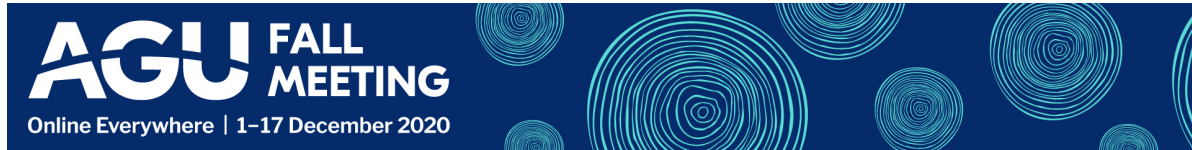


Supervised Machine Learning Application for Event Detection in Induced Microseismic Data

Daniel Lizama(1), Hongkyu Yoon(1)

(1) Geomechanics Department, Sandia National Laboratories, Albuquerque, NM, USA

PRESENTED AT:



MOTIVATION & OBJECTIVES

Motivations

Understanding the contributing factors for induced seismic events caused by subsurface energy activities is essential to mitigate seismic hazard and improve subsurface energy activities. The evolution of induced seismic events in United States and around the world, and the exponential increase of seismic data call for efficient algorithms that can characterize/detect/locate these events in the records and in real time. In this study we use convolutional neural networks (CNNs) to detect and localize microseismic activity induced by CO₂ injection at the Illinois Basin Decatur Basin (IBDP) site.

Objectives

- Analyze IBDP microseismic waveform data and develop an effective data preprocessing methodology.
- Use CNN model to increase detection rate of microseismic activity, compared to human based picks.
- Evaluate training strategies to utilize microseismic data detected at different locations at the IBDP site.

Approach

We compare a 2D CNN architecture, which utilizes the spectrogram of the waveform as training data, over multiple training strategies (pre-trained, re-trained, combined dataset) to analyze its microseismic activity detection ratio performance.

MICROSEISMIC DATA

The Illinois Basin Decatur Project (IBDP) is a passive microseismic monitoring network system dedicated to recordings of noise and coherent signals emitted by discrete subsurface displacement events, including CO₂ subsurface injection [1]. Data streams consist of 10 second long continuous files saved sequentially (8640 files per day) composed of simultaneous recordings of multiple geophones placed along the well, as shown in figure 1 (copied from Will et al., 2016). All sensors recorded at 2KHz sampling rate.

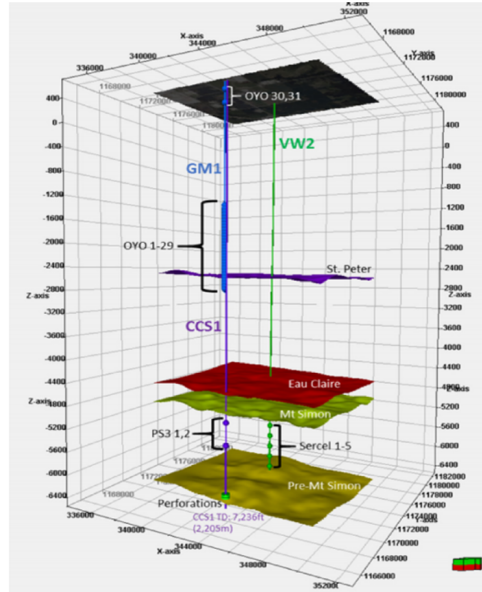


Fig. 1. Subsurface array configuration. Distance units are feet, Z axis is referenced to mean sea level.

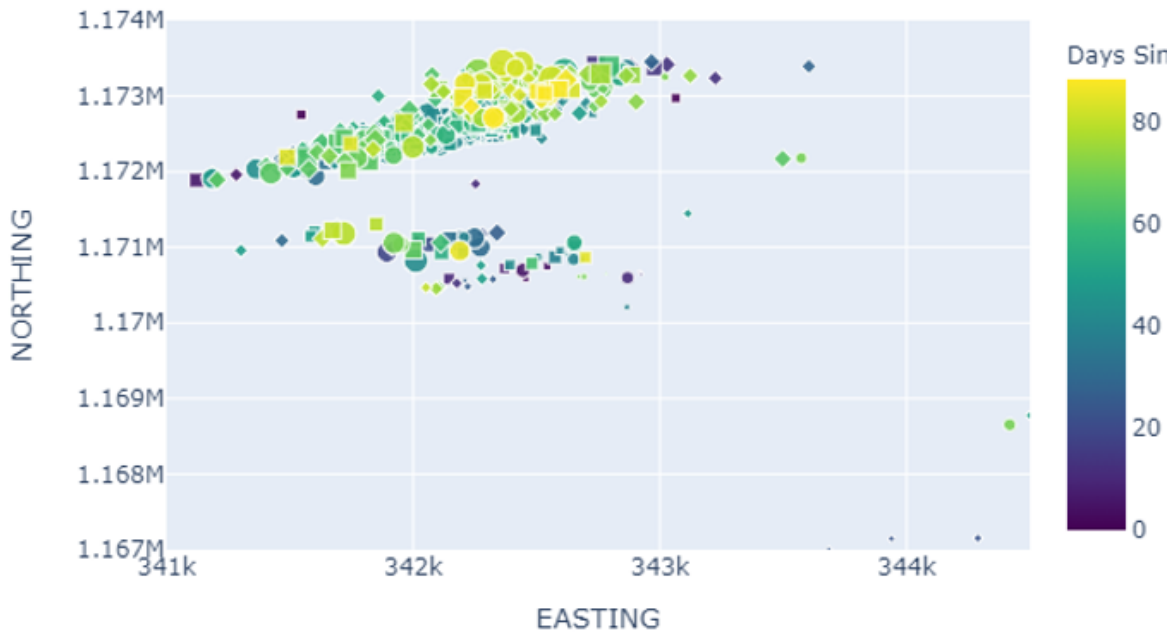


Figure 2 microseismic event source location, time evolution and magnitude (represented by the symbol size).

Certain seismic events are recorded into a catalog along with their source location (see figure 2) and other parameters which could help explain the physiological differences in seismic event waveforms. These known seismic event picks serve as our ground truth and training data for our CNN models.

A 3 channel waveform data from two closest geophones (PS3) to an injection interval are used to construct the labelled event data. A short time fourier transform (STFT) analysis applied to multiple samples seismic event recordings (single channel), as shown in figure 3, shows that the range of frequency for actual seismic event is ~10 to 400 Hz.

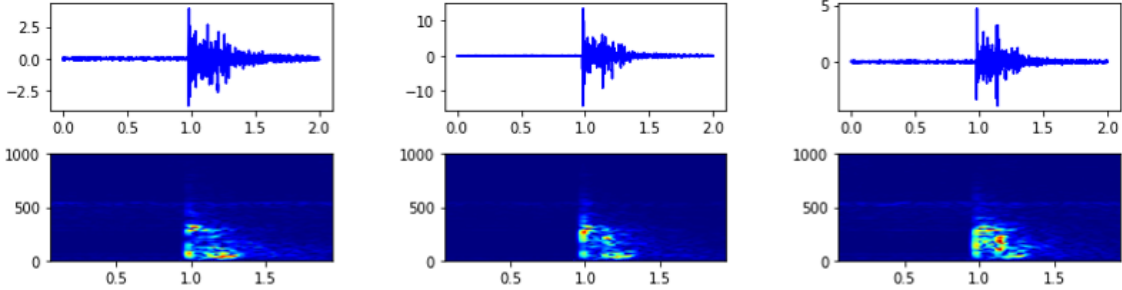


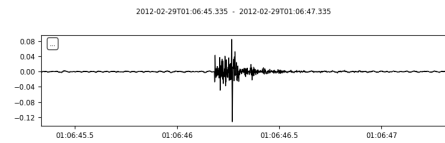
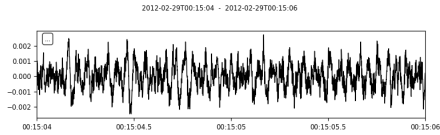
Figure 3- Microseismic event picks raw signal in channel 1 (top), with corresponding spectrogram (bottom)

PRE-PROCESSING & CONVOLUTIONAL NEURAL NETWORKS

Based on the analysis findings, a methodology is developed to effectively incorporate CNN models for microseismic activity detection. Pre-processing is applied to the raw time series data prior to converting each channel to time-frequency (spectrogram) input for the CNN.

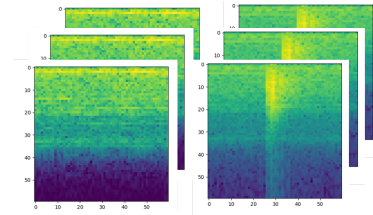
A channel wise temporal filter and detrending is implemented to minimize the contribution of coherent noise and remove other possible distorting variables occurring in the raw data time series. Specifically, we apply:

- Temporal filter: Butterworth bandpass 10-400 Hz
- Signal detrend: subtract signal mean
- Two seconds time window segments of events and noise



Time series dataset are transformed to spectrograms such that their temporal information is augmented to time-frequency information. This is then normalized applying log scaling and remapping to scale the data range Spectrogram methodology:

- $60 \times 60 \times 3$ matrix: window size=128; overlap=1/2
- Log scaling: $\log(\text{spectrogram} + 1e-20)$
- remap scale: remap values range (0 to 1)



Convolutional Neural Networks have proven effective at detecting earthquake related seismic activity [2]. In this study we make use of simple CNN structure, shown in the figure, which take time-frequency inputs and outputs a class prediction (event or noise) of the seismic sample recording. We also want to evaluate the effects of training dataset selection on the prediction performance of the CNN. CNN training strategies:

- Model 1: trained model with dataset from Feb 27 to Mar 12 (684 events and >15,000 noise data)
- Model 2: retrain Model 1 with Jun 24 to Jun 27 dataset
- Model 3: trained model with dataset Feb 27- Mar12 and Jun 24- Jun27
- Training/Evaluating/Testing = 70/20/10 (%)
- PS3 sensor 1 and 2 as training sets

RESULTS & DISCUSSION

Motivations

Overall, Model 1 has consistently more prediction of events over both Feb-March and June periods, suggesting its training dataset accurately captures the features of microseismic data.

Table. Event detection results with three different models

Date	Events Located	Events Model 1	Events Model 2	Events Model 3
Feb 27-29	451	1341	773	965
Mar 1-2	85	1102	1711	844
Mar 3-6	17	67	43	52
Mar 7-9	16	45	31	39
Mar 10-12	43	109	66	83
Jun 24-25	235	429	260	395
Jun 26-27	130	305	283	279

Model 2 (being retrained with June data) had its most significant prediction increase over Mar 1-2 and Jun 26-27 periods.

Model 3 had overall reduction in event predictions compared to Model 1. This suggests there is significant difference in waveform characteristics between Feb-Mar and June datasets given that Models 2 and 3 are not able to generalize as much as Model 1.

From this we can infer that training data quality is more essential to our CNN model than case sample quantity to allow proper feature extraction for general microseismic event detection.

In addition, based on the increased detection of Model 2 over Mar 1-2 and Jun 26-27, it's highly suggested that these two time periods have more common waveform patterns towards which this model adapted to. This could be caused by mechanical changes or failure like increase pore pressure, change in local stress field, fracture opening or closing, or shear slip.

Input_Layer (None,60,60,3)	
Conv2D	(None, 60, 60, 16)
Max Pooling2D	(None, 30, 30, 16)
Batch Normalization	(None, 30, 30, 16)
Conv2D	(None, 30, 30, 32)
Max Pooling2D	(None, 15, 15, 32)
Batch Normalization	(None, 15, 15, 32)
Conv2D	(None, 15, 15, 64)
Max Pooling2D	(None, 8, 8, 64)
Batch Normalization	(None, 8, 8, 64)
Conv2D	(None, 8, 8, 12)
Max Pooling2D	(None, 4, 4, 128)
Batch Normalization	(None, 4, 4, 128)
Flatten	(None, 2048)
Dropout	(None, 2048)
Dense	(None, 100)
Output_Layer	(None, 2)

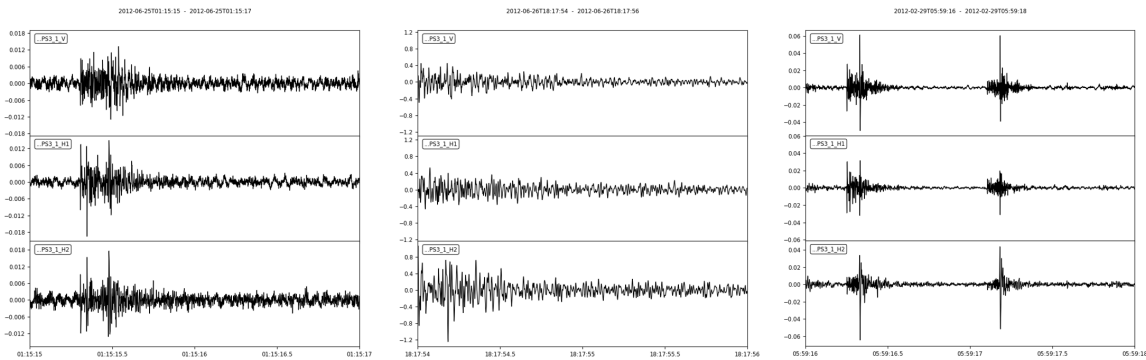


Figure. Examples of newly detected events, varying in SNR.

CONCLUSION/FUTURE WORK/REFERENCES

Conclusions

- Analysis of the seismic time series at hand allowed the development of an effective preprocessing strategy for the CNN to achieve enhanced detection capabilities.
- Overall, CNN based methods achieved higher seismic event detection rate than human based picks, based on informed learning from actual microseismic events.
- Evaluation of training strategies proves that quality and balance of training sample cases allow fundamental event related feature extraction by the CNN, which in terms translates to better generalization and detections.

Future Work

Our findings contribute towards the efforts of effectively automate seismic event detection in real time. Other capabilities could be added to our framework which would allow it to quantify and/or validate (1) event data quality, (2) event data characterization, and (3) event source location in real time.

References

- [1] Will, R., El-Kaseeh, G., Jaques, P., Carney, M., Greenberg, S. and Finley, R., 2016. Microseismic data acquisition, processing, and event characterization at the Illinois Basin–Decatur Project. *International Journal of Greenhouse Gas Control*, 54, pp.404-420.
- [2] T. Perol, M. Gharbi, M. Denolle, Convolutional neural network for earthquake detection and location. *Sci. Adv.* 4, e1700578 (2018).

Acknowledgement: 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.

ABSTRACT

Quantifying in-situ subsurface conditions and predicting fracture development are critical to reducing risks of induced seismicity and improving modern energy activities in the subsurface. In this work, we developed a novel integration of both supervised and unsupervised machine learning methods to identify mechanical failure processes observed under controlled laboratory experiments and estimate reservoir response of fault (re)activation from induced seismic data observed at a field demonstration site of CO₂ injection. For laboratory experiments, we integrated acoustic emission (AE) data collected during three-point bending testing of 3D printed and natural samples, digital image correlation (DIC) and micro-CT data. In addition, synthetic wave data generated from numerical simulations under different failure mechanisms is augmented to constrain classification of the relevant waves related to rock deformation. The methods tested with laboratory data have been applied for microseismic data obtained at Illinois Basin Decatur Project (IBDP) to characterize the different patterns observed in the Precambrian basement rock formation and reservoir formation where CO₂ was injected into the reservoir formation. An unsupervised machine learning has been developed as a fingerprint-based clustering approach with nonnegative matrix factorization and hidden Markov model. The set of clusters resulting from the fingerprint-based approach will be also used to train a supervised approach such as convolutional neural networks to detect the new seismic events that can be used to improve the (hidden) fault identification. This research will improve characterizing seismic waveforms by machine learning approaches and the detection of low-magnitude seismic events leading to the discovery of hidden fault/fracture systems. SNL is managed and operated by NTESS under DOE NNSA contract DE-NA0003525.