GTC/GT/IDA FORGE Project 6-3712 Report for Milestone 1.3

Probabilistic Estimation of Seismic Response Using Physics Informed Recurrent Neural Networks

Jesse Williams, Global Technology Connection, Inc., GA
Zhigang Peng, Georgia Institute of Technology, GA
Sheng Dai, Georgia Institute of Technology, GA
Wencheng Jin, Idaho National Laboratory, ID

1. Introduction

In FORGE project 6-3721, Global Technology Connection, Inc. (GTC), in collaboration with the Georgia Institute of Technology (GT) and Idaho National Laboratory (INL), proposes to develop a predictive tool to estimate the future magnitude of induced seismicity using artificial intelligence (AI) and physics-based modeling. This tool will use the Utah FORGE existing seismic data, injection parameters, and geophysical model predictions to make probabilistic estimations, i.e., a distribution of estimations representing the likelihood of different possible outcomes. Before proceeding with prediction, it is essential to clearly define the model's inputs and outputs. While injection parameters can be directly used as inputs, it is also necessary to determine a method for calculating seismic responses from seismic waveforms to serve as outputs. The following sections detail the progress made in obtaining time-series data with identified events, providing a foundation for constructing these predictive models.

2. Magnitude Frequency Distribution (MFD) Statistics (Task 1.3)

The temporal seismicity magnitude-frequency distribution (MFD) statistics, which includes the a-value, b-value, and magnitude of completeness (Mc), are crucial for understanding seismicity rates (a-value) and can also serve as indicators of subsurface stress conditions (b-value) (Stein and Wysession, 2003). Hence, these are essential parameters in seismic hazard analysis (Reiter, 1990). Studies have shown that integrating temporal b-values, Mc, and the volume of injected fluids can be used in estimating the Maximum Magnitudes (M_{max}) (Van der Elst, 2016; Li et al., 2022), an important parameter that is forecasted during injection.

To estimate the MFD statistics for Utah Forge, we start by using the catalog for detected events during the stimulation experiments conducted in April 2022 (Niemz et al., 2024). We first obtained the Mc for the entire three stages of stimulation using the Maximum Curvature Method (Wiemer and Wyss, 2000), M_c by b-value stability (Cao and Gao, 2002), and the M_c Lilliefors (Herrmann and Marzocchi, 2020). These analyses yield magnitudes of completeness of –1.30, -0.35, and –0.65, respectively. The M_c results from the Lilliefors methods are consistent with the result from Niemz et al., 2024 for the circulation stage. Subsequently, using the maximum likelihood method (Aki, 1965; Bender, 1983; Utsu, 1999), computed temporal b-values within a moving window of 150 data points (Figure 1). The results indicate b-values predominantly ranging between 1.0 and 2.5, which are typical of geothermal systems. In addition, we observe a gradual increase and decrease in the b-value, which correlates with the time of the stimulation stages. We considered using an equal moving time window for b-value computations to standardize time intervals, as using a fixed number of data points results in varying time spans and potential inconsistencies.

However, this approach was constrained by several time windows lacking a sufficient number of events to compute the parameter reliably.

Next, we will conduct a thorough detection on the continuous data to identify additional events within and outside the three stages. This will enable the use of a regular time step for b-value calculations and other MFD statistical parameters.

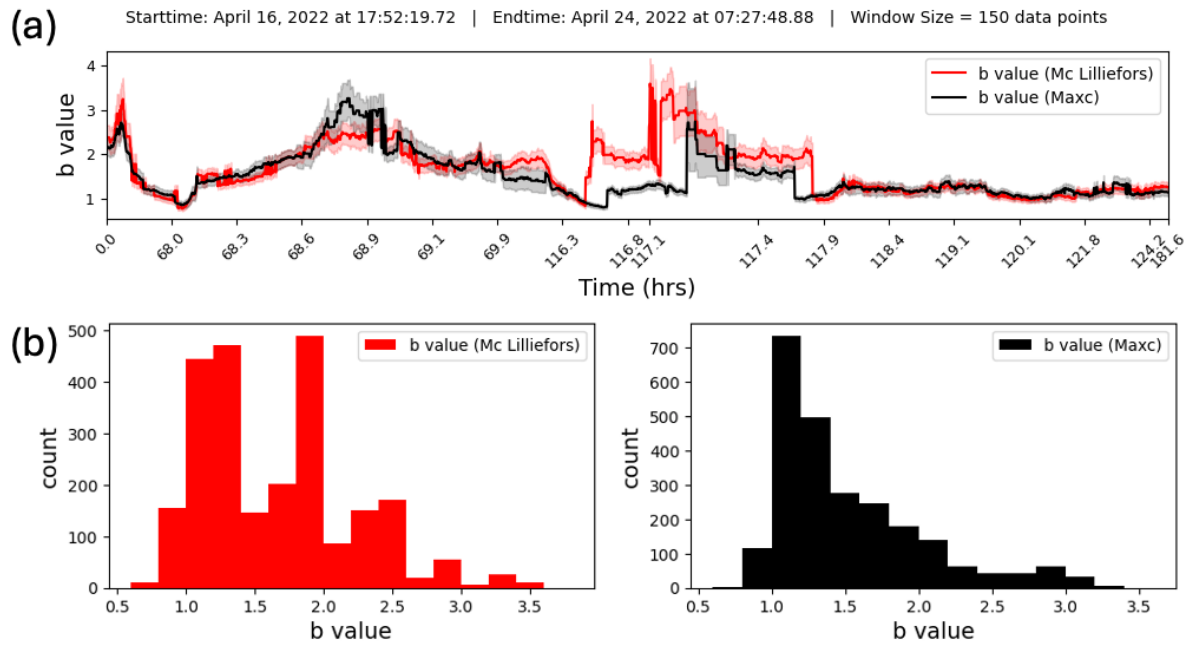

Figure 1: (a) Temporal b-value for the 2022 stimulation based on the catalog from Niemz et al., 2024. Red and black lines are the b-values computed using M_c Lilliefors and the maximum likelihood method, respectively. (b) Summary of the computed b-values using the different methods.

3. Detected Events from continuous data. (Task 1.3)

During the data processing stage, it is necessary to transform seismic waveforms into Magnitude-Frequency Distributions (MFD). Achieving this requires access to a high-resolution seismic catalog. However, due to limitations in the available data, we focused on the 2022 stimulation data to demonstrate the workflow. In previous research (Li et al., 2023), corresponding labels for MFD construction were typically defined as the number of events or the maximum magnitude within a given time window. Unfortunately, the 2022 stimulation data only span one week, encompassing just a few thousand earthquakes across three stimulation stages. This restricted time frame introduces challenges, such as a limited dataset and difficulties in appropriately setting the time window for analysis. Moreover, the currently available catalog is low-resolution, leading to inconsistencies and a lack of smoothness in the corresponding labels required for MFD computation.

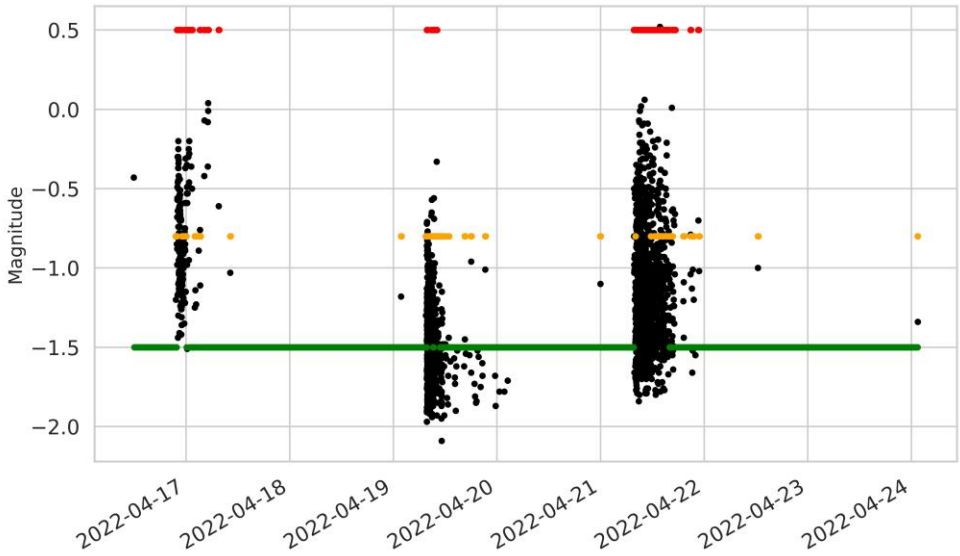

Figure 2. Earthquake catalog (Niemz et al., 2024) of 2022 stimulation data. Black point are the as-reported events. Green, yellow, and red points are categorical severity based on the team's criteria and will be used as targets for later model training.

To address these challenges, we developed a workflow to directly detect potential seismic events from continuous waveforms and generate a high-resolution catalog, which can subsequently be used to compute the corresponding Magnitude-Frequency Distribution. The workflow begins with applying a bandpass filter of 100 to 1800Hz. This step eliminates low-frequency noise from machinery at the site, as the induced microseismic events from the forge are expected to be pulse-like with high-frequency components (Sobot et al., 2024). Thereafter, we obtained the waveform envelope for the continuous waveforms from the individual stations by applying the Hilbert transform. Finally, we stacked the envelopes from the different stations by performing cross-correlation and shifting on the maximum amplitude (Figure 3). This step enhances the signal's representation, focusing on the dominant frequencies relevant to the detection of seismic events. Next, we apply a peak detection algorithm (Marcos, 2021) to identify potential peaks from the calculated envelope. When comparing the detected peaks with the events listed in the existing low-resolution catalog (Figure 3), the results demonstrate that this method successfully detects the cataloged events and identifies additional potential seismic events that are not included in the catalog. This expanded detection capability significantly increases the number of events available for analysis, enabling the computation of a more comprehensive and reliable MFD.

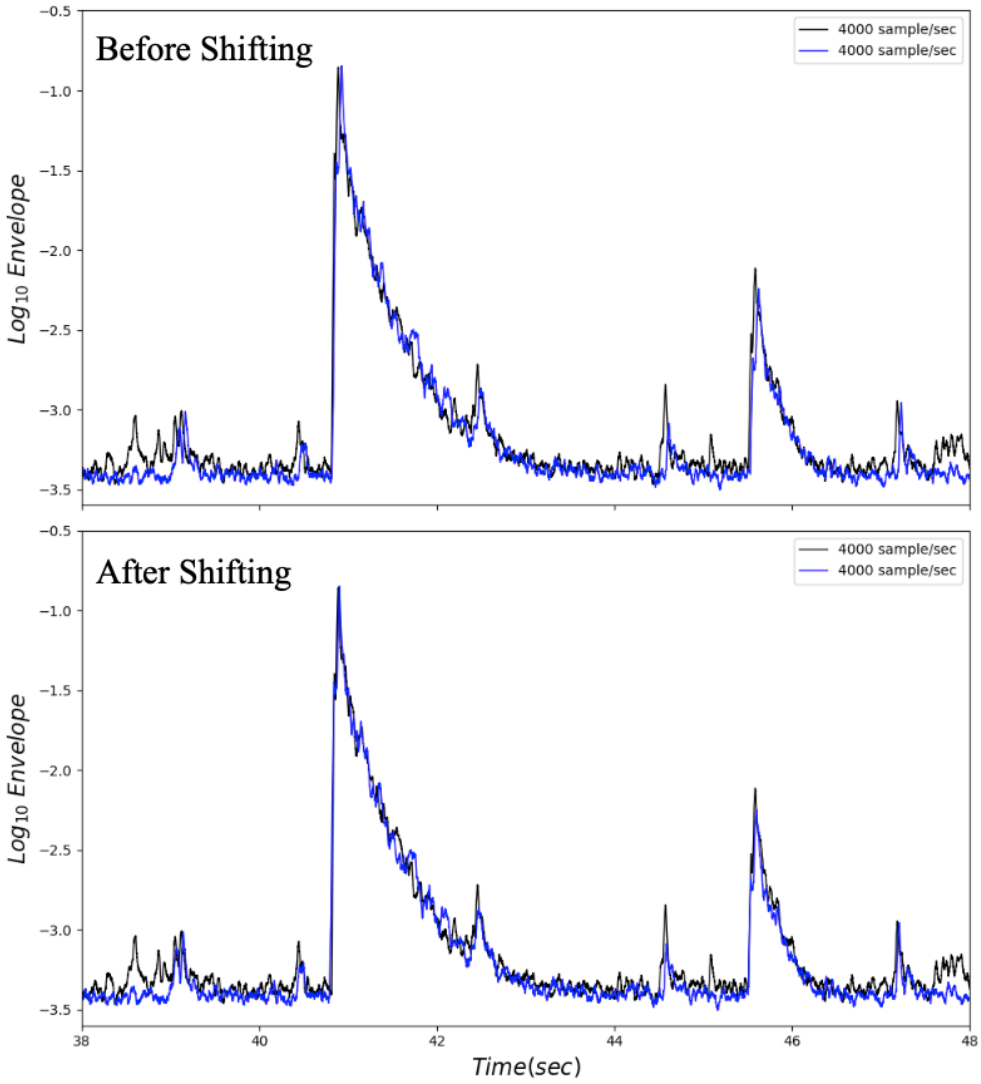

Figure 3: Envelope function before and after cross-correlation and shifting.

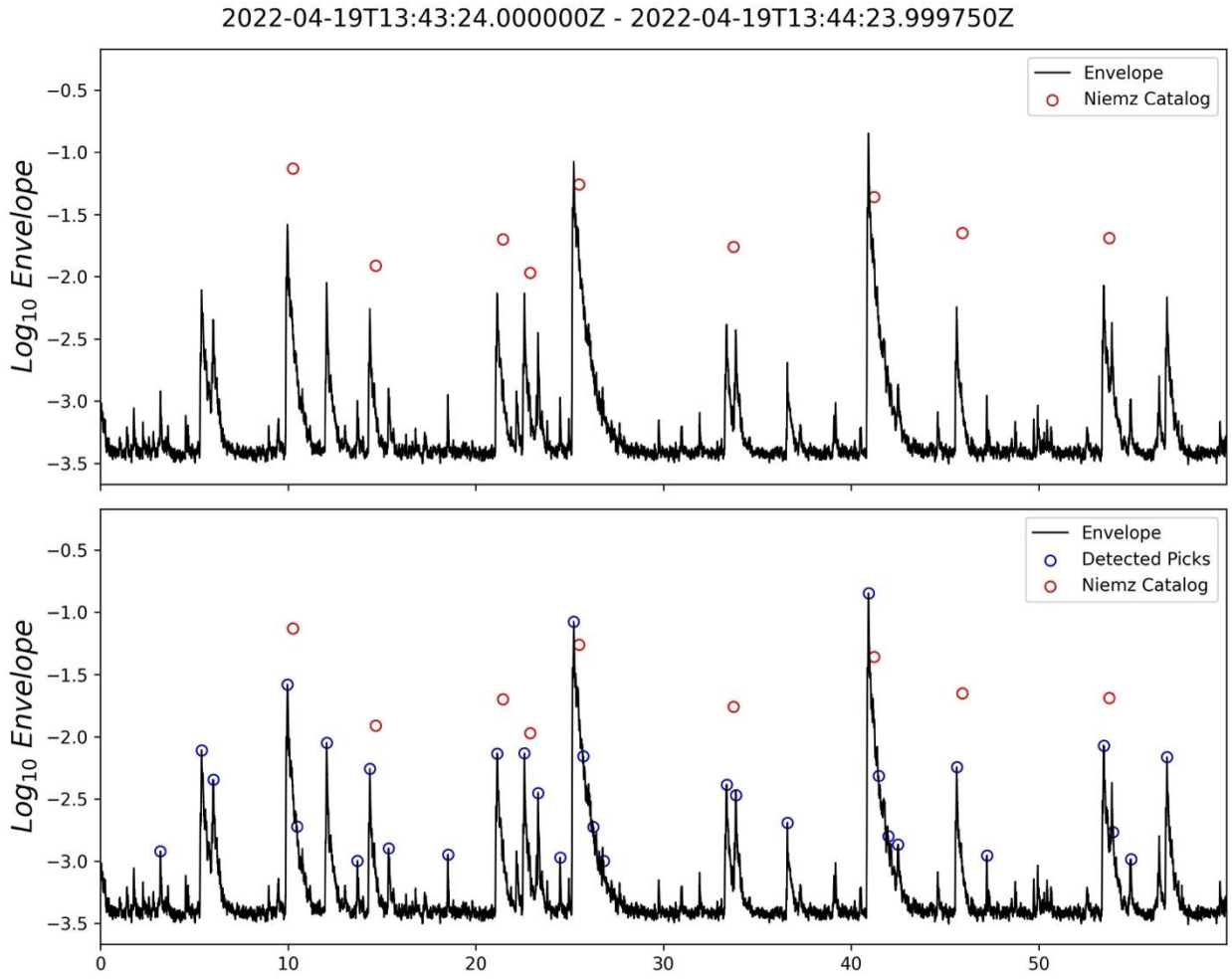

Figure 4: The comparison of detection method results and earthquake catalog.

However, despite this improvement, challenges remain during the processing. Specifically, a gap exists between the potential events detected from the envelope and the events listed in the earthquake catalog. This discrepancy arises because the detected potential events are based on envelope amplitude, which represents a measure of signal strength, while MFD calculations require earthquake magnitudes. To bridge this gap, it is crucial to establish a robust relationship between the envelope amplitude and the earthquake magnitude. This step involves accurately calibrating the amplitude values to approximate magnitudes, ensuring that the resulting MFD reflects the true distribution of earthquake magnitudes. To achieve this, we focused on events that were detected by both our workflow and the existing earthquake catalog. These overlapping events serve as a reliable basis for calibration, as they are consistently represented in both datasets. By leveraging this overlap, we can establish a functional relationship between envelope amplitude and earthquake magnitude.

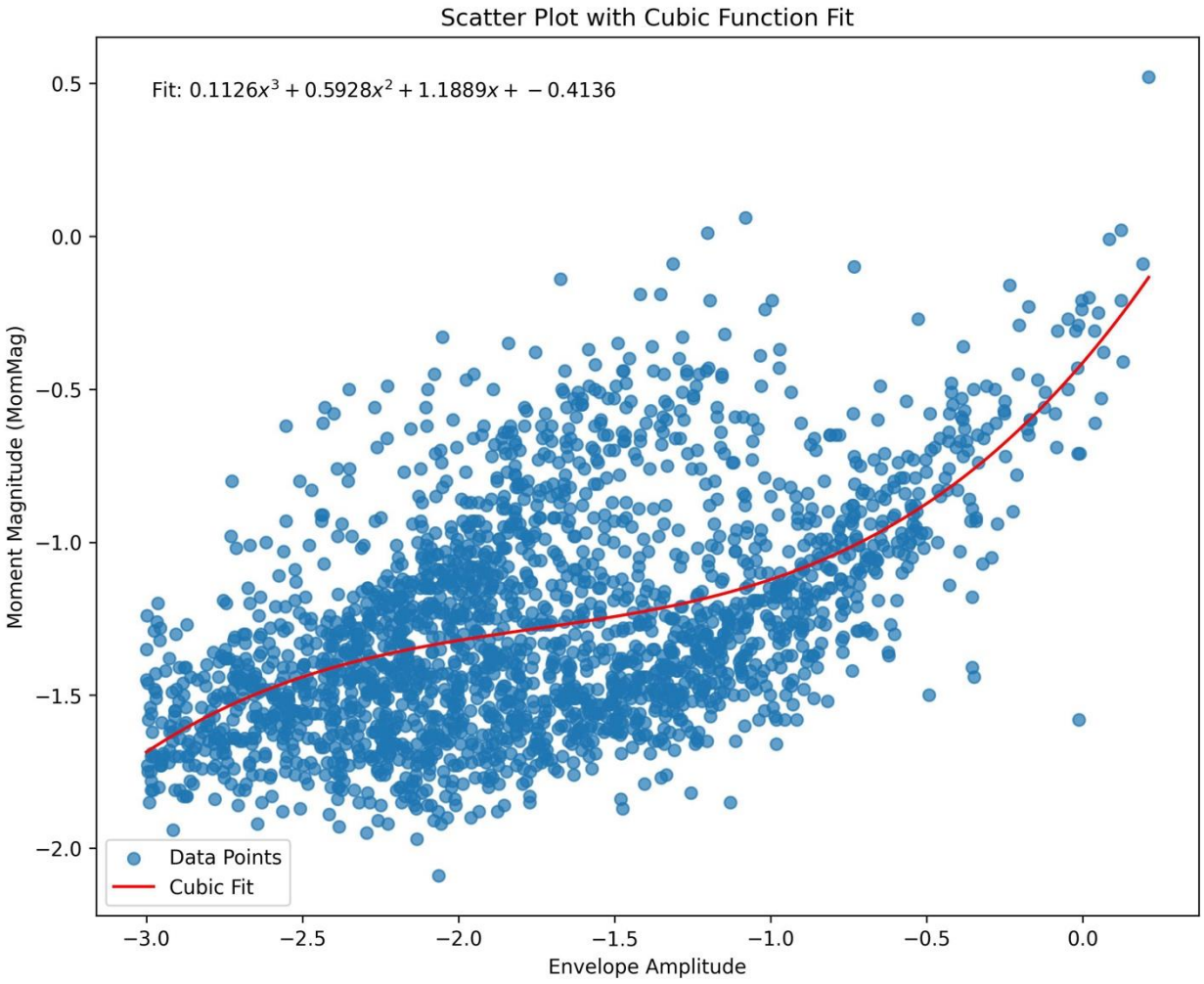

Figure 5: The relationship between the events detected by our workflow and those listed in the earthquake catalog.

After exploring various fitting approaches, we selected a cubic function to model the relationship between the envelope amplitudes and the corresponding earthquake magnitudes from the two catalogs. This decision was based on its superior performance compared to other fitting methods in capturing the nuances of the data. As illustrated in Figure 5, the cubic function effectively aligns the envelope amplitudes with the instantaneous magnitudes of the overlapping events, providing a robust calibration model. This calibration establishes a direct relationship between the envelope amplitude and the moment magnitude. With this relationship in place, we can now estimate the moment magnitudes for other events detected exclusively by our workflow. These calculated magnitudes allow us to construct a complete MFD for the three stimulation stages in 2022 (Figure 5).

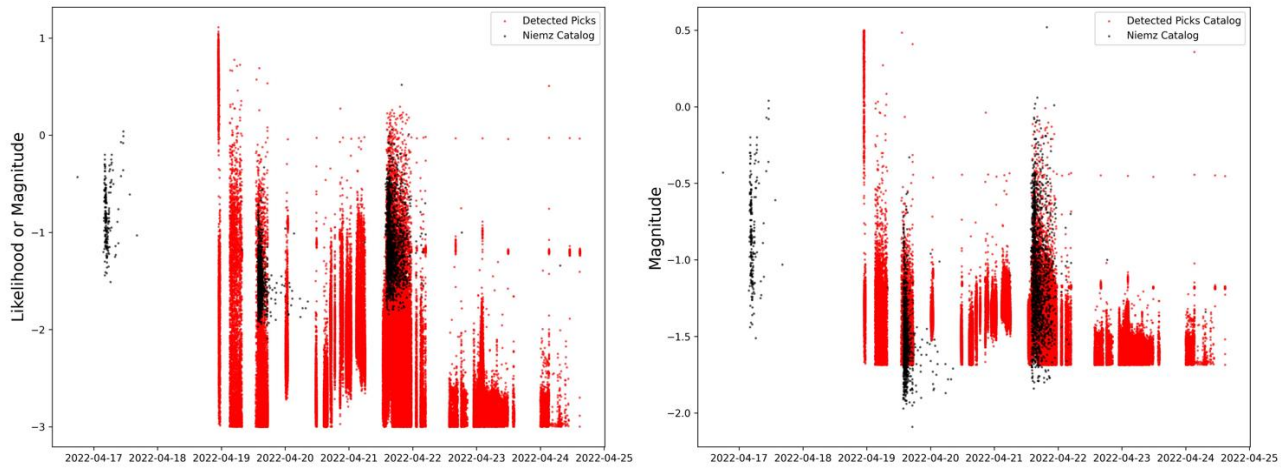

Figure 6: The potential events distribution before and after the transform.

The MFD not only provides insight into the seismic activity during these stages but also enables the generation of key metrics, such as the maximum magnitude and the number of events within specified time windows. As demonstrated in previous research (Li et al., 2023), these metrics are critical for understanding seismic responses and can be used as labels for machine learning models. By setting these derived metrics as labels, we can construct a comprehensive dataset to train models for predicting traffic light states or MFD distributions. Ultimately, this workflow bridges the gap between raw seismic data and actionable predictions.

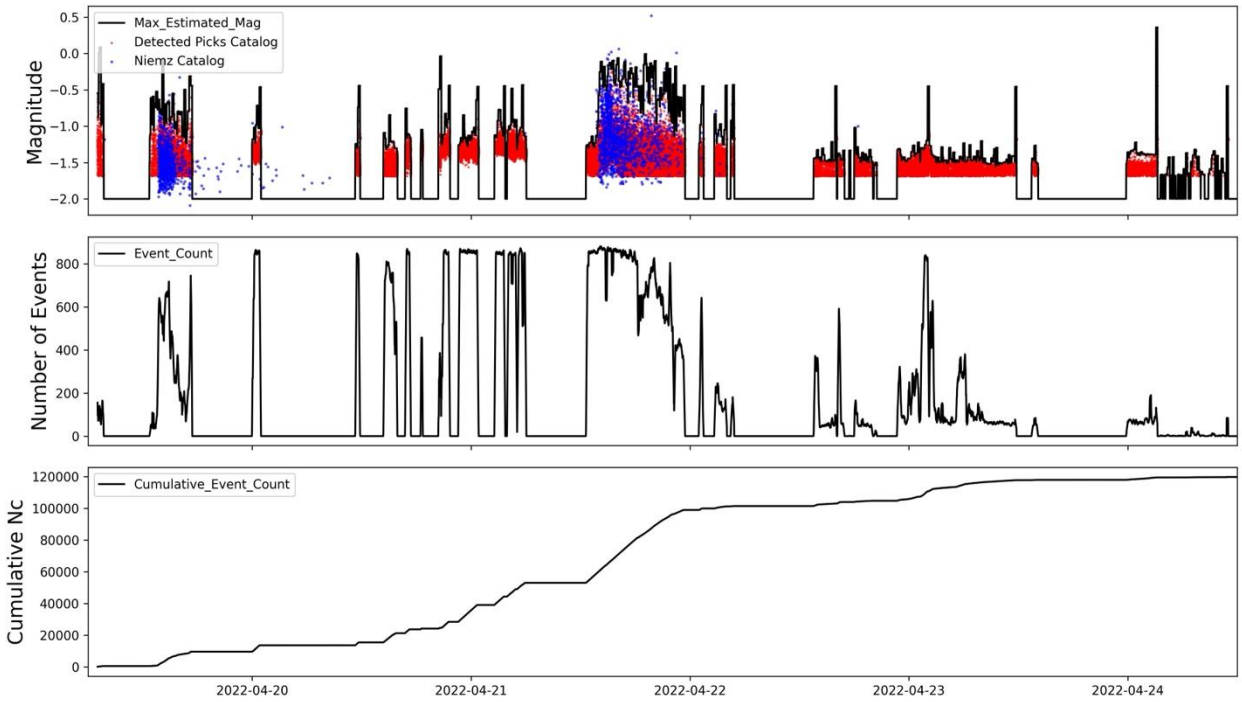

Figure 7: The max magnitude and number of event distributions of 2022 stimulation based on our workflow detected catalog.

4. Conclusion and next step

Here, we analyzed various methods for calculating Magnitude-Frequency Distributions (MFDs), recognizing their critical role in understanding seismic responses. A key observation is that regardless of the approach used, a high-precision earthquake catalog is essential for accurate MFD computation. However, due to data limitations, obtaining such high-resolution catalogs quickly and consistently remains a significant challenge. To address this, we constructed a workflow to generate an "inaccurate high-resolution catalog" directly from seismic waveforms. This workflow leverages advanced signal processing techniques, such as envelope computation and peak detection, to extract potential seismic events from continuous waveforms.

Despite the potential of this workflow, limitations in the dataset, such as the short observation window and low-resolution catalog, restricted its immediate application. As a result, we were unable to utilize the generated catalog to train a machine-learning model to predict MFDs effectively. This limitation underscores the importance of improving both the input data and the workflow itself to achieve more reliable results. In the next step, we will reanalyze the 2022 stimulation data alongside the upcoming 2024 stimulation data. Using these datasets, we will employ our workflow to generate high-resolution earthquake catalogs that can be used to train models for MFD prediction. This step will provide an opportunity to validate the workflow's effectiveness and its ability to enhance model training with more comprehensive and accurate data.

Additionally, we plan to refine the strategies within the workflow to improve the reliability and accuracy of the generated catalogs. These refinements include exploring better fitting functions to establish a more robust relationship between envelope amplitudes and earthquake magnitudes, adjusting the envelope computation method to capture signal characteristics more effectively, and optimizing the peak detection algorithm to minimize false positives and ensure higher catalog fidelity. These enhancements aim to address the current limitations and create a workflow capable of generating earthquake catalogs with greater precision and credibility. By integrating these improvements, our goal is to construct a robust and scalable framework for MFD generation and prediction. This next phase of research will bridge the gap between theoretical advancements and practical implementation, contributing to more accurate and reliable seismic modeling.

References

1. Aki, K. (1965), Maximum Likelihood Estimate of b in the Formula logN= a-bM and its Confidence Limits, Bulletin of the Earthquake Research Institute, 43, 237–239.

2. Bender, B. (1983). Maximum likelihood estimation of b values for magnitude grouped data. Bulletin of the Seismological Society of America, 73(3), 831-851.

3. Cao, A., & Gao, S. S. (2002). Temporal variation of seismic b-values beneath northeastern Japan island arc. Geophysical research letters, 29(9), 48-1.

4. Duarte, M. (2021). Detecta: A Python module to detect events in data (Version 0.0.5). Zenodo. https://doi.org/10.5281/zenodo.4598962

5. 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.

6. Li, Z., Eaton, D. W., & Davidsen, J. (2023). Physics-informed deep learning to forecast M^{\text{max}} during hydraulic fracturing. Scientific Reports, 13(1), 13133. Nature Publishing Group UK, London. https://doi.org/10.1038/s41598-023-40158-6

7. Li, Z., Eaton, D., & Davidsen, J. (2022). Short-term forecasting of M max during hydraulic fracturing. Scientific Reports, 12(1), 12509.

8. Niemz, P., McLennan, J., Pankow, K. L., Rutledge, J., & England, K. (2024). Circulation experiments at Utah FORGE: Near-surface seismic monitoring reveals fracture growth after shut-in. Geothermics, 119, 102947. Elsevier. https://doi.org/10.1016/j.geothermics.2023.102947

9. Reiter, L. (1990). Earthquake hazard analysis. Columbia University Press, New York, 254 pp.

10. Sobot, Tamara, et al. "An active learning framework for microseismic event detection." IGARSS 2024-2024 IEEE International Geoscience and Remote Sensing Symposium. IEEE, 2024.

11. Stein, S., & Wysession, M. (2009). An introduction to seismology, earthquakes, and earth structure. John Wiley & Sons.

12. Utsu, T. (1999). Representation and analysis of the earthquake size distribution: a historical review and some new approaches. Seismicity patterns, their statistical significance and physical meaning, 509-535.

13. Van der Elst, N. J., Page, M. T., Weiser, D. A., Goebel, T. H. & Hosseini, S. M. Induced earthquake magnitudes are as large as (statistically) expected. J. Geophys. Res. Solid Earth 121(6), 4575–4590 (2016).

14. Wiemer, S., & Wyss, M. (2000). Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the western United States, and Japan. Bulletin of the Seismological Society of America, 90(4), 859-869.