SAND2022-16608C

Machine learning applications for event detection and phase arrival time estimation of microseismic waveform data at a CO2 injection site

Daniel Lizama, Hongkyu Yoon

Geomechanics Department, Sandia National Laboratories, Albuquerque, NM, USA

PRESENTED AT

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

MOTIVATION & SUMMARY

Energy storage and recovery activities in the subsurface dramatically increased the frequency of induced seismic activity. Understanding the contributing factors and characteristics of induced seismic events are essential in seismic monitoring for the role of risk assessment. For the case of microseismic level activity, event detection and phase picking become increasingly difficult due to low signal to noise ratio and site-specific frequency. Machine learning/deep learning (ML/DL) models have emerged to improve the detection and phase picking of seismic events more accurately and efficiently [1-4].

In this study, we implement two machine learning approaches for detection and phase picking of microseismic activity induced by CO2 injection at the Illinois Basin Decatur Basin (IBDP) [5] using a limited amount of waveform data. We develop a 2D CNN architecture pipeline for optimized detection of true seismic events and use transfer learning (TL) approach to optimize the well-known PhaseNet [2] architecture for phase picking of microseismic waveform data.

Objectives

- Develop convolutional neural network (CNN) models to improve event detection with a small number of training data
- Evaluate the data preprocessing strategy to improve event detection using CNN models
- Implement TL to optimize PhaseNet prediction on microseismic IBDP data.
- Estimate phase arrival onset for accurate event epicenter approximation.

IBDP DATA & PROCESSING

The Illinois Basin Decatur Project (IBDP) site consists of an array of deep borehole passive microseismic monitoring networks continously recording ground motion. We focus on microseismic waveform data over a short time period (Feb. 27 to Mar. 12) when a total of 612 events are located in the catalog. We use three channel time-series data from the lowest geophone.

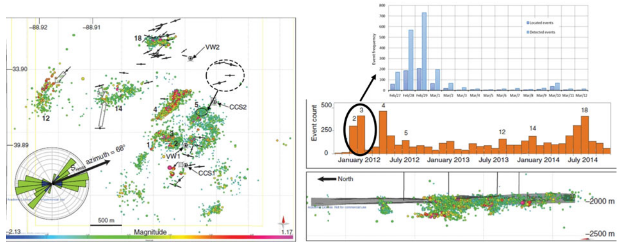

Figure 1. IBDP microseismic event clusters and a number of events over time with the vertical overview of microseismic events [6].

Waveform conversion to spectrogram per channel is as follows:

- Detrend (mean)
- Bandpass (10-400 Hz)
- STFT (win=128)
- Normalized by log scaling [6]
- Rescale (0 to 1)

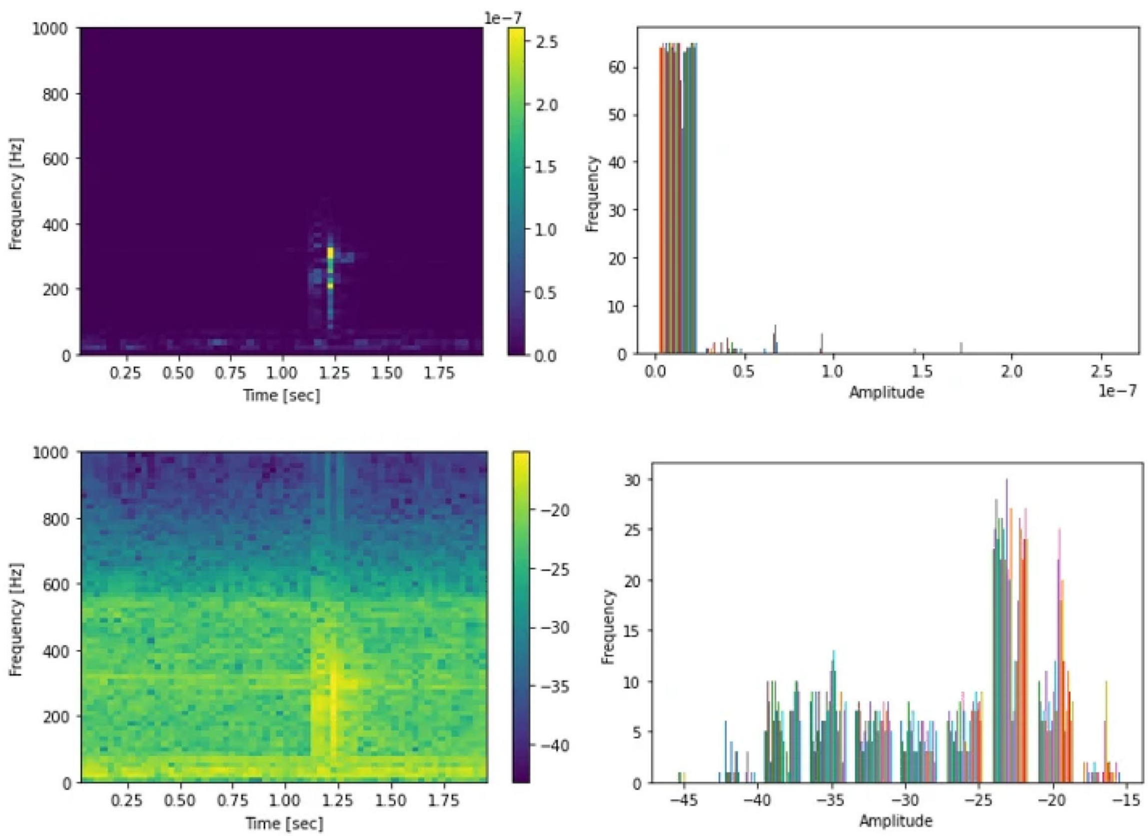

Figure 2. Spectrogram and amplitude histogram (top) before log scaling and (bottom) after log scaling

DETECTION WORKFLOW

Due to the limited number of waveform samples, the models consist of simple CNN architectures with a relatively low number of trainable parameters (~200K) to avoid overfitting. Multilayer perception (MLP) block is used to take MFCC (Mel Frequency Cepstral Coefficient) data as input. Features from both CNN and MLP blocks are concatenated before a final dense layer, leading to the final prediction (event, no event) output.

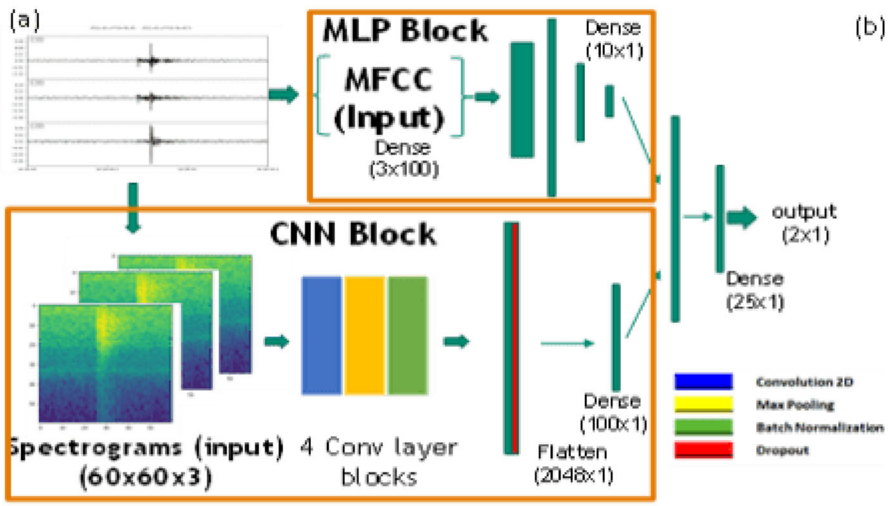

Figure 3. Diagram of DL model architectures and their corresponding input data format.

Key model results:

- Models with proper input normalization optimize better.
- Models predict events registered in the catalog (A-B) and newly detected events (C-D) (Figure 5).

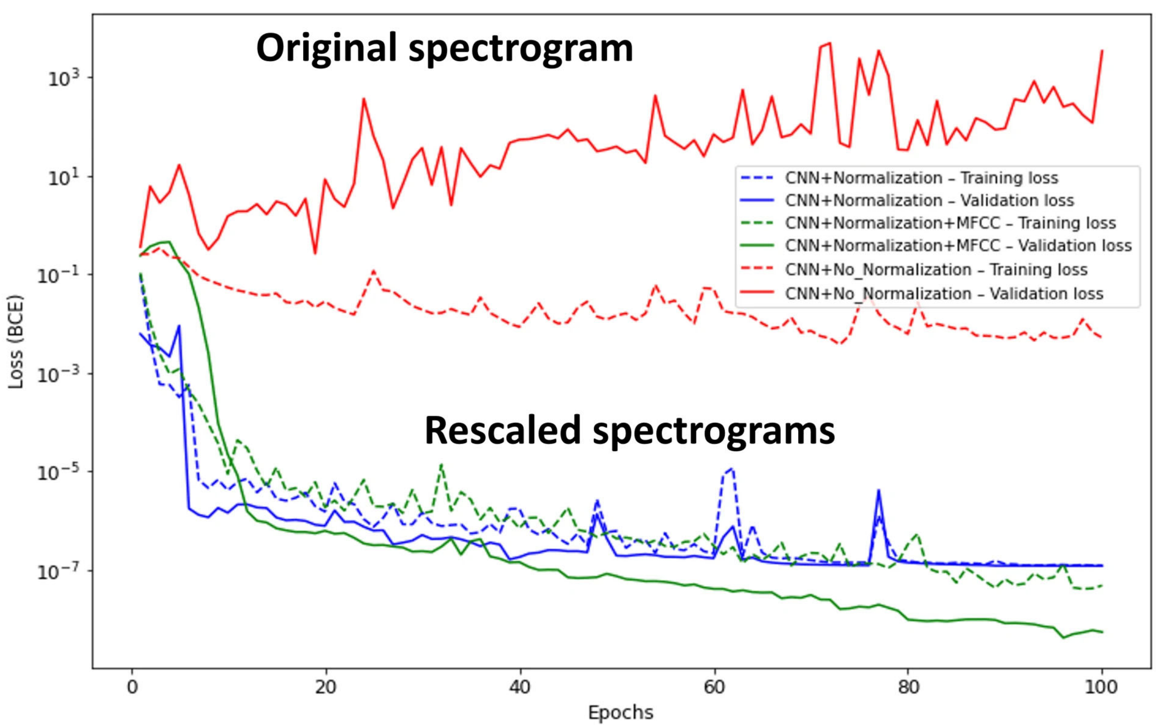

Figure 4. Binary cross-entropy loss values over epochs for each DL model. Dashed and solid curves represent training and validation loss, respectively.

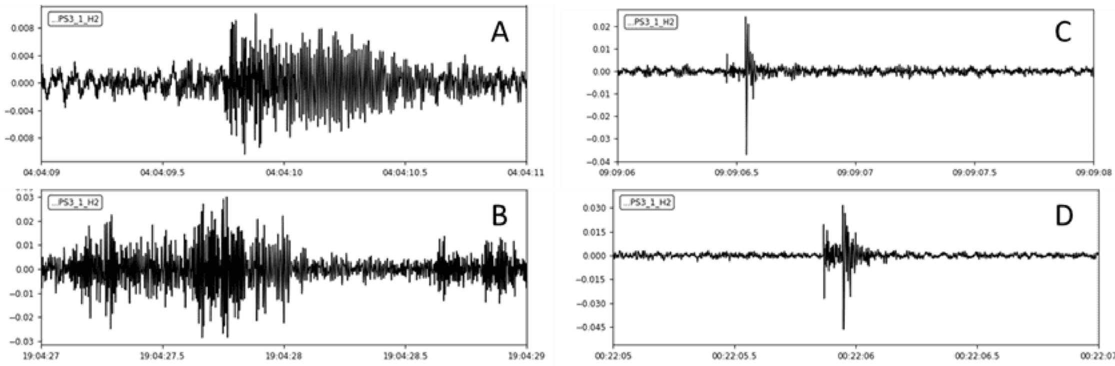

Figure 5. Microseismic detection by DL models on continuous raw waveform data

PHASE ARRIVAL WORKFLOW

We identified events from the catalog with phase arival time labels for training the Phasenet model with transfered weights from the original.

- 419 training samples
- 60 validation samples
- 100 epochs
- Precion results:
  - P-wave: 0.906
  - S-wave: 0.942

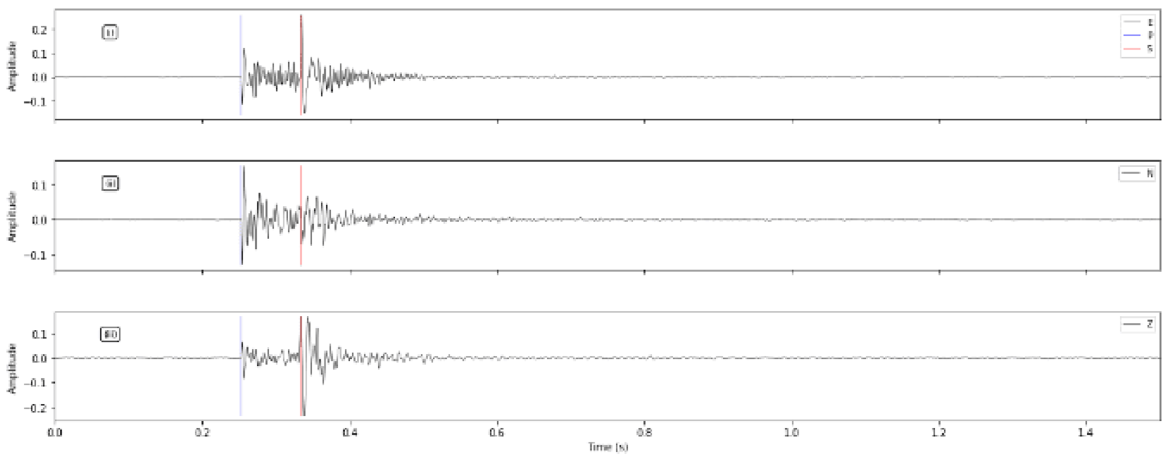

Figure 6. Example of P arrival picks (blue lines) and S arrival picks (red lines) training samples.

Phase picking application:

- Apply prediction window over continuous data within the same period of Feb.27 to Mar 12 (Figure 7).
- Filter to keep prediction windows with both P and S wave probabolity over 90%.
- Figure 8 shows a comparison of picks obtained by PhaseNet against picks obtained by traditional parametric methods (AR-AIC, AICD).

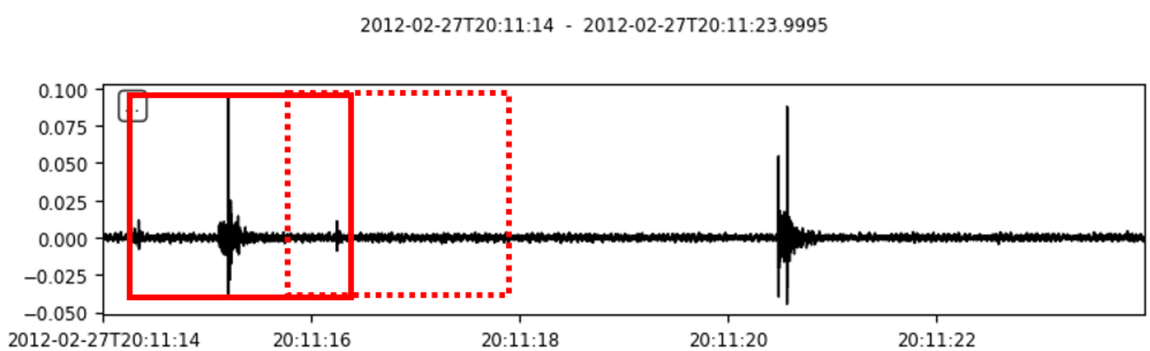

Figure 7. Prediction windowing scheme for continuous recordings. The red square represents the i-th window slice, from the IBDP continuous recordings, to be predicted upon by our DL method.

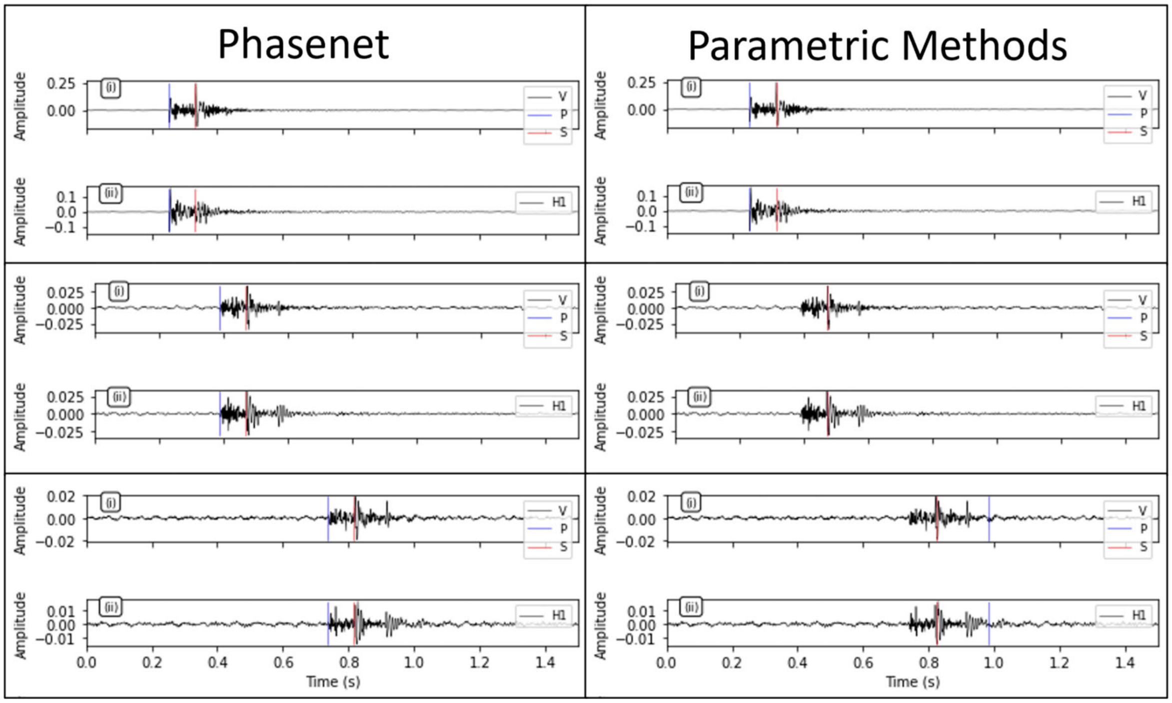

Figure 8. Comparison of PhaseNet picking accuracy over continuous dataset (left) against parameter based automatic pickers (right) applied on the same windows.

RESULTS & FINDINGS

• The DL models detect close to as many or more events than those reported in the IBDP catalog (Figure 9).

• Increase of predicted events suggest there may be more microseismic events of interest.

• The CNN model detects events with characteristics outside of the training distribution, and even outside of its window focal field range as shown in Figure 10 with detected long period and long duration (LPLD) waveforms.

• High confidence phase picks found in new events could enhance current catalog.

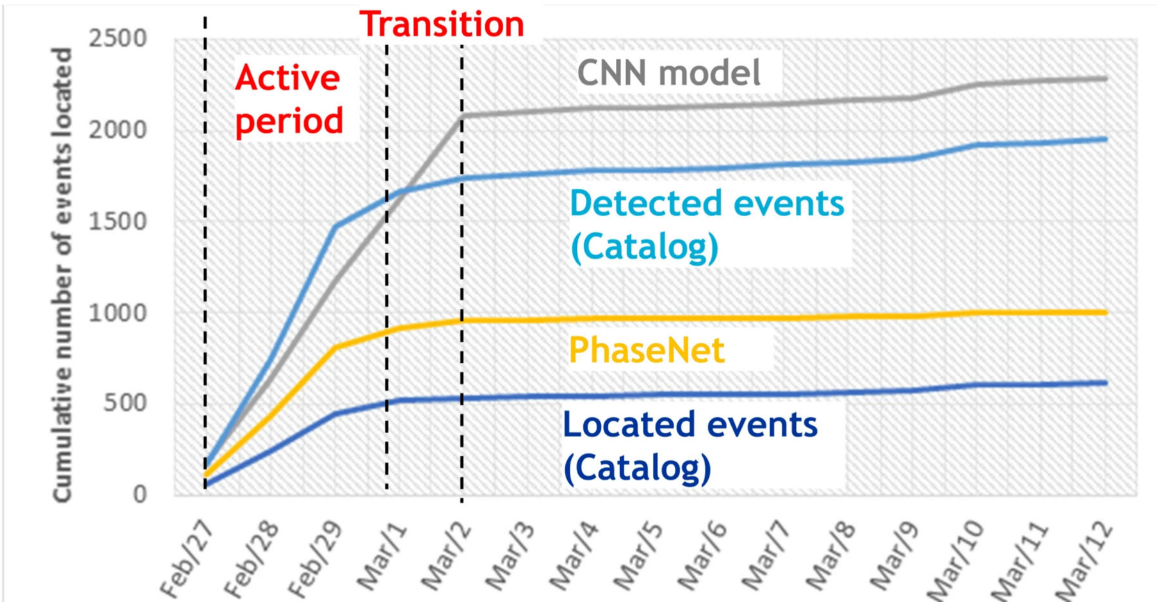

Figure 9. Cumulative number of detected events through the DL models and cataloged events.

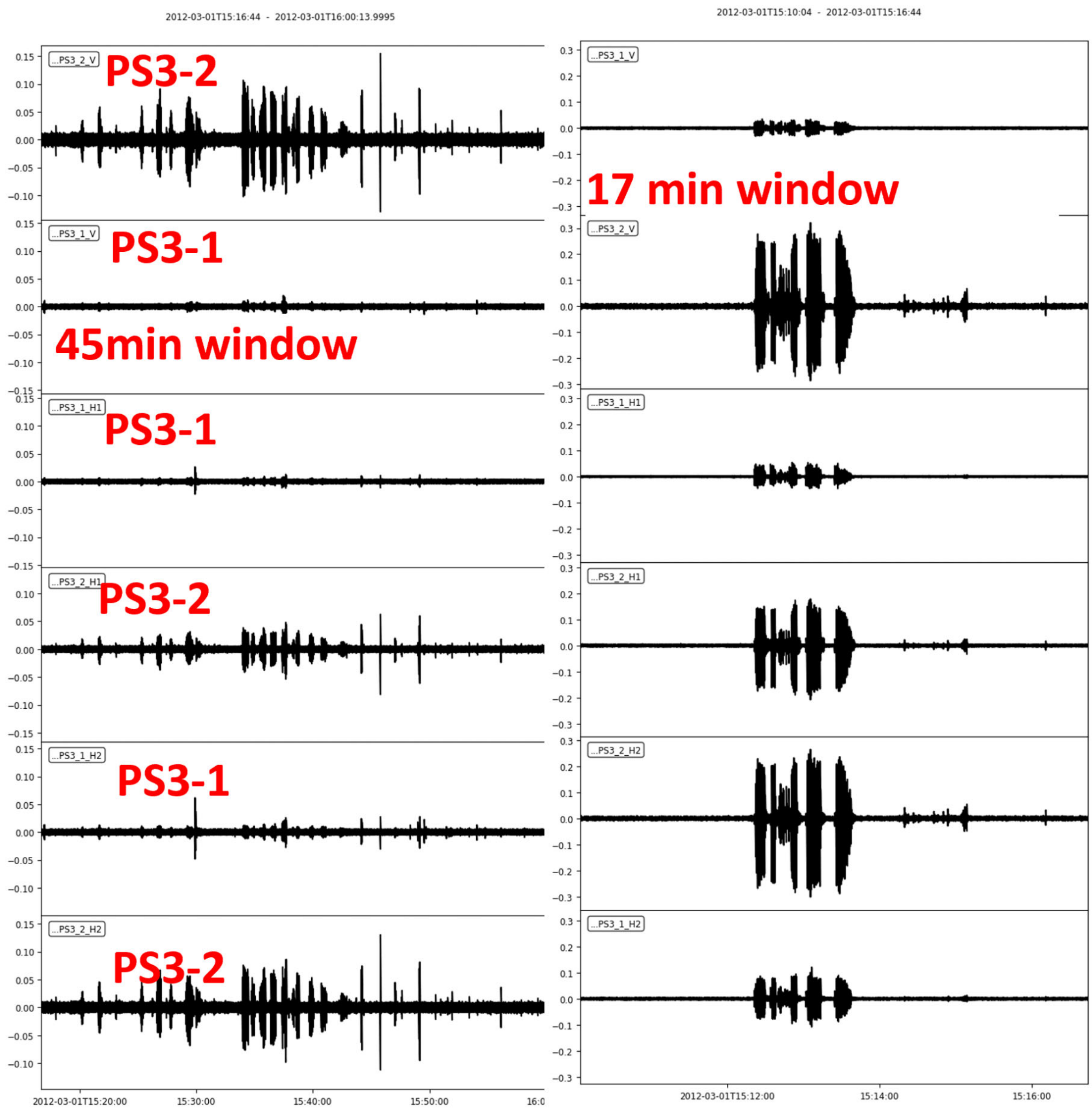

Figure 10. Long period and long duration events detected by trained CNN model (March 1, 2012)

SUMMARY & FUTURE WORK

•The usage of spectrogram and data normalization as a pre-processing improves detection accuracy of CNN models significantly even with 612 event data per cluster.

•Additional physical data in CNN + MLP model seems to improve detection accuracy slightly better.

•Application of trained model for continuous raw microseismic waveform data for Feb-Mar data leads to find LPLD type waveform data, suggesting the fault type is likely different from other faults.

•Significant increase of predicted events suggest there may be more microseismic events of interest unidentified in the IBDP repository.

•We demonstrate the potential of transfer learning for PhaseNet model in the case of limited mircoseismic samples, outperforming conventional phase picking algorithms over continuous unlabeled data.

•Combination of event detection and phase picking DL models will be applied for microseismic event source location analysis.

References

1. Perol et al. (2018). Convolutional neural network for earthquake detection and location. Science Advances, 4(2), p.e1700578.2.

2. Zhu and Beroza (2019). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. Geophysical Journal International, 216(1), pp.261-273.

3. Mousavi et al. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. Nature Communications, 11(1), pp.1-12.

4. Lapins, S., et al. (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): e2021JB021910.

5. Will et al. (2016). Microseismic data acquisition, processing, and event characterization at the Illinois Basin–Decatur Project, International Journal of Greenhouse Gas Control, 54: 404-20.

6.Williams-Stroud et al. (2020). Analysis of Microseismicity and Reactivated Fault Size to Assess the Potential for Felt Events by CO2 Injection in the Illinois Basin, Bulletin of the Seismological Society of America (2020) 110 (5): 2188–2204.

Acknowledgment

This work is supported by the US Department of Energy Office of Fossil Energy and Carbon Management project and Science-Informed Machine Learning to Accelerate Real Time Decisions in Subsurface Applications-Carbon Storage (SMART-CS) initiative.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525.

ABSTRACT

Over the past decade energy storage and recovery activities in the subsurface dramatically increased the frequency of induced seismic activity and high magnitude events increased public concerns related to safety. Understanding the contributing factors and characteristics of induced seismic events are essential components in seismic monitoring for the role of risk assessment. For the case of microseismic level activity, event detection and phase picking become increasingly difficult due to low signal to noise ratio and site-specific frequency which renders typical earthquake monitoring algorithms ineffective. In this study, we implement two machine learning approaches for detection and phase picking of microseismic activity induced by CO2 injection at the Illinois Basin Decatur Basin (IBDP) using a limited amount of waveform data. We develop a 2D CNN architecture pipeline, which analyzes the time-frequency images of waveform for optimized detection of true seismic events and use transfer learning approach to optimize the well-known PhaseNet architecture for phase picking of microseismic waveform data. Data preprocessing and normalization approach dramatically improve CNN model robustness to better adapt to the data of interest and even generalize detections for seismic activity outside of the training data distribution. We also demonstrate the potential of transfer learning by training the PhaseNet model in the case of limited mircoseismic samples, outperforming conventional phase picking algorithms over continuous unlabeled data.

SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.