PROCEEDINGS, 50th Workshop on Geothermal Reservoir Engineering
Stanford University, Stanford, California, February 10-12, 2025
SGP-TR-229

Deep Neural Network-Based Workflow for Accurate Seismic Catalog Generation from Low-Resolution Seismic Data in Enhanced Geothermal System Operations

Xiaoming Zhang[1], Weiqiang Zhu[2], Rebecca O. Salvage[3,4], No'am Zach Dvory[1,5]

[1]Energy and Geoscience Institute, the University of Utah, Salt Lake City, UT, USA

[2]Department of Earth and Planetary Science, University of California Berkeley, Berkeley, CA, USA

[3]Department of Earth, Energy and Environment, University of Calgary, Calgary, Canada

[4]Natural Resources Canada, Geological Survey of Canada, Calgary, Canada

[5]Department of Civil and Environmental Engineering, the University of Utah, Salt Lake City, UT, USA

[1]Xiaoming.Zhang@utah.edu

[5]nzd@egi.utah.edu

Keywords: Enhanced Geothermal Systems (EGS), Deep neural networks (DNN), Real-time seismic monitoring, Fault detection

ABSTRACT

Real-time seismic monitoring is critical for the sustainable development of Enhanced Geothermal Systems (EGS). Accurate knowledge of event locations and induced seismic patterns is essential for effective seismic hazard mitigation. However, acquiring high-resolution seismic data in real-time is challenging, as temporary geophone arrays, while providing detailed records, are deployed for limited durations. To address this limitation, we applied a deep neural network (DNN)-based workflow to far-field sensors, allowing enhanced seismic information extraction and reduced reliance on dense sensor coverage.

In this study, the DNN-based approach was used to process seismic waveforms recorded during the 2024 Utah FORGE stimulation stages using permanent seismic stations located at a distance from the site. Despite the low spatial resolution of these stations, our results show a good agreement with the seismic catalog generated from high-resolution downhole geophones and Distributed Acoustic Sensing (DAS) systems and with the seismic catalog that was obtained from surface geophone temporary dense array records. The event locations, magnitude distributions, and detected fault structures derived from the DNN-based method closely align with those obtained from temporary geophone deployments, demonstrating the effectiveness of deep learning in enhancing seismic catalog generation from low-resolution seismic data. These findings highlight the potential of DNN-based methods for real-time, long-term seismic monitoring and fault detection, providing a cost-effective and scalable solution for seismic hazard mitigation in EGS operations.

1. INTRODUCTION

Enhanced Geothermal Systems (EGS) offer a promising avenue for renewable energy generation by utilizing the Earth's internal heat (Häring et al., 2008; Moore et al., 2019; Olasolo et al., 2016). However, the development of EGS is often accompanied by induced seismicity, which arises due to fluid injection and subsurface stress perturbations (Majer et al., 2007; Niemz, McLennan, et al., 2024). This poses operational and public safety challenges, as induced seismic events can lead to infrastructure damage and public concern, potentially limiting the acceptance and expansion of EGS projects (Ellsworth, 2013; Ellsworth et al., 2019).

Real-time seismic monitoring plays a crucial role in mitigating these risks by enabling operators to track seismic activity, identify active fault structures, and implement adaptive control strategies to minimize seismic hazards. High-resolution seismic data collected from temporary geophone arrays provide detailed insights into microseismic activity and fault slip dynamics, but their deployment is costly and typically limited in duration (Eyre et al., 2019). In contrast, permanent seismic stations offer a more economical and long-term monitoring solution; however, they generally suffer from lower spatial resolution and network density, which can lead to incomplete seismic catalogs and less accurate event locations (Niemz et al., 2024).

Recent advancements in machine learning, particularly the application of deep neural networks (DNNs), offer innovative solutions to these challenges. DNNs have demonstrated remarkable potential in analyzing seismic data, particularly in event detection, phase picking, and hypocenter localization, by extracting valuable information from sparse or noisy datasets (Mousavi et al., 2020; Zhang et al., 2022; Zhu et al., 2022; Zhu & Beroza, 2019). These methods have been successfully applied to improve the completeness of seismic catalogs, detect weak seismic signals, and enhance the accuracy of event location, even with sparse layout of seismic stations (Otake et al., 2020; Yang et al., 2020; Zhang et al., 2020).

This study explores the application of LOC-FLOW (Zhang et al., 2022), a DNN-based workflow, to enhance seismic monitoring during EGS operations. Specifically, we investigate the feasibility of utilizing data from the limited-resolution far-field seismic network at the Utah FORGE geothermal site to accurately detect and locate microseismic events associated with stimulation experiments conducted in

April 2024. By utilizing machine learning techniques, we aim to improve event detection capabilities and enhance the accuracy of seismic event locations, ultimately contributing to more effective seismic hazard mitigation strategies in EGS development.

2. METHODOLOGY

2.1 Data Sources

The seismic data used in this study were collected during the 2024 Utah FORGE stimulation stages (Dyer et al., 2024; Niemz, et al., 2024). The data sources include:

- **Permanent seismic stations:** A network of broadband seismic stations operated by the University of Utah Seismograph Stations (UUSS), located approximately 1–20 km from the FORGE stimulation site. These stations provide real-time seismic signals and are equipped with high-sensitivity sensors capable of detecting regional seismicity.

- **Temporary geophone arrays:** A dense network of 144 autonomous 5 Hz Fairfield Zland surface geophones were deployed (Niemz, Pankow, et al., 2024; Waldhauser, 2001). Each geophone was equipped with batteries providing up to 30 days of recording capability. Downhole geophones were installed at multiple depths in wells 16A, 16B, and Delano-1 to capture microseismic activity during the stimulations (Dyer et al., 2024).

- **DAS:** The distributed acoustic sensing (DAS) data was acquired continuously over the stimulation period at a temporal sampling rate of 10,000 Hz (10 kS/s) and a spatial resolution of approximately 3.35 feet (1.02109 meters), during stimulation activities conducted in April 2024 at wells 16A(78)-32 and 16B(78)-32.

The temporary downhole geophone arrays, surface geophone arrays, and DAS provide high-resolution seismic data, which serve as a benchmark for validating the DNN-based workflow applied to the permanent station data.

2.2 DNN-Based Workflow

The LOC-FLOW workflow consists of the following steps (Zhang et al., 2022):

2.1.1 Preprocessing

Seismic waveforms from permanent stations were retrieved from the Incorporated Research Institutions for Seismology (IRIS) data center using the ObsPy library (Waldhauser, 2001). Preprocessing was crucial to ensure data consistency and remove noise. Key steps include:

- Merging and Trimming: Merged individual waveform segments and trimmed data to the specified time window to ensure continuous time series.

- Resampling: Resampled waveforms to a uniform 100 Hz sampling rate for consistent analysis.

- Response Removal: Removed instrument response using pre-filtering to obtain ground motion in physical units.

2.2.2 Phase Picking

Traditional manual phase picking is time-consuming and prone to human error. PhaseNet employs a deep convolutional neural network (CNN) architecture to analyze three-component seismic waveforms (Zhu & Beroza, 2019). The CNN learns to extract relevant features from the waveforms, such as frequency content, amplitude variations, and phase shifts, to identify the arrival times of P and S waves. Instead of providing a single point estimate for arrival times, PhaseNet generates probability distributions for P arrivals, S arrivals, and noise. This probabilistic output provides a measure of confidence in the predicted arrival times.

2.2.3. Phase Association

Following phase picking with PhaseNet, the Rapid Earthquake Association and Location (REAL) algorithm was employed to associate seismic arrivals and locate the earthquake source (Zhang et al., 2019). REAL associates seismic phases by counting the number of P and S picks, calculating travel-time residuals, and determining the location of an event by finding the grid point with the most picks and the smallest travel-time residual.

2.2.4. Absolute Location

Absolute earthquake locations were determined using the HYPOINVERSE program (Klein, 2002). HYPOINVERSE is a robust, least-squares-based location algorithm that incorporates weights for individual phase picks. It refines initial location estimates and provides associated uncertainties, enhancing the reliability of the final earthquake locations.

2.2.5. Relative location

To further refine these locations and improve their relative accuracy, the hypoDD algorithm was employed (Waldhauser, 2001). hypoDD utilizes the double-difference method, which focuses on the relative arrival time differences between earthquakes at common stations. This approach effectively minimizes the impact of common path effects and improves the resolution of earthquake clusters, resulting in more precise relative locations within the study area.

2.2.6. Fault detection

To identify fault structures, we applied clustering techniques to the seismic catalog. The steps included:

- **Clustering algorithm:** Density-based spatial clustering (DBSCAN) was employed to group events based on their spatial proximity.
- **Fault detection:** Principal Component Analysis (PCA) was used to identify the fault planes that fit the clustered events.
- **Fault interpretation:** The properties of detected fault planes were calculated, such as the center position, length, dip, strike, and dip direction.

3. RESULTS

3.1 Spatial distribution of events

During the FORGE stimulation in April 2024, three primary sources were used to identify event locations: downhole geophones and Distributed Acoustic Sensing (DAS) (Dyer et al., 2024), surface geophones (Niemz, et al., 2024), and permanent seismic stations. The event catalog shown in Figure 1 was generated using downhole geophones and DAS systems, which provide high-quality waveform data for detecting microseismic events. The event locations in this catalog align well with the perforated zones documented in the April 2024 stimulation report (McClennan et al., 2024), validating the catalog's reliability. Figure 2 presents the event distribution obtained from a combination of five permanent seismic stations and a temporary deployment of 144 nodal geophones. Compared to Figure 1, a slight eastward shift in event locations is observed, though the overall distribution remains similar.

The catalogs in Figures 1 and 2 utilize extensive data from temporary geophones and DAS systems. These data sources present different challenges: the geophones provide only short-term monitoring, while the DAS data require computational resources and advanced algorithms for efficient interpretation. For real-time and long-term seismic monitoring, permanent stations offer a more practical solution. However, the number of permanent stations near the FORGE site is limited, and most are located at significant distances (Figure 3). This spatial constraint makes it challenging to generate an accurate event catalog using traditional event detection methods.

To overcome this limitation, we employed LOC-FLOW, a deep learning-based workflow (Zhang et al., 2022), to process raw waveform data from permanent stations and determine earthquake locations. The resulting microseismic catalog is shown in Figure 4. Overall, the LOC-FLOW-generated catalog closely matches the event distribution and timing in Figure 2. A noticeable eastward shift is observed in Figure 4, which is attributed to the 1D velocity model (Waldhauser, 2001) used in this study (Table 1). Future work will incorporate a 3D velocity model to enhance the accuracy of the deep learning-based catalog.

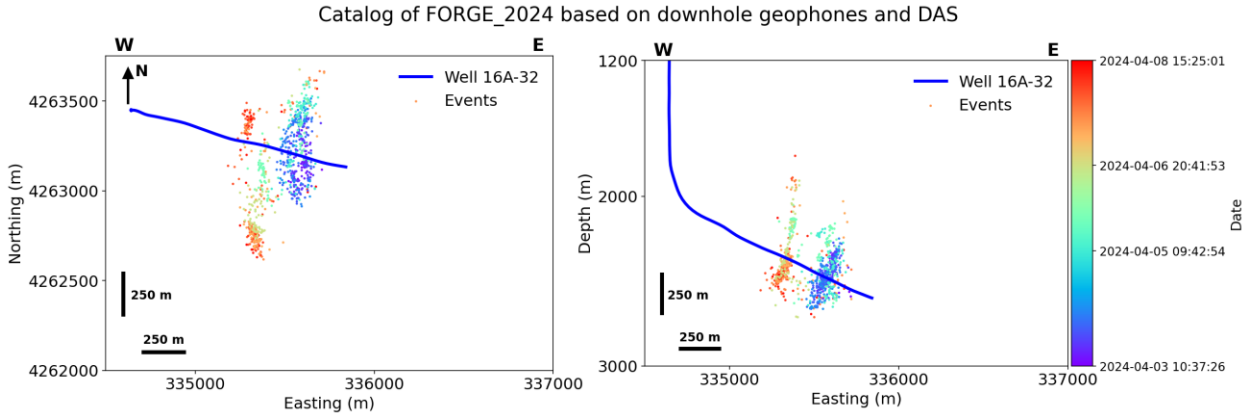

Figure 1: The spatial distribution of microseismic events detected by downhole geophones and DAS.

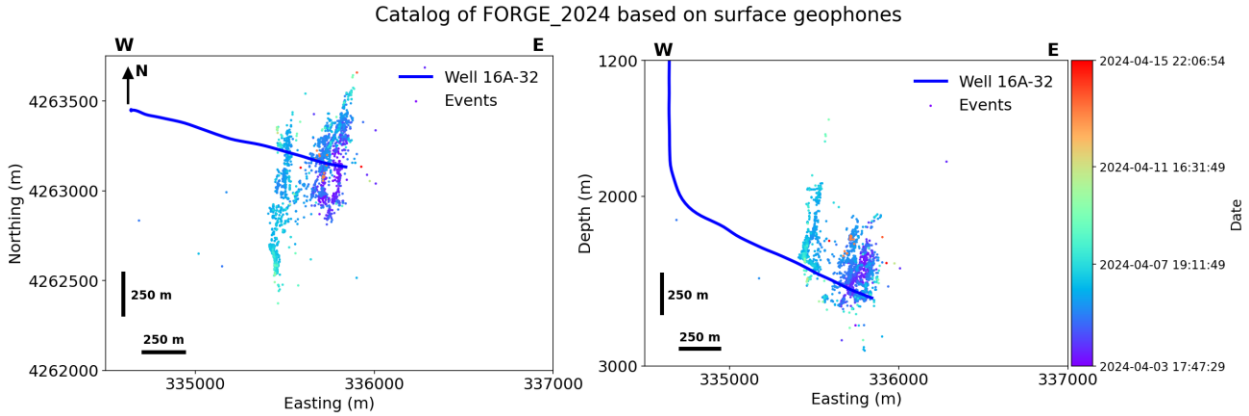

Figure 2: The spatial distribution of microseismic events detected by surface geophones.

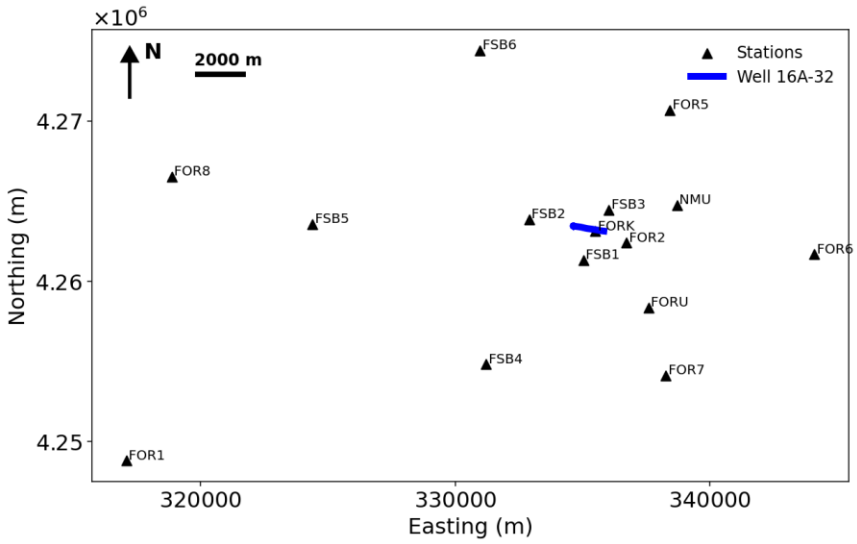

Figure 3: The location of the permanent stations used for generating the event catalog.

Table 1: The 1D velocity model used for generating the event catalog.

Depth [km]	Vp [km/s]	Vs [km/s]
0.0	3.4	1.95
1.69	3.4	1.95
1.69	5.9	3.39
17.25	5.9	3.39
17.25	6.4	3.68
28.15	6.4	3.68
28.15	7.5	4.31
42.15	7.5	4.31
42.15	7.9	4.54

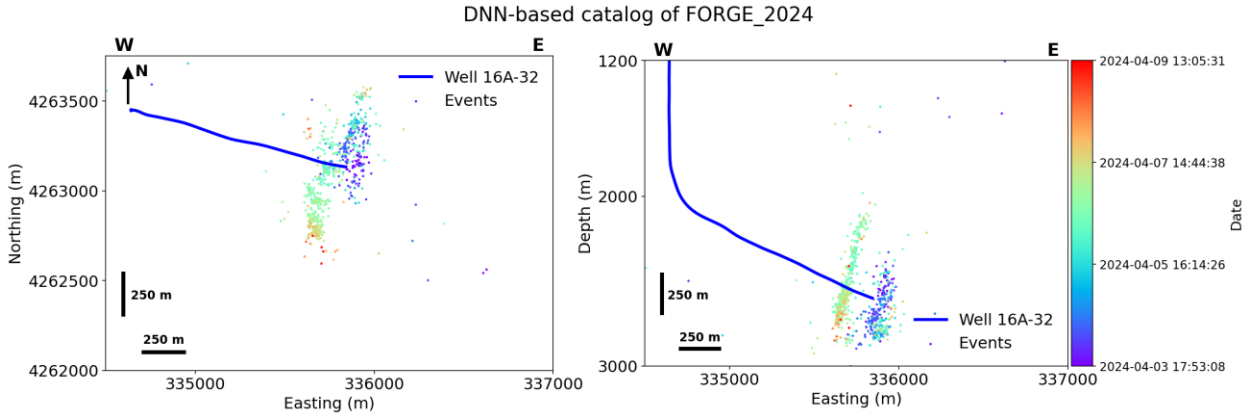

Figure 4: The spatial distribution of microseismic events detected by permanent stations based on deep learning workflow.

3.2 Magnitude

In addition to the spatial distribution of microseismic events, the event magnitude is a crucial parameter for evaluating the reliability and practicality of the event catalog. Figure 5 presents the magnitude distribution derived from downhole geophones and DAS data. The stimulation experiments conducted in April 2024 consisted of multiple stages with intermittent operational periods. This time discontinuity is reflected in Figure 5, where gaps in the event timeline correspond to inactive stimulation periods. The magnitude histogram reveals that most events fall within the range of -0.5 to 0.5, a reasonable range that minimizes the risk of inducing significant seismic activity.

Figure 6 illustrates the magnitude distribution obtained from surface geophones and a subset of permanent stations. Unlike Figure 5, the event timeline appears more continuous due to noise. Additionally, the recorded magnitudes predominantly range between -1.0 and -0.5, deviating from the values obtained using downhole geophones and DAS. This discrepancy suggests potential limitations in using surface geophones for accurate magnitude estimation.

Figure 7 presents the magnitude distribution obtained using the deep learning-based approach applied to data from permanent stations. Both the temporal distribution and the magnitude range closely align with those obtained from downhole geophones and DAS, although the distinct time discontinuity observed in Figure 5 is less apparent. This demonstrates the effectiveness of the deep learning-based method for estimating event magnitudes. Combined with the spatial distribution analysis, these results further validate the reliability of the deep learning-based workflow for seismic monitoring, even when relying on data from distant permanent stations. This highlights the feasibility of real-time event detection using permanent stations, providing a practical and scalable solution for real-time and long-term seismic monitoring.

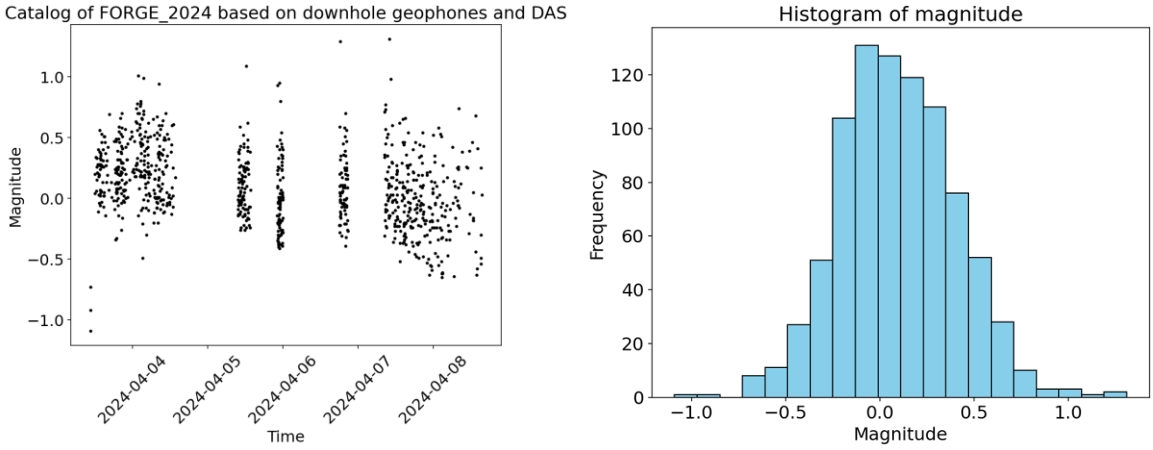

Figure 5: The magnitude distribution of microseismic events derived from downhole geophone and DAS data.

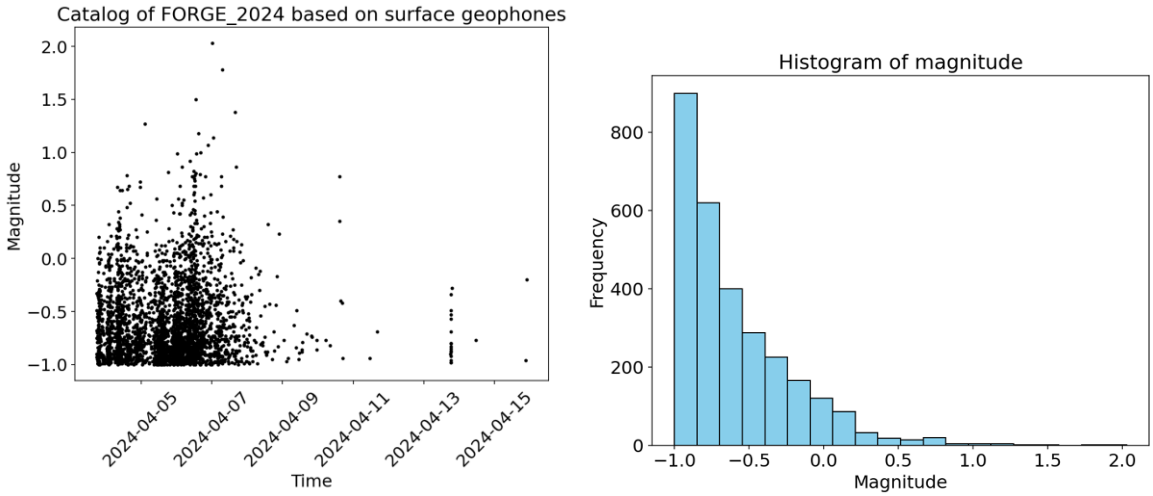

Figure 6: The magnitude distribution of microseismic events derived from surface geophones.

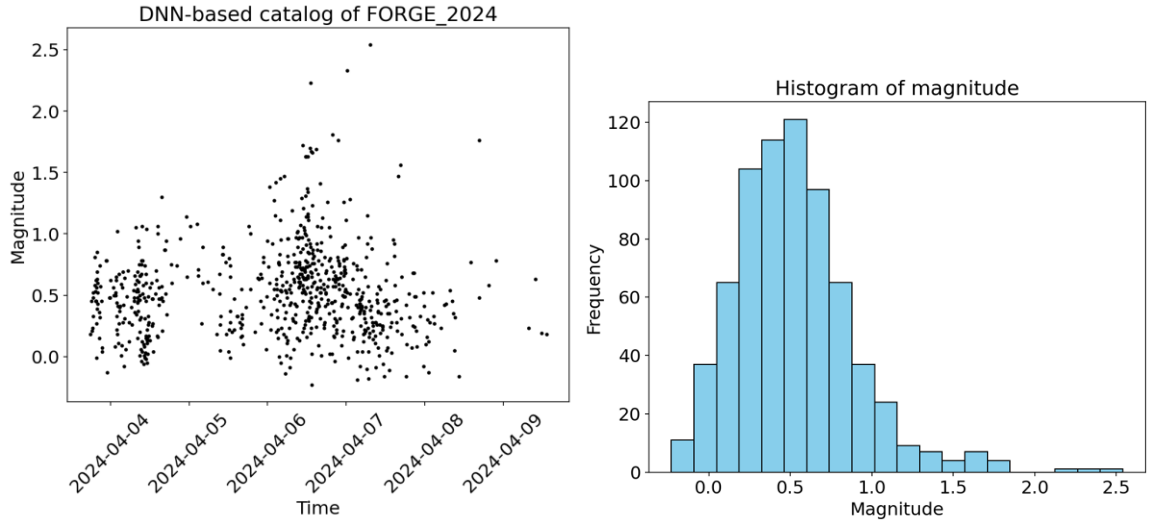

Figure 7: The magnitude distribution of microseismic events derived from permanent stations.

3.3 Detected faults

To further evaluate the reliability of the event catalog generated using permanent station data and the DNN-based workflow, we analyzed the spatial distribution of events to detect fault structures. The detected faults responding to the three data resources are shown in Figure 8, Figure 9, and Figure 10, respectively. The fault detection process involved clustering seismic events and identifying corresponding fault structures. We employed the DBSCAN algorithm, a density-based clustering method that groups events based on spatial proximity while automatically determining the optimal number of clusters. Once the seismic events were grouped, we assumed each cluster represented a fault. To further characterize these faults, Principal Component Analysis (PCA) was applied to fit planar structures to each cluster. Using PCA, we extracted key fault properties, including fault center, length, width, dip, dip direction, and strike. The properties of the detected faults from the three data sources are summarized in Tables 2, 3, and 4, respectively. Only faults with more than 100 associated events are included in these tables, as faults with fewer events may not provide a reliable statistical representation of their seismic characteristics. While these lower-event faults are depicted in the figures for completeness, they are omitted from the tables to ensure that only well-sampled faults are considered for detailed analysis.

The analysis reveals that many faults align in a north-south orientation, consistent with the maximum principal stress direction in the FORGE project region. Additionally, the dip angles of many detected faults fall within the range of 70–80 degrees, characteristic of hydraulic fractures. The detailed fault properties are provided in Tables 2, 3, and 4. A comparison of detected faults across different datasets shows strong agreement between event catalogs derived from different sources:

- Fault #58 in Table 3 (surface geophones) and Fault #10 (permanent stations) in Table 4 correspond to Fault #1 in Table 2 (downhole geophones and DAS).
- Fault #7 in Table 3 and Fault #2 in Table 4 match Fault #4 in Table 2.

- Fault #3 in Table 3 aligns with Fault #6 in Table 2.

More events were detected using surface geophones, consequently resulting in the detection of more faults than expected. This may be attributed to noise in the surface geophone data. In contrast, the permanent station data did not contain enough events to detect a fault that corresponds to Fault #6 in Table 2. This limitation highlights the challenge of using sparse permanent stations for fault detection. Despite this, the deep learning-based workflow applied to permanent station data successfully captured major fault structures, demonstrating its potential for long-term seismic monitoring when high-density temporary arrays are unavailable.

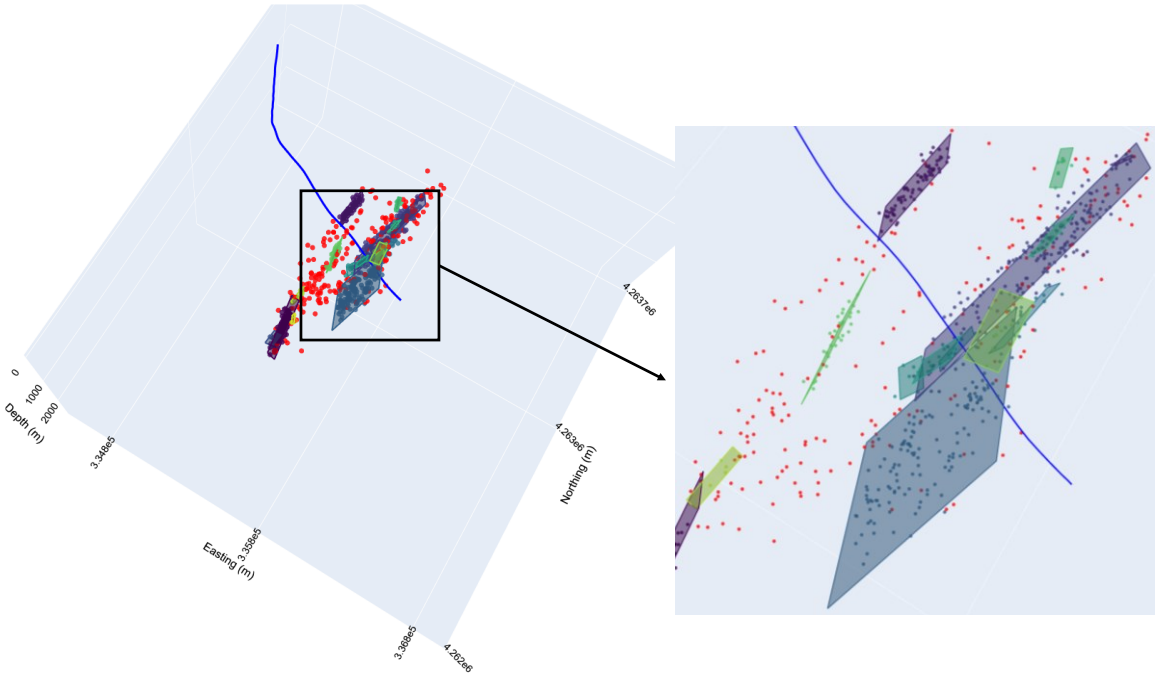

Figure 8: 3D depiction of detected faults using data from downhole geophones and DAS.

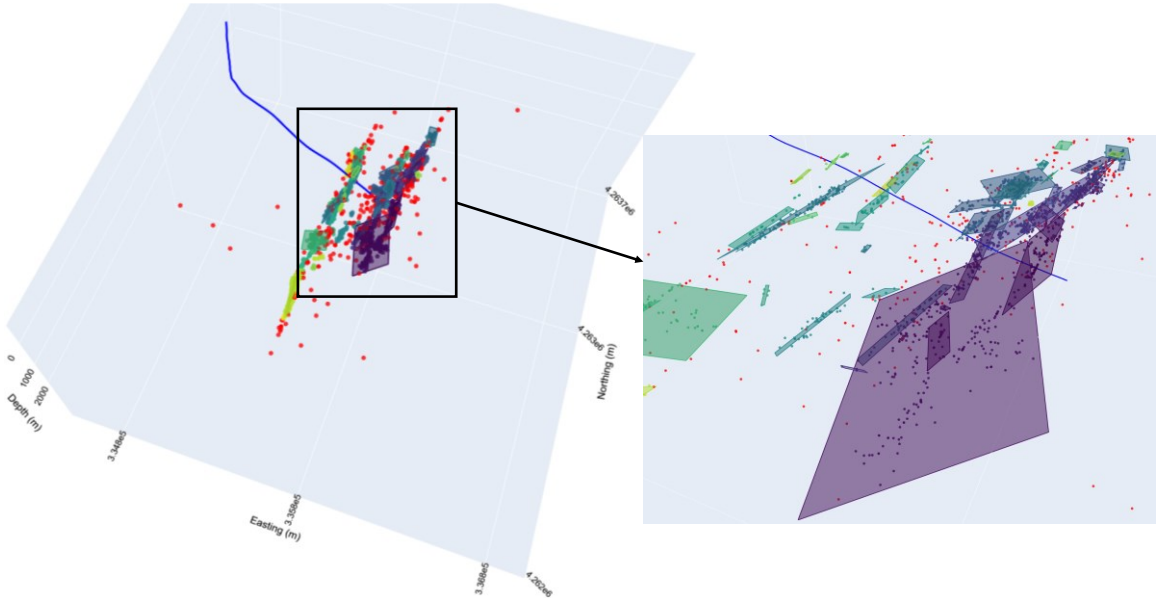

Figure 9: 3D depiction of detected faults using data from surface geophones.

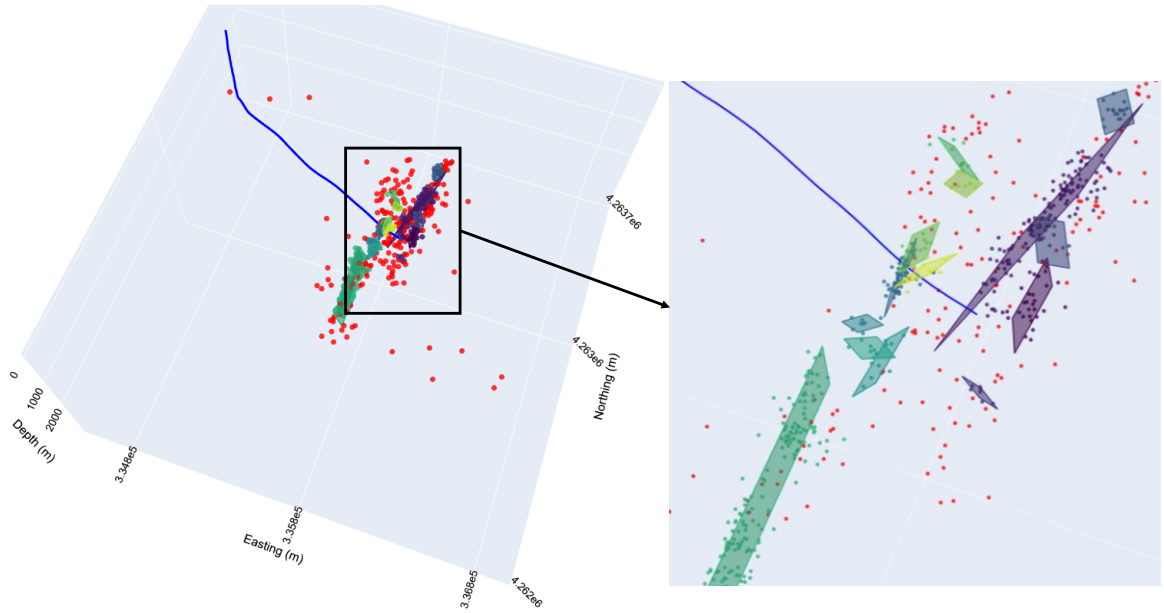

Figure 10: 3D depiction of detected faults using data from permanent stations.

Table 2: The properties of detected faults using data from downhole geophones and DAS.

Fault ID	Events	Center X (m)	Center Y (m)	Depth (m)	Dip	Strike	Dip direction
1	143	335328.9	4262755	2410.4	77.8	165.9	255.9
4	178	335575.7	4263369	2510.5	74	195.2	285.2
6	135	335589.7	4263063	2410.5	59.9	173.8	263.8

Table 3: The properties of detected faults using data from surface geophones.

Fault ID	Events	Center X (m)	Center Y (m)	Depth (m)	Dip	Strike	Dip direction
2	116	335801.9	4263098	2371.7	70	183.6	273.6
3	219	335745.5	4262945	2420.9	58.2	159.9	249.9
6	190	335711.9	4263128	2501.9	76.3	171.7	261.7
7	140	335827.7	4263337	2439.3	76.9	194.3	284.3
11	398	335763.3	4263280	2568.3	89.8	206.6	296.6
29	171	335491.8	4263063	2224.1	70.9	187.8	277.8
58	214	335442.3	4262634	2317.9	79.5	170.2	260.2

Table 4: The properties of detected faults using data from permanent stations.

Fault ID	Events	Center X (m)	Center Y (m)	Depth (m)	Dip	Strike	Dip direction
2	132	335875.8	4263309	2740.3	78.6	200.9	290.9
10	208	335666.7	4262866	2623.3	75.4	174.5	264.5

4. CONCLUSION

This study demonstrates the potential of a deep neural network (DNN)-based workflow to generate accurate seismic catalogs and identify fault structures using low-resolution seismic data from permanent stations. The application of this approach to the Utah FORGE 2024 stimulation experiments highlights its effectiveness in detecting microseismicity and mapping fault zones. This study compared seismic catalogs generated from three different data sources: downhole geophones and DAS systems, surface geophones, and permanent seismic stations processed with the DNN-based workflow (LOC-FLOW).

The event catalog from downhole geophones and DAS systems closely matches the FORGE 2024 April stimulation activities in terms of event spatial distribution, magnitude distribution, and detected fault structures. The catalog generated using surface geophones is similar to the downhole geophone catalog but contains more noise, leading to higher event counts and smaller magnitudes. The presence of noise also resulted in detecting more faults than expected. The DNN-based workflow applied to permanent station data produced an event catalog that showed strong agreement with the downhole geophone catalog. A systematic eastward shift in event locations was observed, likely due to the use of a 1D velocity model instead of a more accurate 3D velocity model. Despite this, the magnitude distribution and detected faults matched well with the downhole geophone catalog. Due to station sparsity, the total number of detected events was lower.

To further enhance the accuracy of microseismic event locations using permanent stations, future research will focus on implementing a 3D velocity model within the deep learning-based workflow. This improvement is expected to reduce spatial discrepancies and further refine event detection and fault characterization.

ACKNOWLEDGEMENT

This project was supported by the Utah FORGE initiative, sponsored by the U.S. Department of Energy, under the project titled “Cutting-edge application of machine learning, geomechanics, and seismology for real-time decision-making tools during stimulation”.

REFERENCES

Dyer, B., Karvounis, D., Meier, P., Fiori, R., & Jaques, P. (2024). Utah FORGE: GES Well 16A(78)-32 and Well 16B(78)-32 Stimulation Seismic Event Catalogs. University of Utah Seismograph Stations. https://doi.org/10.15121/2455012

Ellsworth, W. L. (2013). Injection-Induced Earthquakes. *Science*, *341*(6142), 1225942. https://doi.org/10.1126/science.1225942

Ellsworth, W. L., Giardini, D., Townend, J., Ge, S., & Shimamoto, T. (2019). Triggering of the Pohang, Korea, Earthquake (Mw 5.5) by Enhanced Geothermal System Stimulation. *Seismological Research Letters*, *90*(5), 1844–1858. https://doi.org/10.1785/0220190102

Eyre, T. S., Eaton, D. W., Zecevic, M., D’Amico, D., & Kolos, D. (2019). Microseismicity reveals fault activation before Mw 4.1 hydraulic-fracturing induced earthquake. *Geophysical Journal International*, *218*(1), 534–546. https://doi.org/10.1093/gji/ggz168

Häring, M. O., Schanz, U., Ladner, F., & Dyer, B. C. (2008). Characterisation of the Basel 1 enhanced geothermal system. *Geothermics*, *37*(5), 469–495.

Klein, F. W. (2002). *User’s guide to HYPOINVERSE-2000, a Fortran program to solve for earthquake locations and magnitudes*. US Geological Survey.

Majer, E. L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B., & Asanuma, H. (2007). Induced seismicity associated with Enhanced Geothermal Systems. *Geothermics*, *36*(3), 185–222. https://doi.org/https://doi.org/10.1016/j.geothermics.2007.03.003

McClennan, J., Swearingen, L., & England, K. (2024). Utah FORGE: Wells 16A(78)-32 and 16B(78)-32 Stimulation Program Report - May 2024. Energy and Geoscience Institute at the University of Utah. https://doi.org/10.15121/2483880

Moore, J., McLennan, J., Allis, R., Pankow, K., Simmons, S., Podgorney, R., et al. (2019). The Utah Frontier Observatory for Research in Geothermal Energy (FORGE): an international laboratory for enhanced geothermal system technology development. In *44th Workshop on Geothermal Reservoir Engineering* (pp. 11–13). Stanford University.

Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature Communications*, *11*(1), 3952. https://doi.org/10.1038/s41467-020-17591-w

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. https://doi.org/https://doi.org/10.1016/j.geothermics.2024.102947

Niemz, P., Pankow, K., Isken, M., Whidden, K., McLennan, J., & Moore, J. (2024). Utah FORGE: 2024 Stimulations Microseismic Event Catalog from Seismic Surface Network. University of Utah Seismograph Stations. Retrieved from https://gdr.openei.org/submissions/

Olasolo, P., Juárez, M. C., Morales, M. P., & Liarte, I. A. (2016). Enhanced geothermal systems (EGS): A review. *Renewable and Sustainable Energy Reviews*, *56*, 133–144.

Otake, R., Kurima, J., Goto, H., & Sawada, S. (2020). Deep Learning Model for Spatial Interpolation of Real-Time Seismic Intensity. *Seismological Research Letters*, *91*(6), 3433–3443. https://doi.org/10.1785/0220200006

Waldhauser, F. (2001). *HypoDD-A program to compute double-difference hypocenter locations*.

Yang, S., Hu, J., Zhang, H., & Liu, G. (2020). Simultaneous Earthquake Detection on Multiple Stations via a Convolutional Neural Network. *Seismological Research Letters*, *92*(1), 246–260. https://doi.org/10.1785/0220200137

Zhang, M., Ellsworth, W. L., & Beroza, G. C. (2019). Rapid Earthquake Association and Location. *Seismological Research Letters*, *90*(6), 2276–2284. https://doi.org/10.1785/0220190052

Zhang, M., Liu, M., Feng, T., Wang, R., & Zhu, W. (2022). LOC-FLOW: An end-to-end machine learning-based high-precision earthquake location workflow. *Seismological Society of America*, *93*(5), 2426–2438.

Zhang, X., Zhang, J., Yuan, C., Liu, S., Chen, Z., & Li, W. (2020). Locating induced earthquakes with a network of seismic stations in Oklahoma via a deep learning method. *Scientific Reports*, *10*(1), 1941. https://doi.org/10.1038/s41598-020-58908-5

Zhu, W., & Beroza, G. C. (2019). PhaseNet: a deep-neural-network-based seismic arrival-time picking method. *Geophysical Journal International*, *216*(1), 261–273. https://doi.org/10.1093/gji/ggy423

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), e2021JB023283. https://doi.org/https://doi.org/10.1029/2021JB023283