern Network [Dataset]. International Federation of Digital Seismograph Networks, doi: 10.7914/SN/ZJ_2012.

Hansen, S. E., Reusch, A., Parker, T., Bloomquist, D., Carpenter, P., Graw, J. H., & Brenn, G. R. (2015). The Transantarctic Mountains Northern Network (TAMNNET): Deployment and Performance of a Seismic Array in Antarctica. *Seismological Research Letters*, *86*, 1636-1644, doi: 10.1785/0220150117

Hansen, S. M., & Schmandt, B. (2015). Automated detection and location of microseismicity at Mount St. Helens with a large-N geophone array. *Geophysical Research Letters*, *42*(18), 7390-7397, doi: 10.1002/2015GL064848

Harper, J. T., Humphrey, N.F., & Pfeffer, W. T. (1998). Crevasse patterns and the strain-rate tensor: A high resolution comparison, *Journal of Glaciology*, *44*, 68-77, doi: 10.3189/S0022143000002367

Heeszel, D. S., Wiens, D. A., Nyblade, A. A., Hansen, S. E., Kanao, M., An, M., & Zhao, Y. (2013). Rayleigh wave constraints on the structure and tectonic history of the Gamburtsev Subglacial Mountains, East Antarctica. *Journal of Geophysical Research: Solid Earth*, *118*(5), 2138-2153, doi: https://doi.org/10.1002/jgrb.50171

Helmstetter, A. (2022). Repeating Low Frequency Icequakes in the Mont-Blanc Massif Triggered by Snowfalls. *Journal of Geophysical Research: Earth Surface*, *127*(12), doi: 10.1029/2022JF006837

Herrmann, M., & Marzocchi, W. (2021). Inconsistencies and lurking pitfalls in the magnitude–frequency distribution of high-resolution earthquake catalogs. *Seismological Research Letters*, 92(2A), 909-922, doi: 10.1785/0220200337

Ho, L. M. (2024). Transfer_Learning_and_Seisbench [Software], GitHub, https://github.com/longmho/Transfer_Learning_and_Seisbench.

Holdsworth, G. (1969). Primary transverse crevasses, *Journal of Glaciology*, *8*, 107-129.

Hudson, T. S., Kufner, S. K., Brisbourne, A. M., Kendall, J. M., Smith, A. M., Alley, R. B., et al. (2023). Highly variable friction and slip observed at Antarctic ice stream bed. *Nature Geoscience*, *16*, 612-618, doi: 10.1038/s41561-023-01204-4

Hutton, L. K., & Boore, D. M. (1987). The ML scale in south California, *Bulletin of the Seismological Society of America*, *77*(6), 2074-2094, doi: 10.1785/BSSA0740051827

Jiang, C., Fang, L., Fan, L., & Li, B. (2021). Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes. *Earthquake Science, 34*(5), 425-435, doi: 10.29382/eqs-2021-0038

Kaminuma, K. (1987). Seismic activity of Erebus volcano, Antarctica. *Pure and Applied Geophysics*, *125*, 993-1008, doi: 10.1007/BF00879364

Kato, A., Obara, K., Igarashi, T., Tsuruoka, H., Nakagawa, S., & Hirata, N. (2012). Propagation of slow slip leading up to the 2011 M w 9.0 Tohoku-Oki earthquake. *Science*, 335(6069), 705-708, doi: 10.1126/science.1215141

Kato, A., & Nakagawa, S. (2014). Multiple slow-slip events during a foreshock sequence of the 2014 Iquique, Chile Mw 8.1 earthquake. *Geophysical Research Letters*, 41(15), 5420-5427, doi: 10.1002/2014GL061138

Kennett, B. L. N., & Engdahl, E. R. (1991). Traveltimes for global earthquake location and phase identification. *Geophysical Journal International*, 105(2), 429-465, doi: 10.1111/j.1365-246X.1991.tb06724.x

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

Kumar, S., Vig, R., & Kapur, P. (2018). Development of Earthquake Event Detection Technique Based on STA/LTA Algorithm for Seismic Alert System. *Journal of the Geological Society of India*, 92, 679-686, doi: 10.1007/s12594-018-1087-3

Lapins, S., Goitom, B., Kendall, J. M., Werner, M. J., Cashman, K. V., & Hammond, J. O. (2021). A little data goes a long way: Automating seismic phase arrival picking at Nabro volcano with transfer learning. *Journal of Geophysical Research: Solid Earth, 126*(7), doi: 10.1029/2021JB021910.

Li, X., Shang, X., Wang, Z., Dong, L., & Weng, L. (2016). Identifying P-phase arrivals with noise: An improved Kurtosis method based on DWT and STA/LTA. *Journal of Applied Geophysics*, *133*, 50-61, doi: 10.1016/j.jappgeo.2016.07.022

Li, Z., Peng, Z., Hollis, D., Zhu, L., & McClellan, J. (2018). High-resolution seismic event detection using local similarity for Large-N arrays. *Scientific Reports*, *8*(1), doi: 10.1038/s41598-018-19728-w.

Li, C., Peng, Z., Chaput, J. A., Walter, J. I., & Aster, R. C. (2021). Remote triggering of Icequakes at Mt. Erebus, Antarctica by large teleseismic earthquakes. *Seismological Research Letters*, *92*(5), 2866-2875, doi: 10.1785/0220210027

Liao, W., Chen, X., Lu, X., Huang, Y., & Tian, Y. (2021). Deep transfer learning and time-frequency characteristics-based identification method for structural seismic response. *Frontiers in Built Environment*, *7*, doi: 10.3389/fbuil.2021.627058.

Liu, M., Li, H., Zhang, M., & Wang, T. (2020). Graphics processing unit-based match and locate (GPU-M&L): An improved match and locate method and its application. *Seismological Research Letters*, *91*(2A), 1019-1029, doi: 10.1785/0220190241

Lomax, A., Virieux, J., Volant, P., & Berge-Thierry, C. (2000). Probabilistic earthquake location in 3D and layered models. In *Advances in Seismic Event Location*, eds. C.H. Thurber & N. Rabinowitz, 101-134, doi: 10.1007/978094-015-9536-0_5.

Lomax, A., Michelini, A., Curtis, A., & Meyers, R. A. (2009). Earthquake location, direct, global-search methods. *Encyclopedia of Complexity and Systems Science*, *5*, 2449-2473, doi:10.1007/978-0-387-30440-3

Long, M., Wang, J., Ding, G., Sun, J., & Yu, P. S. (2013). Transfer Feature Learning with Joint Distribution Adaptation. *Proceedings of the IEEE International Conference on Computer Vision*, 2200–2207.

Long, M., Cao, Z., Wang, J., & Jordan, M. I. (2015). Learning transferable features with deep adaptation networks. *Proceedings of the 32nd International conference on machine learning*, 97-105.

Lough, A. C., Wiens, D. A., Barcheck, G.C., Anandakrishnan, S., Aster, R. C., Blankenship, D. D., et al. (2013). Seismic detection of an active subglacial magmatic complex in Marie Byrd Land, Antarctica. *Nature Geoscience*, *6*(12), 1031-1035, doi: 10.1038/ngeo1992

Lough, A. C., Wiens, D. A., & Nyblade, A. (2018). Reactivation of ancient Antarctic rift zones by intraplate seismicity. *Nature Geoscience*, *11*(7), 515-519, doi: 10.1038/s41561-018-0140-6

Lucas, E. M., Nyblade, A. A., Aster, R. C., Wiens, D. A., Wilson, T. J., Winberry, J. P., & Huerta, A. D. (2023). Tidally Modulated Glacial Seismicity at the Foundation Ice Stream, West Antarctica. *Journal of Geophysical Research: Earth Surface*, *128*(7), doi: 10.1029/2023JF007172.

Lund, B. (2015). Palaeoseismology of Glaciated Terrain. Encyclopedia of Earthquake Engineering, Springer, Berlin, doi: 10.1007/978-3-642-36197-5_25-1.

Ma, H., Chu, R., Sheng, M., & Yu, Z. (2020). Template matching for simple waveforms with low signal-to-noise ratio and its application to icequake detection. *Earthquake Science*, *33*, 256-263, doi: 10.29382/eqs-2020-0256-01

Ma, X., & Chen, T. (2022). Small Seismic Events in Oklahoma Detected and Located by Machine Learning–Based Models. *Bulletin of the Seismological Society of America*, *112*(6), 2859-2869.

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta (BBA)-Protein Structure*, *405*(2), 442-451, doi: 10.1016/0005-2795(75)90109-9

Meier, M. A., Kodera, Y., Böse, M., Chung, A., Hoshiba, M., Cochran, E., et al. (2020). How often can earthquake early warning systems alert sites with high-intensity ground motion?. *Journal of Geophysical Research: Solid Earth*, *125*(2), doi: 10.1029/2019JB017718

Meng, X., Yu, X., Peng, Z., & Hong, B. (2012). Detecting earthquakes around Salton Sea following the 2010 Mw7. 2 El Mayor-Cucapah earthquake using GPU parallel computing. *Procedia Computer Science*, *9*, 937-946, doi: 10.1016/j.procs.2012.04.100

Meng, X., Peng, Z., & Hardebeck, J. (2013). Seismicity around Parkfield correlates with static shear stress changes following the 2003 Mw6.5 San Simeon earthquake, *Journal of Geophysical Research*, 118(7), 3576-3591, doi:10.1002/jgrb.50271

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*, 179,464-179,476, doi: 10.1109/ACCESS.2019.2947848

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), doi: 10.1038/s41467-020-17591-w

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), doi: 10.1029/2021JB023499

Myers, S. C., Johannesson, G., & Hanley, W. (2007). A Bayesian hierarchical method for multiple-event seismic location. *Geophysical Journal International*, *171*(3), 1049–1063, doi: 0.1111/j.1365-246X.2007.03555.x

Nath, P. C., & Vaughan, D. G. (2003). Subsurface crevasse formation in glaciers and ice sheets. *Journal of Geophysical Research: Solid Earth*, *108*(B1), doi: 10.1029/2001JB000453

Nettles, M., & Ekström, G. (2010). Glacial earthquakes in Greenland and Antarctica. *Annual Review of Earth and Planetary Sciences*, *38*, 467-491, doi: 10.1146/annurev-earth-040809-152414

Neves, M., Chuang, L.Y., Li. W., Peng, Z., Figueiredo, P. M., & Ni. S. (2024). A dense microearthquake catalog reveals the complex rupture of the extremely shallow M5.1 Sparta, North Carolina Earthquake. *Communications Earth & Environment.*

Pan, S. J. & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345-1359, doi: 10.1109/TKDE.2009.191

Pan, S. J., Tsang, I. W., Kwok, J. T., & Yang, Q. (2011). Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks*, *22*(2), 199–210, doi: 0.1109/TNN.2010.2091281

Peng, Z., & Zhao, P. (2009). Migration of early aftershocks following the 2004 Parkfield earthquake. *Nature Geoscience*, *2*(12), 877-881, doi: 10.1038/ngeo697

Peng, Z., Walter, J. I., Aster, R. C., Nyblade, A., Wiens, D. A., & Anandakrishnan, S. (2014). Antarctic icequakes triggered by the 2010 Maule earthquake in Chile. *Nature Geoscience*, *7*(9), 677-681, doi: 10.1038/ngeo2212

Perol, T., Gharbi, M., & Denolle, M. (2018). Convolutional neural network for earthquake detection and location. *Science Advances*, *4*(2), doi: 10.1126/sciadv.170057

Pisarska-Jamroży, M., Belzyt, S., Börner, A., Hoffmann, G., Hüneke, H., Kenzler, M., et al. (2018). Evidence from seismites for glacio-isostatically induced crustal faulting in front of an advancing land-ice mass (Rügen Island, SW Baltic Sea). *Tectonophysics*, *745*, 338-348, doi: 10.1016/j.tecto.2018.08.004

Podolskiy, E. A., & Walter, F. (2016). Cryoseismology. *Reviews of Geophysics*, *54*(4), 708-758.

Powers, D.M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, *2*(1), 37-63, doi: 10.48550/arXiv.2010.16061

Pyle, M. L., Wiens, D. A., Nyblade, A. A., & Anandakrishnan, S. (2010). Crustal structure of the Transantarctic Mountains near the Ross Sea from ambient seismic noise tomography. *Journal of Geophysical Research: Solid Earth*, *115*(B11), doi: 10.1029/2009JB007081

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, doi: 10.1093/gji/ggx238

Riel, B., Minchew, B., & Bischoff, T. (2021). Data-Driven Inference of the Mechanics of Slip Along Glacier Beds Using Physics-Informed Neural Networks: Case Study on Rutford Ice Stream, Antarctica. *Journal of Advances in Modeling Earth Systems*, *13*(11), doi: 10.1029/2021MS002621

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention, 9351*, 234-241, doi: 10.1007/978-3-319-24574-4_28

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, doi: 10.1029/2017JB015251

Rowe, C. A., Aster, R. C., Kyle, P. R., Schlue, J. W., & Dibble, R. R. (1998). Broadband recording of Strombolian explosions and associated very-long-period seismic signals on Mount Erebus Volcano, Ross Island, Antarctica. *Geophysical Research Letters*, *25*(13), 2297-2300, doi: 10.1029/98GL01622

Rowe, C. A., Aster, R. C., Kyle, P. R., Dibble, R. R., & Schlue, J. W. (2000). Seismic and acoustic observations at Mount Erebus volcano, Ross Island, Antarctica, 1994–1998. *Journal of Volcanology and Geothermal Research*, *101*(1-2), 105-128, doi: 10.1016/S0377-0273(00)00170-0

Schaff, D. P., & Beroza, G. C. (2004). Coseismic and postseismic velocity changes measured by repeating earthquakes. *Journal of Geophysical Research: Solid Earth*, *109*(B10), doi: 10.1029/2004JB003011

Shelly, D. R., Beroza, G. C., & Ide, S. (2007). Non-volcanic tremor and low-frequency earthquake swarms. *Nature*, *446*(7133), 305-307, doi: 10.1038/nature05666

Si, X., Wu, X., Li, Z., Wang, S., & Zhu, J. (2024). An all-in-one seismic phase picking, location, and association network for multi-task multi-station earthquake monitoring. *Communications Earth & Environment*, 5(1), 22, doi: 10.1038/s43247-023-01188-4

Skoumal, R. J., Brudzinski, M. R., & Currie, B. S. (2015). Earthquakes induced by hydraulic fracturing in Poland Township, Ohio. *Bulletin of the Seismological Society of America*, *105*(1), 189-197, doi: 10.1785/0120140168

Smith, L. N. (2018). A disciplined approach to neural network hyper-parameters: Part 1--learning rate, batch size, momentum, and weight decay. *US Naval Research Laboratory Technical Report*, 5510-026, doi: 10.48550/arXiv.1803.09820

Steffen, R. (2013). The influence of glacial isostatic adjustment on intraplate seismicity in northeastern Canada. *Doctoral Thesis*, University of Calgary, Canada, doi: 10.11575/PRISM/28209

Steffen, R., Steffen, H., Weiss, R., Lecavalier, B. S., Milne, G. A., Woodroffe, S. A., & Bennike, O. (2020). Early Holocene Greenland-ice mass loss likely triggered earthquakes and tsunami. *Earth and Planetary Science Letters*, *546*, doi: 10.1016/j.epsl.2020.116443.

Stewart, I. S., Sauber, J., & Rose, J. (2000). Glacio-seismotectonics: ice sheets, crustal deformation and seismicity. *Quaternary Science Reviews*, *19*(14-15), 1367-1389, doi: 10.1016/S0277-3791(00)00094-9

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In *Artificial Neural Networks and Machine Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27* (pp. 270-279). Springer International Publishing.

Tan, D., Fee, D., Hotovec-Ellis, A. J., Pesicek, J.D., Haney, M.M., Power, J. A., & Girona, T. (2023). Volcanic earthquake catalog enhancement using integrated detection, matched-filtering, and relocation tools. *Frontiers in Earth Science*, *11*, doi: 10.3389/feart.2023.1158442

Tepp, G. (2018). A Repeating Event Sequence Alarm for Monitoring Volcanoes. *Seismological Research Letters*, *89*(5), 1863-1876, doi: 10.1785/0220170263

Trnkoczy, A. (2009). Understanding and parameter setting of STA/LTA trigger algorithm. In *New manual of seismological observatory practice,* Deutsches GeoForschungsZentrum GFZ, 1-20.

Vaezi, Y., & Van der Baan, M. (2015). Comparison of the STA/LTA and power spectral density methods for microseismic event detection. *Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society*, *203*(3), 1896-1908, doi: 10.1093/gji/ggv419

Waldhauser, F., & Schaff, D. (2007). Regional and teleseismic double-difference earthquake relocation using waveform cross-correlation and global bulletin data. *Journal of Geophysical Research: Solid Earth*, *112*(B12), doi: 10.1029/2007JB004938

Walter, J. I., E. E. Brodsky, S. Tulaczyk, S. Y. Schwartz, & R. Pettersson (2011). Transient slip events from near-field seismic and geodetic data on a glacier fault, Whillans Ice Plain, West Antarctica, *Journal of Geophysical Research: Solid Earth*, *116*, doi:10.1029/2010JF001754

Walter, J. I., I. Svelitsky, J. Fineberg, S. Tulaczyk, E. E. Brodsky, C. G. Barcheck, & S. P. Carter (2015). Rupture speed dependence on initial stress profiles: Insights from glacier and laboratory stick-slip, *Earth and Planetary Science Letters*, *411*, 112-120, doi: 10.1016/j.epsl.214.11.025

Walter, J. I., Ogwari, P., Thiel, A., Ferrer, F., & Woelfel, I. (2021). easyQuake: Putting machine learning to work for your regional seismic network or local earthquake study, *Seismological Research Letters*, *92*(1): 555–563, doi: 10.1785/0220200226

Walter, J. I. (2022). easyQuake [Software], GitHub, https://github.com/jakewalter/easyQuake.

Wessel, P., Smith, W. H., Scharroo, R., Luis, J., & Wobbe, F. (2013). Generic mapping tools: improved version released. *Eos, Transactions American Geophysical Union*, *94*(45), 409-410, doi: 10.1002/2013EO450001

Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*(1), 79-82, doi: 10.3354/cr030079

Wimez, M., & Frank, W. B. (2022). A recursive matched-filter to systematically explore volcanic long-period earthquake swarms. *Geophysical Journal International*, *231*(2), 912-920, doi: 10.1093/gji/ggac221

Winberry, J. P., Anandakrishnan, S., Alley, R. B., Wiens, D. A., & Pratt, M. J. (2014). Tidal pacing, skipped slips and the slowdown of Whillans Ice Stream, Antarctica. *Journal of Glaciology*, *60*(222), 795-807, doi: 10.3189/2014JoG14J038

Winberry, J. P., Huerta, A. D., Anandakrishnan, S., Aster, R.C., Nyblade, A.A., & Wiens, D. A. (2020). Glacial earthquakes and precursory seismicity associated with Thwaites Glacier calving. *Geophysical Research Letters*, *47*(3), doi: 10.1029/2019GL086178

Whidden, K. M., Petersen, G. M., Mesimeri, M., & Pankow, K. L. (2023). Analysis of the 2021 Milford, Utah earthquake swarm: Enhanced earthquake catalog and migration patterns. *Frontiers in Earth Science*, *11*, doi: 10.3389/feart.2023.1057982

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, doi: 10.1785/0220180312

Woollam, J., Münchmeyer, J., Tilmann, F., Rietbrock, A., Lange, D., Bornstein, T., & Soto, H. (2022). Seisbench-A toolbox for machine learning in seismology. *Seismological Research Letters*, *81*(3), 1695–1709, doi: 10.1785/0220210324

Wu, Y., Lin, Y., Zhou, Z., Bolton, D. C., Liu, J., & Johnson, P. (2018). DeepDetect: A Cascaded Region-Based Densely Connected Network for Seismic Event Detection, *IEEE Transactions on Geoscience and Remote Sensing*, *57*, 62-75, doi: 0.1109/TGRS.2018.2852302

Yao, D., Peng, Z., Kaneko, Y., Fry, B., & Meng, X. (2021). Dynamic triggering of earthquakes in the North Island of New Zealand following the 2016 Mw 7.8 Kaikōura earthquake. *Earth and Planetary Science Letters*, *557*, doi: 10.1016/j.epsl.2020.116723

Yuan, C., Ni, Y., Lin, Y., & Denolle, M. (2023). Better Together: Ensemble Learning for Earthquake Detection and Phase Picking. *IEEE Transactions on Geoscience and Remote Sensing*, doi: 10.1109/TGRS.2023.3320148

Yoon, C. E., O'Reilly, O., Bergen, K. J., & Beroza, G. C. (2015). Earthquake detection through computationally efficient similarity search. *Science Advances*, *1*(11), doi: 10.1126/sciadv.1501057

Zhou, H. (1994). Rapid three-dimensional hypocentral determination using a master station method. *Journal of Geophysical Research*, *99*(B8), doi: 10.1029/94JB00934

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, doi: 10.1785/0220180319

Zhou, R., Yao, X., Hu, G., & Yu, F. (2021). Learning from unlabelled real seismic data: Fault detection based on transfer learning. *Geophysical Prospecting*, *69*(6), 1218-1234.

Zhu, W. & Beroza, G. C. (2019). PhaseNet: A deep-neural-network-based seismic arrival-time picking method. *Geophysical Research Letters*, *45*(13), 6644-6653, doi: 10.1093/gji/ggy423

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 Mw 7.9 Wenchuan Earthquake. *Physics of the Earth and Planetary Interiors*, *293*, doi: 10.1016/j.pepi.2019.05.004

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), doi: 10.1029/2021JB023283

Zoet, L. K., Anandakrishnan, S., Alley, R. B., Nyblade, A. A., & Wiens, D. A. (2012). Motion of an Antarctic glacier by repeated tidally modulated earthquakes. *Nature Geoscience*, *5*(9), 623-626, doi: 10.1038/ngeo1555

Zoet, L. K., Alley, R. B., Anandakrishnan, S., & Christianson, K. (2013). Accelerated subglacial erosion in response to stick-slip motion. *Geology*, *41*(2), 159-162 doi: 10.1130/G33624.1

Journal of Geophysical Research: Machine Learning and Computation

Supporting Information for

Evaluating Automated Seismic Event Detection Approaches: An Application to Victoria Land, East Antarctica

L.M. Ho[1], J.I. Walter[2], S.E. Hansen[1], J.L. Sánchez-Roldán[3], and Z. Peng[4]

[1]The University of Alabama, Tuscaloosa, AL 35487
[2]Oklahoma Geological Survey, University of Oklahoma, Norman, OK 73019
[3]Universidad Complutense de Madrid, Madrid, Spain, 28040
[4]Georgia Institute of Technology, Atlanta, GA 30332

Contents of this file

Introduction

This supporting information provides several supplemental figures, which complement the discussion in the main text. These include our SNR threshold analysis for the MF approach, example events that were identified once transfer learning was applied, the velocity model used in our event relocation, and maps of all relocated events, regardless of their uncertainty thresholds. Additionally, a table summarizing the number of relocated events is provided.

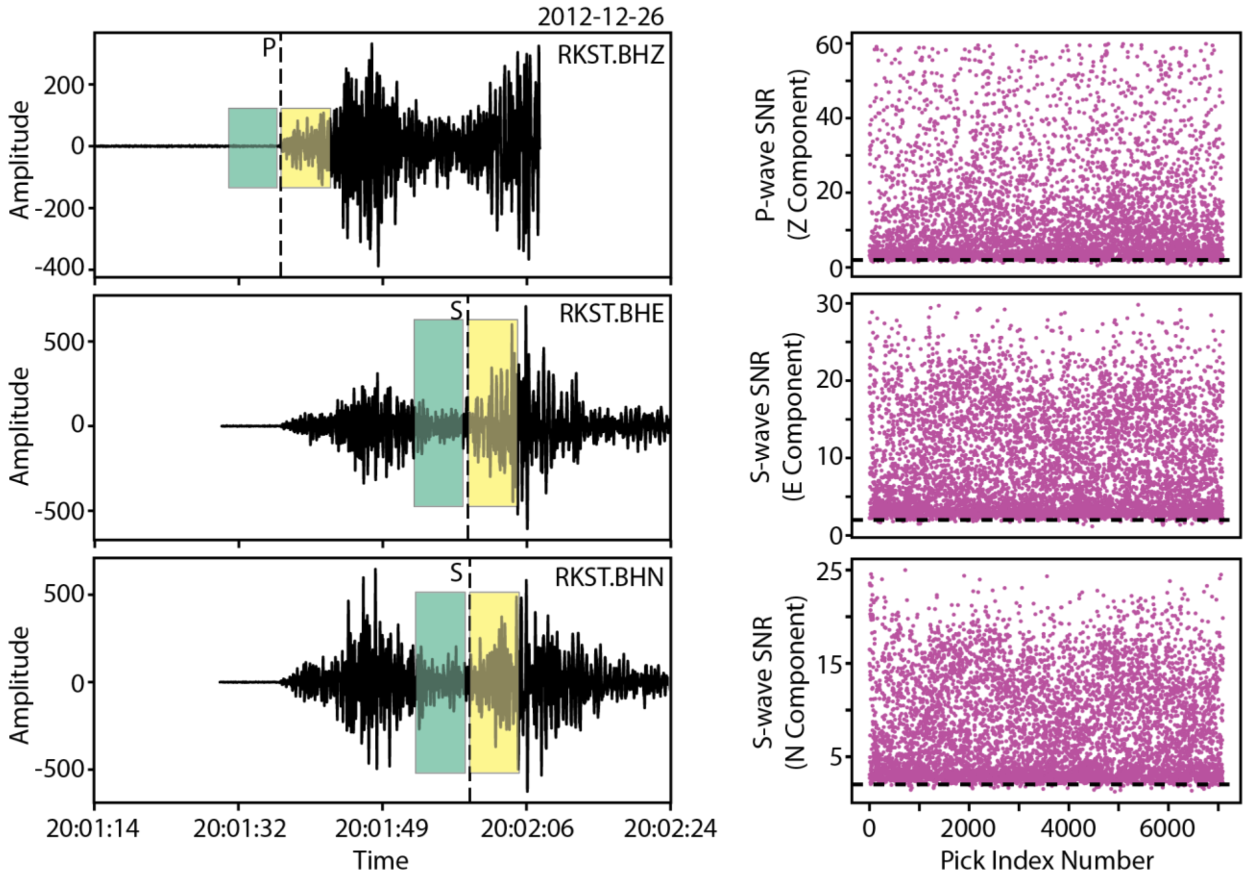

Figure S1. (left) Example three-component data from station RKST (Fig. 1) with the P- and S-wave arrivals from a December 26, 2012 event marked by vertical dashed lines. Green and yellow boxes highlight the portion of the waveform used to define the SNR noise and signal windows, respectively. (right) Scatter plots showing the pick-specific SNR values for all events in the initial MF catalog. Horizontal dashed lines mark the selected SNR of 2.0 applied to both our P- and S-wave picks. The x-axis just reflects the index (identification) numbers associated with each pick.

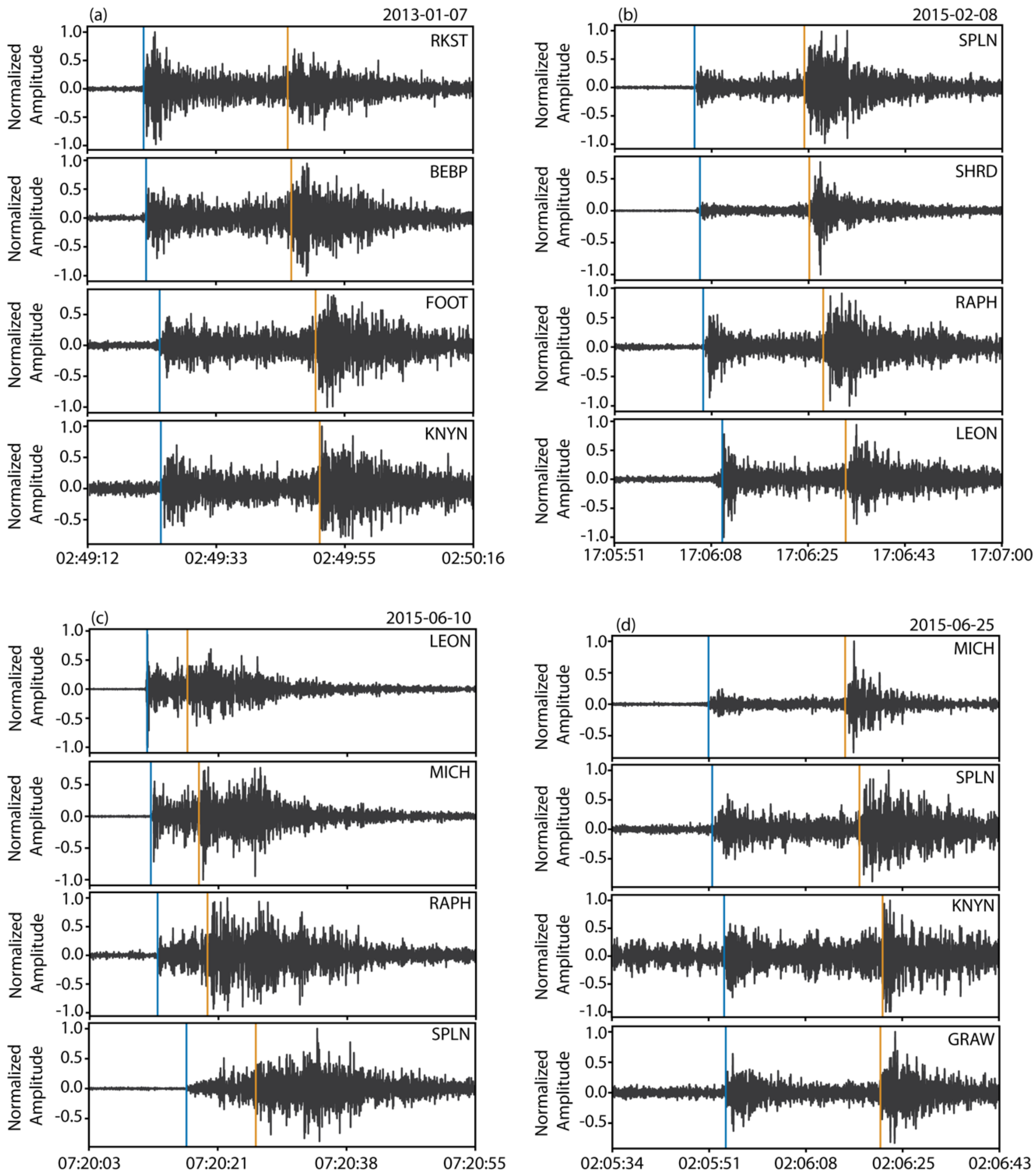

Figure S2. Four example events that were not identified by the STA/LTA technique and hence were not included in the initial SL catalog; however, once transfer learning was applied, the events were detected by the fine-tuned GPD model. The fixed STA and LTA window lengths prevented the method from detecting these seismic signals. In each panel, the blue and orange lines mark the P- and S-wave picks, respectively, and station names are listed on the right. Only vertical channel records are shown for simplicity.

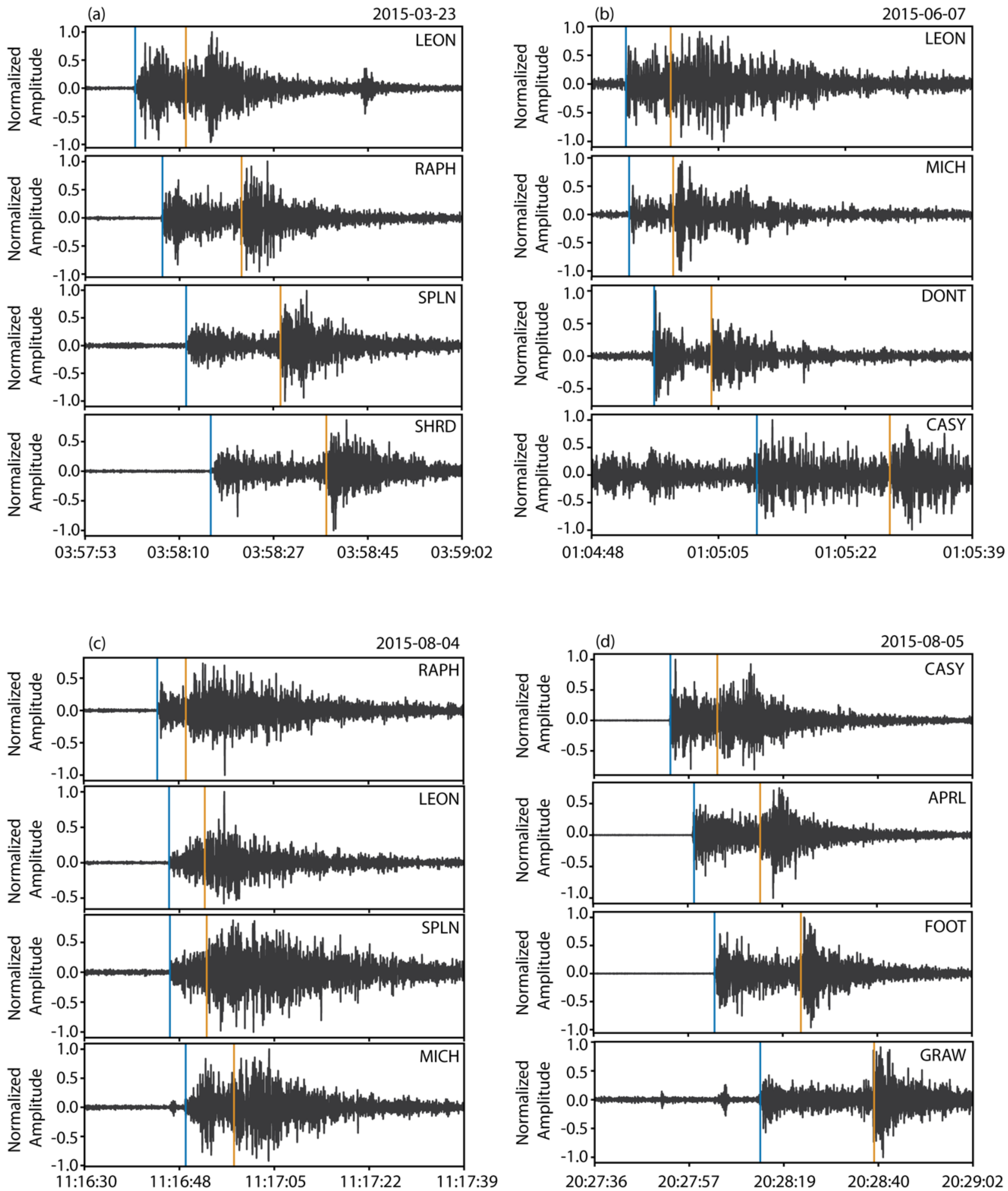

Figure S3. Four example events that were not identified by the MF approach and hence were not included in the initial MF catalog; however, once transfer learning was applied, the events were detected by the fine-tuned GPD model. The events were missed prior to transfer learning because they did not sufficiently correlate with any of the MF template events. Panels are plotted in the same fashion as in Figure S2.

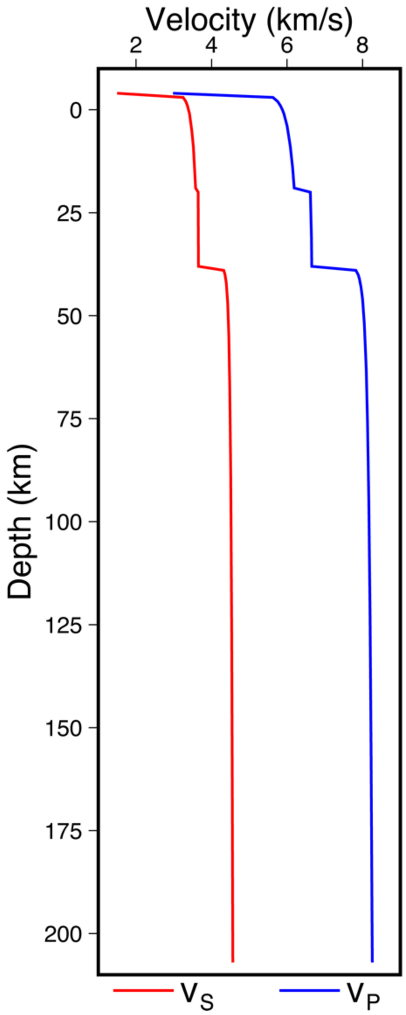

Figure S4. Seismic velocity model used to determine NonLinLoc relocations. S-wave velocity (V_S) is indicated by the red line, and P-wave velocity (V_P) is indicated by the blue line.

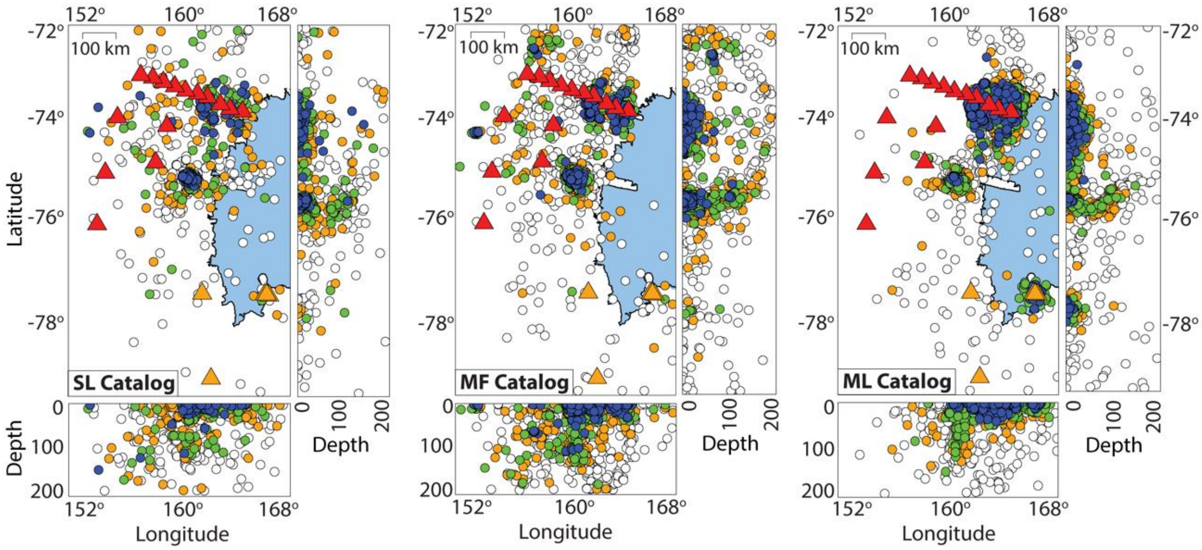

Figure S5. Seismic event relocations from NonLinLoc after the fine-tuned GPD model was applied to each catalog. From left to right: SL catalog, MF catalog, and ML catalog. Circles denote event locations, which are color-coded by their R_e group assignments. Blue: group A; green: group B; orange: group C; no fill: group D. See Table S1 for further details. TAMNNET stations are denoted by red triangles, and other stations are denoted by orange triangles.

Table S1. The number of events (in parentheses) within each Quality Group for each catalog after relocation. Group A and B events are plotted in Figure 8, and all events are plotted in Figure S5.

Catalog	Quality Group
Updated ML Catalog	A (332)
	B (568)
	C (506)
	D (691)
Updated MF Catalog	A (593)
	B (961)
	C (707)
	D (1300)
Updated IN Catalog	A (196)
	B (266)
	C (176)
	D (403)