

Cindy S Y Lim¹, Sacha Lapins², Margarita Segou³, and Maximilian J Werner²

¹Affiliation not available

²School of Earth Sciences, University of Bristol, UK

³British Geological Survey, The Lyell Centre

February 28, 2024

Abstract

Deep learning (DL) phase picking models have proven effective in processing large volumes of seismic data, including successfully detecting earthquakes missed by other standard detection methods. Despite their success, the applicability of existing extensively-trained DL models to high-frequency borehole datasets is currently unclear. In this study, we compare four established models (GPD, U-GPD, PhaseNet and EQTransformer) trained on regional earthquakes recorded at surface stations (100 Hz) in terms of their picking performance on high-frequency borehole data (2000 Hz) from the Preston New Road (PNR) unconventional shale gas site, in the United Kingdom (UK). The PNR-1z dataset, which we use as a benchmark, consists of continuously recorded waveforms containing over 38,000 seismic events previously catalogued, ranging in magnitudes from -2.8 to 1.1. Remarkably, three of the four DL models recall a good fraction of the events and two might satisfy the monitoring requirements of some users without any modifications. In particular, PhaseNet and U-GPD demonstrate exceptional recall rates of 95% and 76.6%, respectively, and detect a substantial number of new events (over 15,800 and 8,300 events, respectively). PhaseNet's success might be attributed to its exposure to more extensive and diverse instrument dataset during training, as well as its relatively small model size, which might mitigate overfitting to its training set. U-GPD outperforms PhaseNet during periods of high seismic rates due to its smaller window size (400-samples compared to PhaseNet's 3000-sample window). All models start missing events below M_w -0.5, suggesting that the models could benefit from additional training with microseismic datasets. Nonetheless, PhaseNet may satisfy some users' monitoring requirements without further modification, detecting over 52,000 events at PNR. This suggests that DL models can provide efficient solutions to the big data challenge of downhole monitoring of hydraulic-fracturing induced seismicity as well as improved risk mitigation strategies at unconventional exploration sites.

Deep Learning Phase Pickers: How Well Can Existing Models Detect Hydraulic-Fracturing Induced Microseismicity from a Borehole Array?

Cindy S.Y. Lim¹, Sacha Lapins¹, Margarita Segou², Maximilian J. Werner¹

¹ *COMET, School of Earth Sciences, University of Bristol, Queens Road, BS8 1RJ, Bristol, UK.*

Email: cindy.lim@bristol.ac.uk

² *British Geological Survey, The Lyell Centre, EH14 4AP, Edinburgh, UK*

Received 2024 February 6; in original form 2024 February 6

SUMMARY

Deep learning (DL) phase picking models have proven effective in processing large volumes of seismic data, including successfully detecting earthquakes missed by other standard detection methods. Despite their success, the applicability of existing extensively-trained DL models to high-frequency borehole datasets is currently unclear. In this study, we compare four established models (GPD, U-GPD, PhaseNet and EQTransformer) trained on regional earthquakes recorded at surface stations (100 Hz) in terms of their picking performance on high-frequency borehole data (2000 Hz) from the Preston New Road (PNR) unconventional shale gas site, in the United Kingdom (UK). The PNR-1z dataset, which we use as a benchmark, consists of continuously recorded waveforms containing over 38,000 seismic events previously catalogued, ranging in magnitudes from -2.8 to 1.1. Remarkably, three of the four DL models recall a good fraction of the events and two might satisfy the monitoring requirements of some users without any modifications. In particular, PhaseNet and U-GPD demonstrate exceptional recall rates of 95% and 76.6%, respectively, and detect a substantial number of new events (over 15,800 and 8,300 events, respectively). PhaseNet's success might be attributed to its exposure to more extensive and diverse instrument dataset during training, as well as its relatively small

model size, which might mitigate overfitting to its training set. U-GPD outperforms PhaseNet during periods of high seismic rates due to its smaller window size (400-samples compared to PhaseNet’s 3000-sample window). All models start missing events below M_w -0.5, suggesting that the models could benefit from additional training with microseismic datasets. Nonetheless, PhaseNet may satisfy some users’ monitoring requirements without further modification, detecting over 52,000 events at PNR. This suggests that DL models can provide efficient solutions to the big data challenge of downhole monitoring of hydraulic-fracturing induced seismicity as well as improved risk mitigation strategies at unconventional exploration sites.

Key words: Induced seismicity – Machine learning – Earthquake monitoring and test-ban treaty verification – Neural networks, fuzzy logic – Computational seismology – Downhole methods.

1 INTRODUCTION

Hydraulic-fracturing induced seismicity (HFIS) can pose serious risks to a country’s infrastructure, energy security and communities (Li et al., 2019; Atkinson et al., 2020; Schultz et al., 2020). Examples of significant HFIS include the 2011 moment magnitude M_w 4.8 Eagle Ford earthquake in Texas, United States (Frohlich & Brunt, 2013), the 2016 M_w 4.1 Fox Creek earthquake in Alberta, Canada (Schultz et al., 2017), and the 2017 M_w 4.7 Changning earthquake in South Sichuan, China (Lei et al., 2017). The risks extend to wastewater disposal sites (Chen et al., 2017), enhanced geothermal sites (Grigoli et al., 2018) and potentially carbon capture and storage (CCS) sites (Verdon & Stork, 2016).

Microseismic monitoring of HFIS by operators is critical for potential risk mitigation measures and is a legal requirement in many countries (e.g., Wong et al., 2015; Kao et al., 2016; Clarke et al., 2019). In addition, high resolution datasets of small events can lead to a better understanding, modelling and forecasting of the mechanisms and hazards of HFIS (e.g., Eyre et al., 2019; Kettlety & Verdon, 2021; Mancini et al., 2022). To that end, arrays of high-frequency geophones installed in boreholes close to stimulation wells are particularly useful for detecting and characterising the

quickly attenuating high-frequency content of small events (Klinger & Werner, 2022; Holmgren et al., 2023).

However, borehole arrays with high sampling frequencies can generate substantial volumes of data, as monitoring arrays may include tens to hundreds of seismic stations. Examples of large borehole datasets include the ToC2ME dataset in Fox Creek, Alberta (Eaton et al., 2018), the Horn River basin dataset in British Columbia, Canada (Verdon & Budge, 2018), and the FORGE geothermal experiment dataset in Utah, United States (Shi et al., 2022). Processing large datasets requires significant time and costly computational resources, posing an issue in generating event catalogues that are crucial for informing our understanding of subsurface activities, especially in real-time.

Deep learning (DL) phase pickers offer a solution to efficiently picking seismic events in large volumes of continuous seismic data. Perol et al. (2018), Ross et al. (2018), Zhu & Beroza (2019) and Mousavi et al. (2020) demonstrated that DL models can pick events more efficiently than standard approaches and human analysts. Specifically, DL neural networks such as GPD (Ross et al., 2018), U-GPD (Lapins et al., 2021), EQTransformer (Mousavi et al., 2020) and PhaseNet (Zhu & Beroza, 2019) have undergone extensive training to proficiently pick seismic phases within large datasets. Moreover, DL phase pickers have detected new events overlooked by conventional methods, thereby contributing, for instance, to uncovering the complexities of fault structures (Tan et al., 2021) and earthquake swarm dynamics, providing new insights into aseismic crustal processes (Ross & Cochran, 2021).

The aim of this paper is to assess the performances of various existing DL models when picking microseismic phases in high-frequency borehole data without training. Most existing DL models, including GPD, U-GPD, EQTransformer and PhaseNet, are trained using 100 Hz data collected from surface seismic stations. These training sets mainly consist of larger, regional-sized earthquakes ($M_w > 0$) but also encompass a number of smaller, local earthquakes ($M_w < 0$). However,

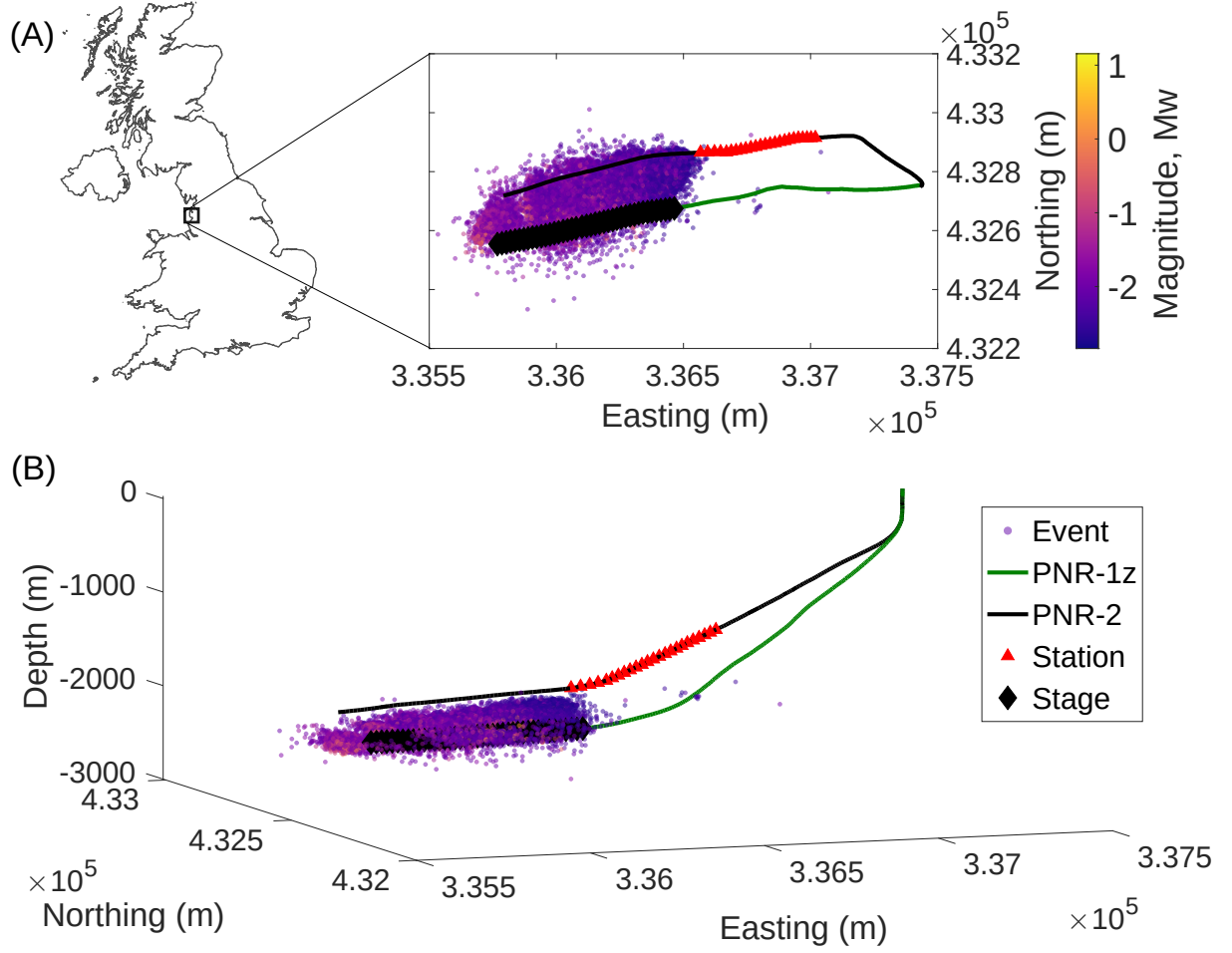


Figure 1. A geographical map showing the (A) plan view of the Preston New Road unconventional shale gas exploration site and a (B) 3-D section of the site including seismicity (coloured circles), geophones (red triangles), stages (black diamonds) on the PNR-1z (green line) and PNR-2 (black line) wells.

it is not clear that the models can generalise to detect microseismicity in high frequency data in a borehole setting. In this study, we apply and compare these models using the Preston New Road 1z (PNR-1z) shale gas exploration dataset (Clarke et al., 2019), where the largest earthquake recorded was of magnitude M_w 1.1. This benchmark catalogue contained over 38,000 events recorded on an array of 24 borehole geophones during injection activities at a sampling rate of 2000 Hz (Fig. 1). Furthermore, we assess whether DL models identify additional microseismic events, and we compare each model in terms of phase detection and picking ability.

Table 1. The DL phase pickers used in this study.

Models	GPD	U-GPD	EQTransformer	PhaseNet
Input window size	400	400	6000	3000
Magnitude range	-0.81 to 5.7	-0.81 to 5.7 and -0.4 to 3.6	-0.5 to 7.9	0 to 5
Architecture type	CNN + FCNN	U-Net	CNN + LSTM + attention	U-Net
N of params	1,741,003	672,419	376,935	269,675

2 DL MODELS

We evaluate four DL models (Table 1): the Generalized Seismic Phase Detection (GPD) model by Ross et al. (2018), the U-GPD model by Lapins et al. (2021), EQTransformer (EQT) by Mousavi et al. (2020) and PhaseNet by Zhu & Beroza (2019). We select these models because they have publicly available code, working GitHub repositories, and the models were extensively trained on large datasets for generalised phase picking. Additionally, a previous study by Münchmeyer et al. (2022) found that GPD, PhaseNet and EQT were the best performing models for earthquake detection, phase classification and onset time determination for regional and teleseismic datasets (not including downhole, high frequency datasets). These models, however, were primarily trained on 100 Hz 3-component seismograms from surface seismic stations and were not specifically trained, nor tested, on high-frequency borehole datasets.

GPD is a convolutional neural network (CNN) trained on 4.5 million waveforms (1.5 million each for P, S and noise). Ross et al. (2018) trained this model with magnitudes M from -0.81 to 5.7 recorded by the Southern California Seismic Network. The model uses a 400-sample sliding window on continuous three component data to input into convolutional layers for feature extraction. These extracted features are then input into a fully connected neural network (FCNN) for phase classification. The model outputs single class probability values for P, S and noise (i.e., model output dimensions are 3×1) for each window. When the phase probability is above a user-defined threshold, a phase is declared in the middle of the window.

U-GPD is a DL model that modifies the GPD model by using a fully convolutional U-Net

architecture and fine-tuning the base GPD weights with an additional dataset. Lapins et al. (2021) fine-tuned the weights using a limited volcano-seismic dataset (from the Nabro volcano, Eritrea) sampled at 100 Hz. The Nabro dataset contains 2,498 event waveforms with local magnitudes ranging from -0.4 to 3.6. U-GPD uses the same input as GPD (3×400 sliding window) but replaces the FCNN with additional convolutional layers. The new layers were initialised with randomised weights and then trained on the Nabro dataset. As U-GPD is a fully-convolutional model, its output differs from GPD as U-GPD estimates a class probability for each sample in the window (i.e., model output dimensions are 3×400). The probability traces help the network pick more precisely as it removes the ambiguity arising from a single class prediction over an entire signal window (Lapins et al., 2021).

The PhaseNet model, similar to U-GPD, is a U-Net fully convolutional neural network developed by Zhu & Beroza (2019). However, unlike GPD and U-GPD, PhaseNet uses a larger 3000-sample sliding window as its input. PhaseNet’s training dataset consists of over 600,000 event waveforms from the Northern California Earthquake Data Center. Their training dataset comprises a magnitude range from M 0 to 5. A unique characteristic of this training dataset is that Zhu & Beroza (2019) included different types of instruments (e.g. accelerometers, high gain and low gain seismometers) during training, which might help generalise its ability to phase pick in different data.

EQT is an attention-based DL model trained using 1 million labelled earthquake waveforms and 300,000 noise waveforms Mousavi et al. (2020) from the STanford EArthquake Dataset (STEAD), a global dataset of earthquakes. Out of the models in this study, EQT has the largest data input window of 6000 samples. The model has an encoder, which extracts high level representations of the data from the continuous seismic signals and three decoder branches that use these data representations to generate three probability traces (for the presence of an earthquake, the P-wave and the S-wave, respectively). EQT consists of convolutional layers (CNN), long-short-term-memory (LSTM) layers and a hierarchical attention mechanism. EQT’s attention mechanism

visually weights sections of data in its 6000-sample input window both globally (full waveform) and locally (where it expects the P and S phase). Earthquake magnitudes in STEAD range from M -0.5 to 7.9. In the EQT repository, there are two versions of this model. For our study, we use the original EQT version that has been optimised to maximise the number of detections.

Applying these diverse DL models allows for a comprehensive comparison considering variations in model architecture, input window sizes, training data and the range of trained earthquake magnitudes (Table 1). This assessment aims to identify the model specifications that are most advantageous for monitoring microseismicity in high frequency borehole data.

3 DATA

3.1 Preston New Road (PNR-1z) continuous downhole dataset

Cuadrilla Resources Ltd. monitored the hydraulic stimulation of the PNR-1z well using a borehole geophone array in the PNR-2 well (Clarke et al., 2019). Figure 1 shows the downhole array of 24 Avalon Geochain Slimline 15 Hz geophones that continuously recorded seismic data of the operations that took place from 15 October to 17 December 2018. Hydraulic fracturing operations were carried out from 15 October to 2 November, followed by an injection hiatus from 3 November to 7 December 2018 and then further stimulations from 8 to 17 December 2018.

The PNR-1z continuous downhole dataset was recorded at 2000 Hz from 8 October to 18 December 2018 and a total data volume of 3.92 Tb was acquired, available in 16-second SEG-Y files. Each SEG-Y file contains 3-component seismic data for 24 stations, resulting in 72 traces per file. These continuous traces serve as input to the DL models.

From visual inspection of the waveforms, we observe that the amplitudes of events above moment magnitude M_w -0.237 are clipped because the operators set a high gain on the recording instruments to detect more microseismic events (James Verdon, personal communication). Al-

though a lot of the larger events were clipped, this should not greatly affect phase detection as DL models primarily focus on the onset of phase arrivals (Ross et al., 2018). Still, we assess whether clipping affects phase detection by the deep learning models.

3.2 Existing catalogue and injection data

We use the Coalescence Microseismic Mapping (CMM) catalogue as a benchmark to compare with the other DL models. Schlumberger (Cuadrilla’s contracted processor) produced an event origin time catalogue, with locations and magnitudes, from the PNR-1z dataset, using proprietary code based on the CMM method (Drew et al., 2013). The CMM catalogue contains event origin times and locations of a total of 38,452 events with magnitudes that span $-2.839 \leq M_w \leq 1.155$. Fig. 2 shows the events and their magnitudes overlaid with the injection rate data over time.

The CMM method is a computationally intensive multi-station simultaneous detection and location approach that generates characteristic functions for each station using STA/LTA ratios (Drew et al., 2013). The migration of signals from multiple stations to a coherent source in time and space makes CMM a robust method for earthquake detection, with the added advantage of simultaneously determining hypocentres. While it has been proven to be a robust method for microseismic monitoring (Smith et al., 2015), it is constrained by the sliding time-windows it employs on continuous data, detecting only the largest event (energy maxima) within a fixed time window. We set the CMM as a good benchmark for these single station DL models while also exploring whether they can outperform the CMM method and provide a more efficient alternative for monitoring microseismicity.

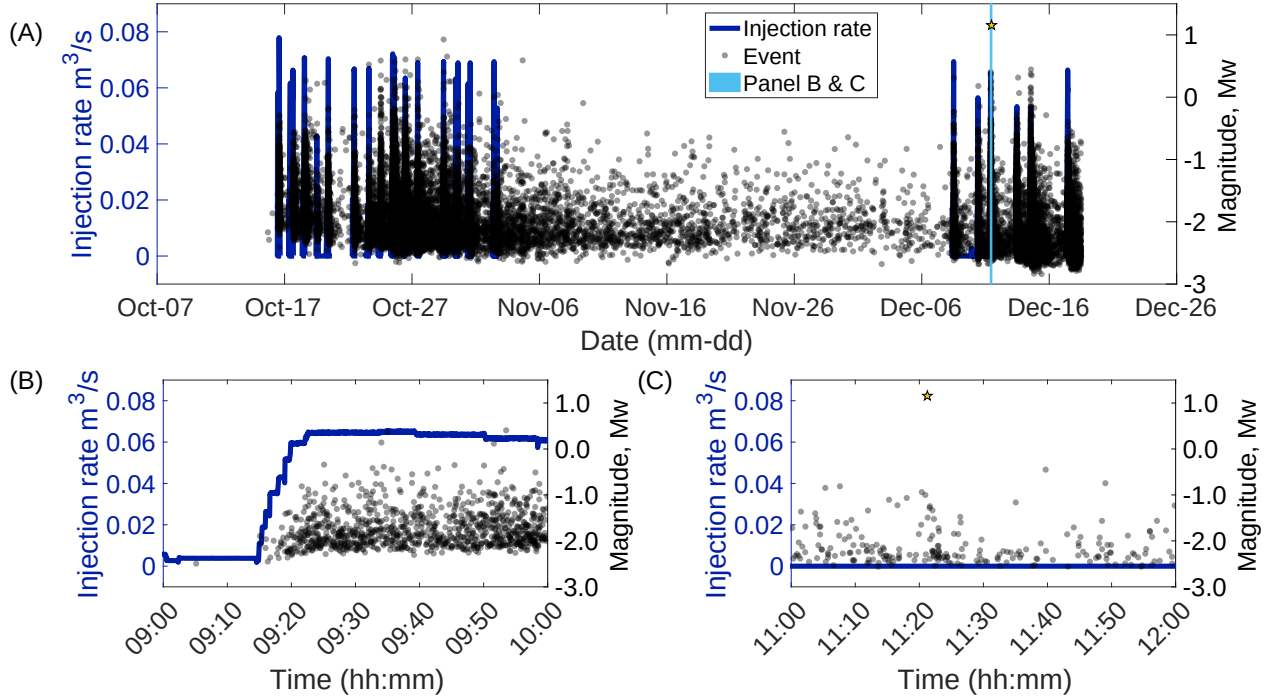


Figure 2. Temporal plot of the injection rates (blue line, left axis) and CMM catalogued events (black circles) with their respective moment magnitudes, M_w , (right axis) for (A) the whole duration of the PNR-1z continuous downhole data, (B) during a period of high injection rates (11 December 2018, 9am-10am) and (C) during a period without injection (11 December 2018, 11am-12pm). The largest earthquake (M_w 1.1) of the catalogue is shown as a yellow star.

4 METHODS

4.1 Data pre-processing

We pre-process the raw continuous data by rotating the waveforms from their respective station orientations to the E, N and Z components. For each model's input, we then apply a 50 Hz Butterworth high-pass filter. This ensures that the pre-trained models can pick microseismic phases within the noisy downhole data as microseismicity typically has low signal-to-noise ratio (SNR). We choose a 50 Hz high-pass filter by examining the Fast Fourier Transform (FFT) event spectra on the geophones closest to (deepest) and farthest from (most shallow) the source (perforations in the PNR-1z well).

4.2 Classification test

We first compare each model’s ability to classify seismic phases (P, S and noise) in the PNR-1z continuous downhole waveform data from single stations, which we refer to as phase classification tests. Our test dataset comprises 750 sections of data from 250 random events (i.e., 250 P, 250 S and 250 noise) on random stations. We then record the class labels produced by each model for each section. Each section is a 3-second (6000 sample) window of 3-component continuous data containing a phase arrival in the window. We filter, visually inspect, and manually select each section from the CMM catalogue. We also randomly select stations to avoid bias, especially as some stations are less noisy or closer to events than other stations (e.g., shallow stations are further away from the events). If a model produces more than one label in the section, we record the class label closest to the manually determined pick time.

We calculate recall, precision and F1-scores for each phase as well as overall metrics to assess the classification performance of each model. For overall recall and precision, we compute the average values for each class (P, S, and noise). Overall F1-score is calculated by using the average P and R values.

Recall R measures the completeness of positive predictions and is calculated as

$$R = \frac{(TP)}{(TP + FN)} \quad (1)$$

where TP is the number of true positives and FN is the number of false negatives.

For example, when calculating the recall value for the ‘P’ phase of a model, we determine that: true positives (TP) are the number of ‘P’ labels correctly identified as ‘P’; false negatives (FN) are the number of ‘P’ labels that the model misclassified as ‘S’ or ‘noise’; and their sum ($TP + FN$) is the total number of actual ‘P’ labels in the test dataset, which is fixed at 250.

R values close to 1 signify that the model produces a low number of false negatives. For example, a high R value for the P phase shows that the model does not miss a lot of P labels that are in the test set.

Precision P measures the accuracy of positive predictions by the model and is calculated as

$$P = \frac{(TP)}{(TP + FP)} \quad (2)$$

where FP is the number of false positives. When calculating the precision value for the ‘P’ phase of a model, we define: true positives (TP) are the number of ‘P’ labels correctly identified as ‘P’; false positives (FP) are the number of ‘S’ or ‘noise’ labels that the model misclassified as ‘P’; their sum (TP + FP) is the total number of ‘P’ classifications that the model predicted in the test dataset.

High precision P (close to 1) signifies a low number of false positive classifications (i.e., a high P shows that the model does not falsely classify S or noise labels as P phases). The F1-score weighs the P and R values equally and represents the harmonic mean of both values. F1-scores close to 1 suggest a good balance of high P and R for the model. The F1-score is given by

$$F1 = \frac{2TP}{2TP + FP + FN} = 2 \times \frac{P \times R}{P + R} \quad (3)$$

4.3 Catalogue comparison

4.3.1 PNR-1z downhole dataset

To complement the phase classification tests, we also compare the origin time catalogue generated from arrival picks of each model when applied to the PNR-1z continuous downhole dataset. The initial (CMM) method catalogue serves as our benchmark. We analyse the proportion of events recalled from the CMM catalogue, new candidate seismic events (i.e., prior to quality control of the events) and the events missed by the DL models.

Our workflow is structured to group phase picks into events, associate phases (P to S) and then

locate events to produce an origin time catalogue. We start by constructing an event catalogue using model phase picks across multiple stations. Here, we define a candidate seismic event as a group of P phase picks from at least four different stations within a fixed 0.2 second time window. We choose a 0.2 second window because it corresponds to the longest travel time for a phase of the same event to travel from one end to the other end of the borehole array, given the available velocity model (https://github.com/cindylimsy/PNR_run). We set the event defining threshold to at least four different stations because it is the typical minimum number of stations required to constrain location in time and space.

Next, we employ a straightforward approach for P to S phase association. We impose a fixed-time window from the initial P arrival time at each station. Specifically, we use a time difference of 0.3 seconds. This assumes that any S phase that is picked within the time range $t_p < t_s \leq t_p + 0.3$ will be associated with the initial P pick. The fixed time difference of 0.3 seconds theoretically corresponds to a 1.8 km radius. This time window for phase association adequately encompasses all the events in the target area (see Fig. S1, Supporting Information). Finally, we use the Non-Linear Location (NonLinLoc) algorithm (Lomax et al., 2000) to obtain origin times. For the full workflow, we conduct the model tests with one GeForce GTX 1080 GPU on a cluster node with one CPU.

Each DL model has a user-defined detection threshold parameter. When the probability of a phase (P or S) crosses that threshold, the model picks and labels the arrival in time. As Zhu & Beroza (2019) recommended the lowest threshold (0.3) compared to the other models, we set the same probability threshold for most of the models (GPD, U-GPD and PhaseNet). After initial parameter testing, we find that EQT requires a very low threshold to be able to detect any events on the PNR-1z dataset, so we use a threshold of 0.01 for EQT.

We compare the event origin times derived from DL model phase picks with the CMM catalogue. We assume that DL events within a window of ± 0.25 seconds of an origin time listed in the

benchmark (CMM) catalogue are considered matches. We label any event that is not a match in the CMM catalogue as a new candidate event. In Fig. S2 (Supporting Information), we show that the small origin time residuals between the PhaseNet and CMM catalogues validate the adequacy of the ± 0.25 -second window for matching DL phase pick-derived event origin times with the catalogued events.

4.3.2 Selected time series

To provide a more detailed comparison, we focus on two different periods of interest to assess each model: on 11 December 2018, from 9am to 10am, and from 11am to 12pm. We select an hour of data when the injection rate and seismicity were high (9am to 10am, Fig. 2b) and an hour when the seismicity was relatively quiet and injection was paused (11am to 12pm, Fig. 2c). We can thus observe which models are able to pick phases during a very active period (i.e. with high event rates) and during a quieter period of seismicity. The 11am to 12pm period also contains the largest (M_w 1.1) earthquake in the dataset. We additionally use this period to assess whether the models detect more aftershocks after the M_w 1.1 event.

4.4 Candidate seismic event validation using LinMEF: a Linear Moveout Event Filter

As DL models have demonstrated the ability to detect a substantial number of new events (Beroza et al., 2021; García et al., 2022), we develop an automated validation method to ensure that the new events originate from our target area, i.e., from within the zone of the hydraulic fracturing stimulations. We visually evaluate the subsets of the new events to ensure that we do not exclude potentially interesting events (see Results section: 5.2)

Our method exploits the expected move-out of phase arrivals across a near-linear borehole array (Fig. 3). By imposing the near-linear moveout pattern, we filter the model detections (candidate events) to focus on events arriving from the PNR-1z well stimulations. This method leverages the

information gathered from multiple geophones, taking advantage of the wealth of data obtained from a borehole array of individual geophone observations. We restrict our approach to just the P picks of each new event, as the DL models picked more P phases compared to S phases.

Because the hydraulic fracturing stimulations occur below the borehole array, we make two key assumptions: firstly, the P picks should exhibit moveout from the deepest to the most shallow geophone, and secondly, the P-wave velocity across the array should approximately be the velocity of a P-wave travelling through shale (4,700 m/s). These assumptions allow us to filter candidate events that deviate from the expected near-linear moveout, thereby retaining only the seismicity originating from the stimulated volume of rock (i.e., from the in-zone).

To validate the thousands of new machine learning-picked detections, we utilise the above information and apply a method we call the Linear Moveout Event Filter (LinMEF) to the seismic phase picks of each event. LinMEF is a simple and robust method consisting of three steps.

Firstly, for each new event, we use the group of P picks across the borehole array associated with the specific event origin time. We obtain a list of time differences t_{diff} between each pick time in the group, t_{pick} , and the earliest arrival time, t_0 :

$$t_{diff} = t_{pick} - t_0 \quad (4)$$

When plotting each t_{diff} against each corresponding station depth, we observe that the data points roughly follow a straight line for events arriving from the in-zone. The inverse of this gradient is representative of the P-wave velocity.

Secondly, after calculating t_{diff} from all available picks across the array for all events, we choose a random subset of 250 newly identified candidate events as our calibration dataset. We visually check the pick moveouts for each event and labelled this calibration dataset as either 'target' or 'non-target' events depending on their moveout. 'Target' events refer to events that: have a near-linear moveout; arrive from the bottom to the top of the array; and have apparent velocities

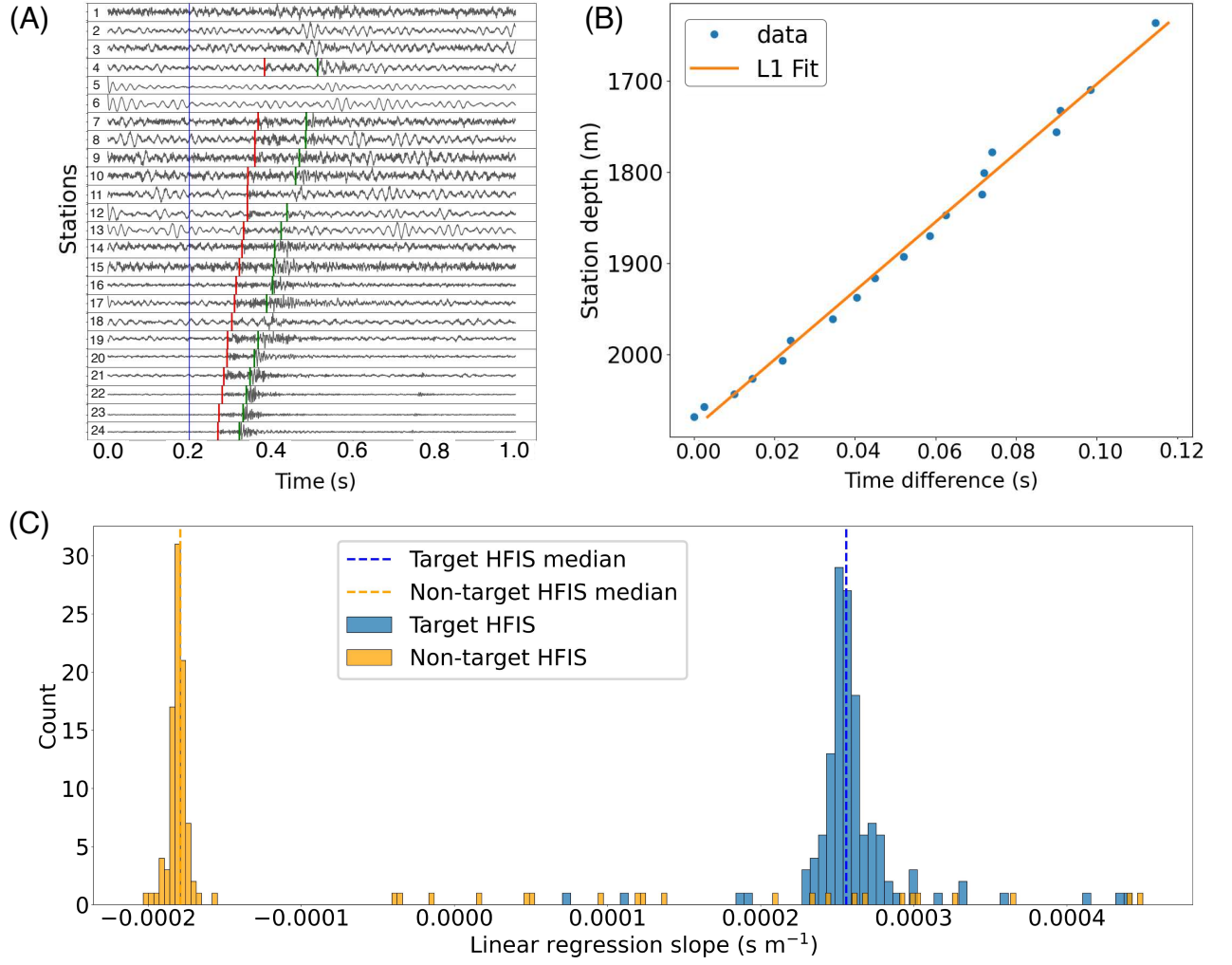


Figure 3. (A) An example of a new event picked by PhaseNet with P (red) and S (green) arrival picks arriving across the borehole array (geophones numbered on the side of each seismogram from 1 to 24). The vertical blue line represents the event origin time. (B) Time differences between the earliest P arrival pick time (t_0) and each individual pick time (t_{pick}) of the new event against station depth. We fit an L1-norm best-fit line to the datapoints. (C) A histogram of estimated linear regression slopes (slowness) of 250 random new candidate seismic events. The target events (blue) and non-target events (orange) labelled from manual visual inspection. The dotted lines represent the median value of the gradients for each label (target and non-target HFIS).

close to the P-wave velocity in shale (4,700 m/s) shown in Fig. 3(a).

‘Non-target’ events are candidate events that characteristically deviate from the ‘target’ events.

We further discuss these non-target events in the Results section: 5.2.

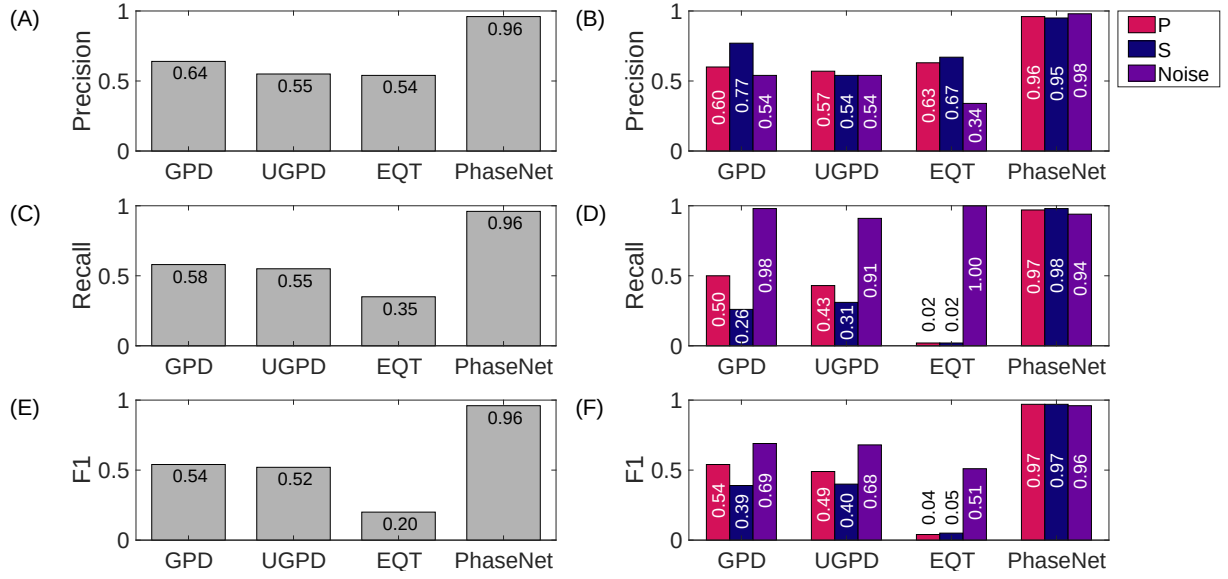


Figure 4. The left column of panels shows the (A) overall precision, (C) overall recall and (E) overall F1-scores of each model. The right column shows the individual P (pink), S (dark blue) and Noise (purple) phases of the (B) precision, (D) recall and (F) F1-scores.

We use this dataset to constrain the range of gradients (i.e., the approximate inverse of P-wave velocity) of the P picks that belong to ‘target’ and ‘non-target’ labelled events. We estimate the L1-norm best-fit line of the arrival pick times across all stations for every event. The L1-norm reduces undesired effects from pick time outliers on the arrival time gradient estimates, as compared to the L2-norm. Fig. 3(c) shows that the linear regression slopes of ‘target’ and ‘non-target’ labelled events fall into a bimodal distribution.

Lastly, we calculate the 90% range (using the 5th and 95th percentiles) of the resulting gradients around the ‘target events’ labelled distribution to define our event filter. We subsequently classify any new candidate events with a linear moveout slope within this 90% range as ‘target’ events and any that fall outside this range as ‘non-target’ events.

5 RESULTS

5.1 Picking assessment: classification test results

We compute precision (P), recall (R), and F1-scores from the confusion matrices of each model (see Supporting Information). The classification test results show that PhaseNet outperforms other models in terms of precision and recall (Fig. 4). Notably, EQT exhibits low P and R , resulting in the lowest F1-score, demonstrating a tendency to misclassify phases as noise. The U-GPD and GPD models have P and R values ranging between 0.44 and 0.64, implying occasional misclassifications. Overall, PhaseNet consistently demonstrates the highest P , R and F1-score values, highlighting its accuracy in classifying known seismic phases. Most models exhibited superior performance in classifying individual P phases compared to S phases (Fig. 4f). Higher F1-scores for P phase classifications also suggest that the DL models more successfully identified P phases than S.

Visual inspection of PhaseNet’s phase misclassifications revealed heightened noise levels around the phase in the data or when a larger event was present within the normalised window, making the smaller magnitude target phase located in the middle of the window appear smaller and more challenging to detect.

In terms of the picking precision of each model, PhaseNet consistently demonstrates the most precise and accurate picks (Fig. 5). In contrast, the GPD model exhibits less precise picking due to its output of a single phase probability value for each data window shift, leading to ambiguity in the arrival timing. The PhaseNet and U-GPD models, which follow a U-Net styled architecture, show precise picking with temporally continuous probability traces. Fig. 5 illustrates the phase arrival picks from each model, clearly indicating the precise phase picking by the U-GPD and PhaseNet models, while the GPD model produces imprecise picks. It should be noted that the EQT model, although capable of picking phases precisely, frequently misses phase arrivals in general.

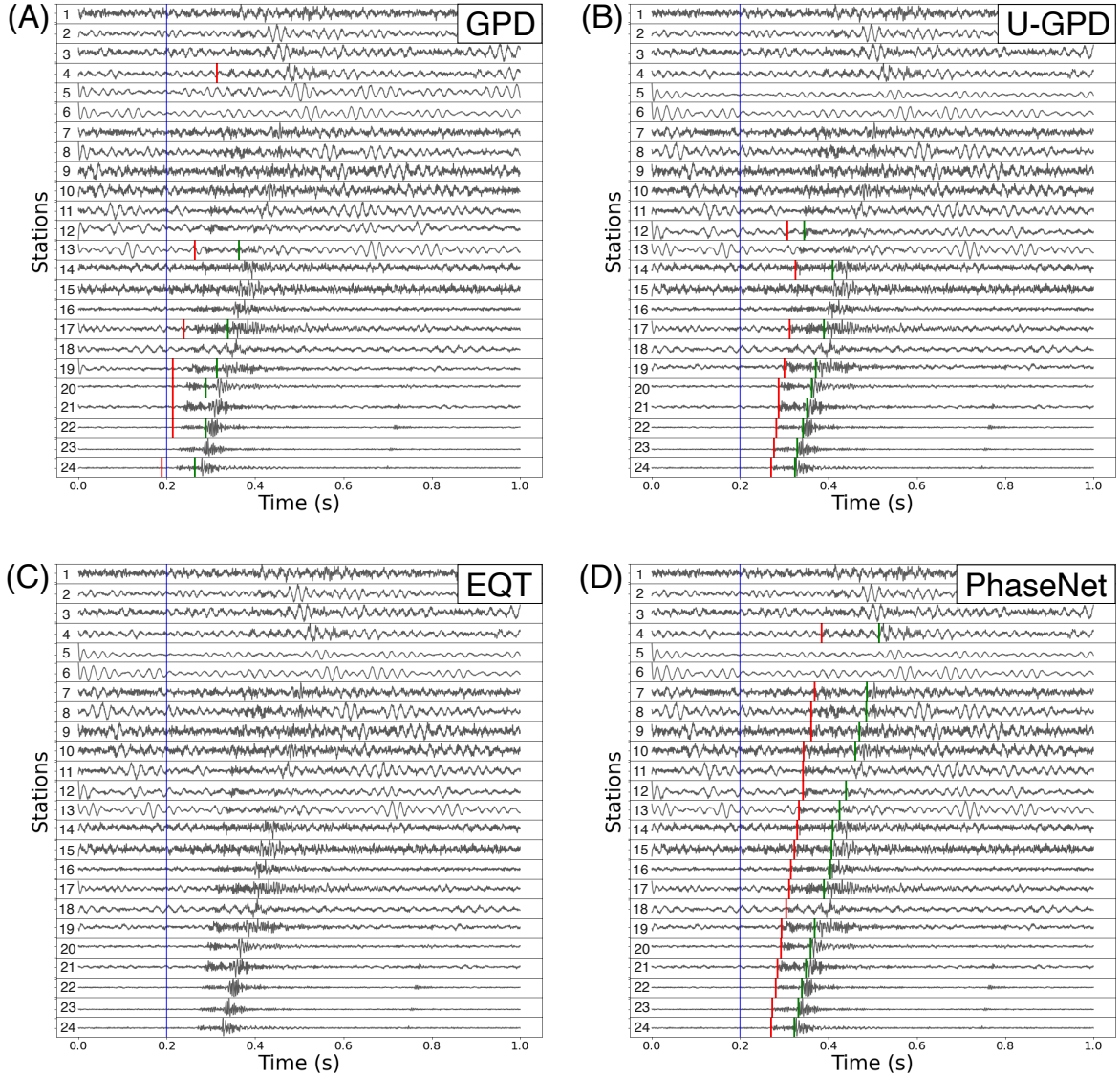


Figure 5. Seismograms (Z component) of the same event across the borehole array from Stations 1 (top row, most shallow) to 24 (bottom row, deepest) showing the phase classification and picking precision with P picks (red vertical lines) and S picks (green vertical lines) from the (A) GPD, (B) U-GPD, (C) EQTransformer and (D) PhaseNet models. The blue vertical line across all stations is the inferred event origin time.

5.2 Characterisation of non-target events

We observe that the DL models consistently detect events or phase arrivals that we do not believe to be induced seismic events associated with the injection. Figs. 5 and 6 illustrate the two types of candidate events detected, namely 'target events' and 'non-target events', respectively. The non-target events belong to at least four types of events that do not exhibit characteristics indicative of

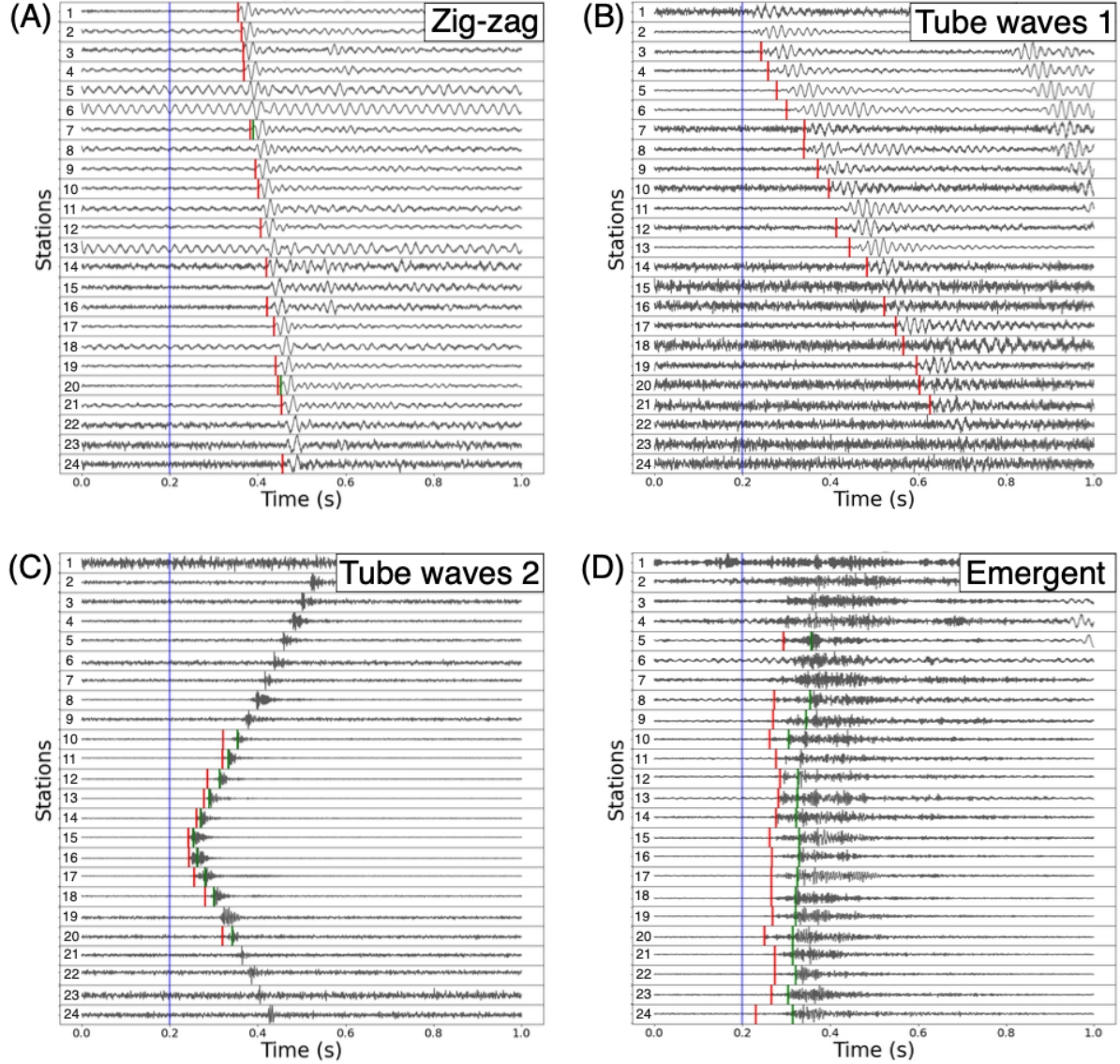


Figure 6. Seismograms of the four types of non-target events: (A) 'zig-zag', (B) 'tube waves 1', (C) 'tube waves 2' and (D) 'emergent' events across the borehole array from Stations 1 (top row, most shallow) to 24 (bottom row, deepest) showing the P picks (red vertical lines) and S picks (green vertical lines) from PhaseNet. The blue vertical line across all stations is the inferred event origin time.

induced seismicity resulting from the hydraulic fracturing stimulations through the PNR-1z bore-
hole. We label these signals as 'zig-zag' events, 'tube waves 1', 'tube waves 2' and 'emergent'
events. Through visual observations and event characterisation based on apparent velocities and
angle of incidence, we determine that these signals are likely unrelated to our target events.

'Zig-zag' events are waves that arrive from the opposite direction (from shallow to deep) com-

pared to induced seismicity (from deep to shallow). They have an average apparent velocity of 7,500 m/s. This is significantly faster than the P-wave velocity in shale (4,700 m/s) and in the steel casing (4,500 m/s). Most of these events do not have S picks, and they we only occur during working hours, between 8am to 4pm.

'Tube waves 1' are low frequency and travel at an average apparent velocity of 1,445 m/s, which is slower than the P-wave velocity in shale but close to the velocity of P-waves in water (1,500 m/s). Tube waves are a product of Rayleigh waves interacting with the wellbore, propagating along the walls of a fluid-filled borehole (Sheriff, 2002; Gadallah & Fisher, 2008).

'Tube waves 2' are higher frequency tube waves that appear to arrive at the broadside of the borehole array and travel from between the geophones on the PNR-2 well throughout the array. They do not arrive at any particular single station and their average velocity is similar to that of tube waves (1,445 m/s).

The 'emergent' events, which occur at high frequencies and do not follow a consistent move out, present a unique challenge. We observe a gradual energy buildup in these arrivals, rendering it challenging for the models to precisely identify the phase onset. The emergent nature of these events complicates picking, which results in an inconsistent moveout pattern. These events arrive almost instantaneously, though the arrival patterns vary randomly along the array. In addition, both the frequency content and amplitude of these events deviate characteristically among the geophones. None of these events correspond to any catalogued regional earthquakes in the UK. Given the unusual observations, we conclude that these 'emergent' events are not HFIS.

We visually label a random sample of 250 candidate seismic events detected by the best-performing DL model, PhaseNet. Our observations show that 48.4% (121) of these events were induced seismicity with a near-linear P-wave moveout (i.e., target events), while 40.8% (102) were 'zig-zag', 2.8% (7) corresponded to 'tube waves 1' or 'tube waves 2', and 2.4% (6) were 'emer-

gent' events. To verify the detection of HFIS using these DL models, we implement the Linear Moveout Event Filter (LinMEF) method (detailed in the Methods section: 4.4) to eliminate these non-target events from the DL event catalogues.

5.3 Model catalogue comparison

Fig. 7 shows that PhaseNet successfully recalls the most previously catalogued events (36,375 events, 94.6%) from the full continuous dataset, missing only 5.4% (2077 events) of the CMM events. Additionally, PhaseNet detects the highest number of undetected events (+41.2%), corresponding to 15,835 new events after filtering out non-target events using LinMEF. Comparatively, U-GPD outperforms GPD as it recalls more catalogued events (76.6% and 59.5%, respectively) and detects more new events (8,302 events, +21.6% and 1,918 events, +4.99%, respectively). On the other hand, EQT struggles to detect events, identifying less than 1% of the CMM catalogue (392 events) and 8 new events.

Most of the models begin to miss events at magnitudes below $M_w -0.5$ (Fig. 8). Additionally, we note the magnitude ranges for the recalled and missed events of each model. Fig. 8 shows that EQT was unable to detect events over the $-3 \geq M_w \geq 1$ range, suggesting its limitation in recognising microseismic phases within the higher frequency dataset. We also remark that the DL models' phase picks are not affected by clipping by visually checking the picks of 100 of the largest events with clipped amplitudes. The modelled P and S phase picks are still picked accurately and precisely.

Results from the selected hours show that during the injection period (Fig. 9), both PhaseNet and U-GPD successfully recall a majority (>90%) of the CMM catalogue, while also detecting the highest number of new events within the hour. During high injection rates, U-GPD detects more new events than PhaseNet, whereas the reverse is true during the quieter period. We also observe increases in seismicity rates during both selected hours for PhaseNet and U-GPD compared to the

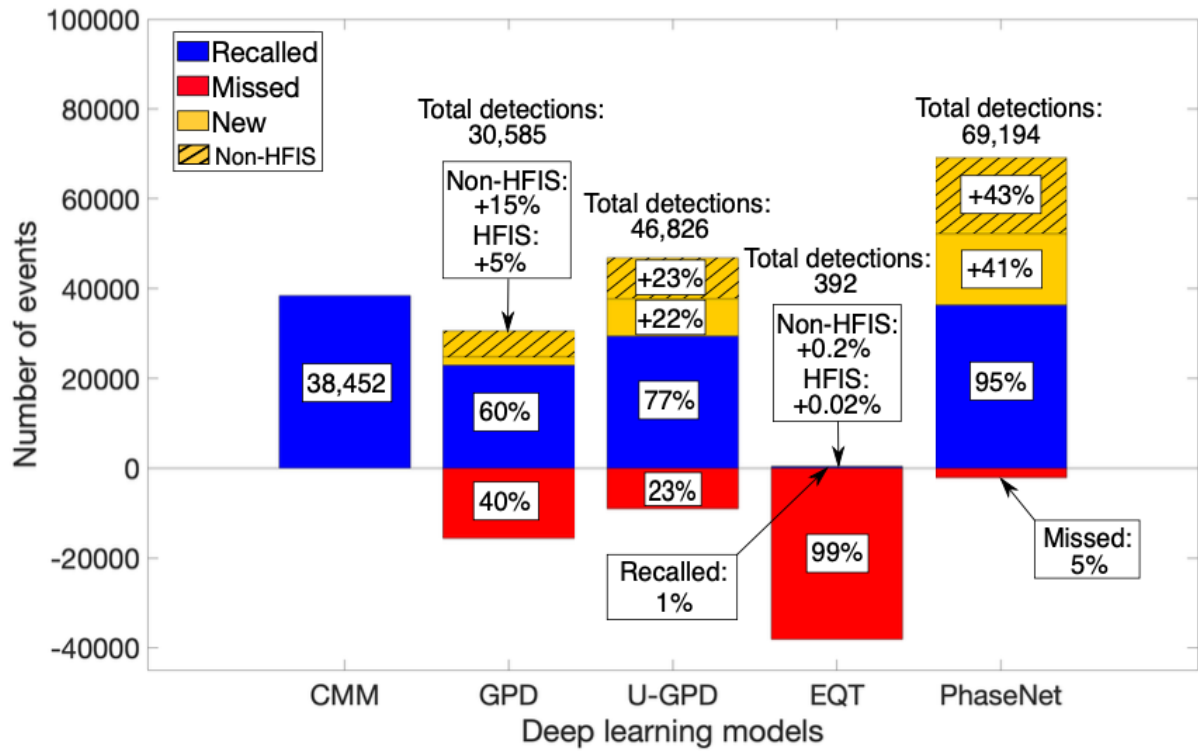


Figure 7. A comparison of the total number of model event detections in the full continuous PNR-1z dataset, including the new (yellow), recalled (blue) and missed (red) events. We filtered the new candidate seismic events using the LinMEF method, so the cross-hatched yellow sections represent the number of non-target events, and the non-hatched yellow sections illustrate the number of target events (i.e., QC’ed new events). The annotations in squares refer to the percentage with respect to the total number of CMM-detected events.

CMM method.

Figs. 10(b) and (d) demonstrate that PhaseNet outperforms U-GPD during the quieter period, with the exception of struggling to detect small events after the initial largest event when a cluster of aftershocks occurred. This suggests that PhaseNet’s performance is adversely affected by high seismicity rates. Furthermore, we notice that PhaseNet begins missing more small events during high seismic activity, likely due to the smaller amplitudes of these events as normalisation may be biased towards the increased presence of larger events.

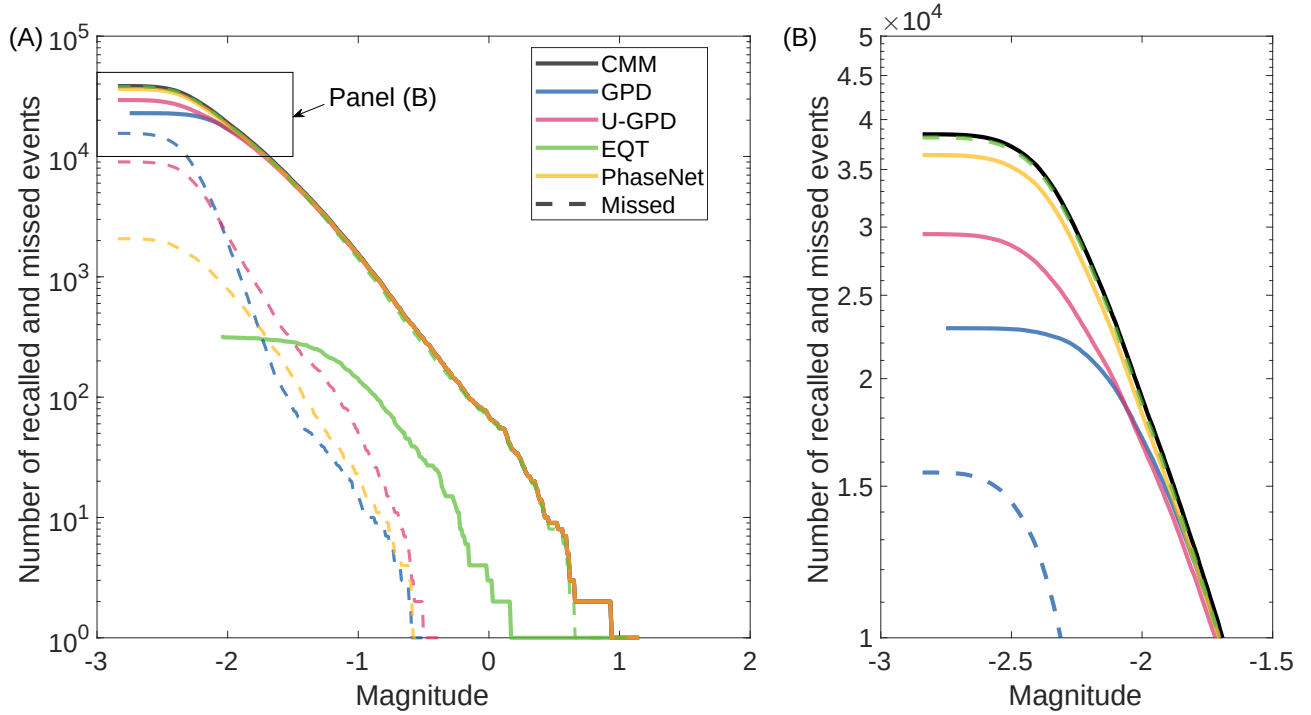


Figure 8. Cumulative frequency of each model against magnitude for the recalled (solid lines) and missed (dotted lines) events from the GPD (blue), U-GPD (red), EQT (green) and PhaseNet (orange) models compared to the CMM (black) catalogue. We used the estimated magnitudes from the CMM catalogue so new events are not included in this plot.

5.4 Model runtimes and computational efficiency

We present the model runtime statistics and performance metrics based on one hour of continuous borehole data (Table 2). We use the NVIDIA Tesla K80 12 GB GPU on Google Colab for these tests. Our findings reveal that that all DL models take less time to process than the given data window length (one hour of data). In ascending order, EQT, PhaseNet, GPD and U-GPD demonstrate processing speeds surpassing real time by 2.88, 3.72, 4.94 and 12.75 times faster, respectively. Although these figures might seem modest, it's crucial to consider the high sampling frequency of our test data (2000 Hz). One hour of 2000 Hz data is equivalent to 20 hours of 100 Hz data. When scaled to standard 100 Hz data, the processing speed of the models are 57.6, 74.4, 98.8 and 255 times faster, respectively. These results highlight the remarkable efficiency of these models, highlighting their suitability to real-time microseismic monitoring.

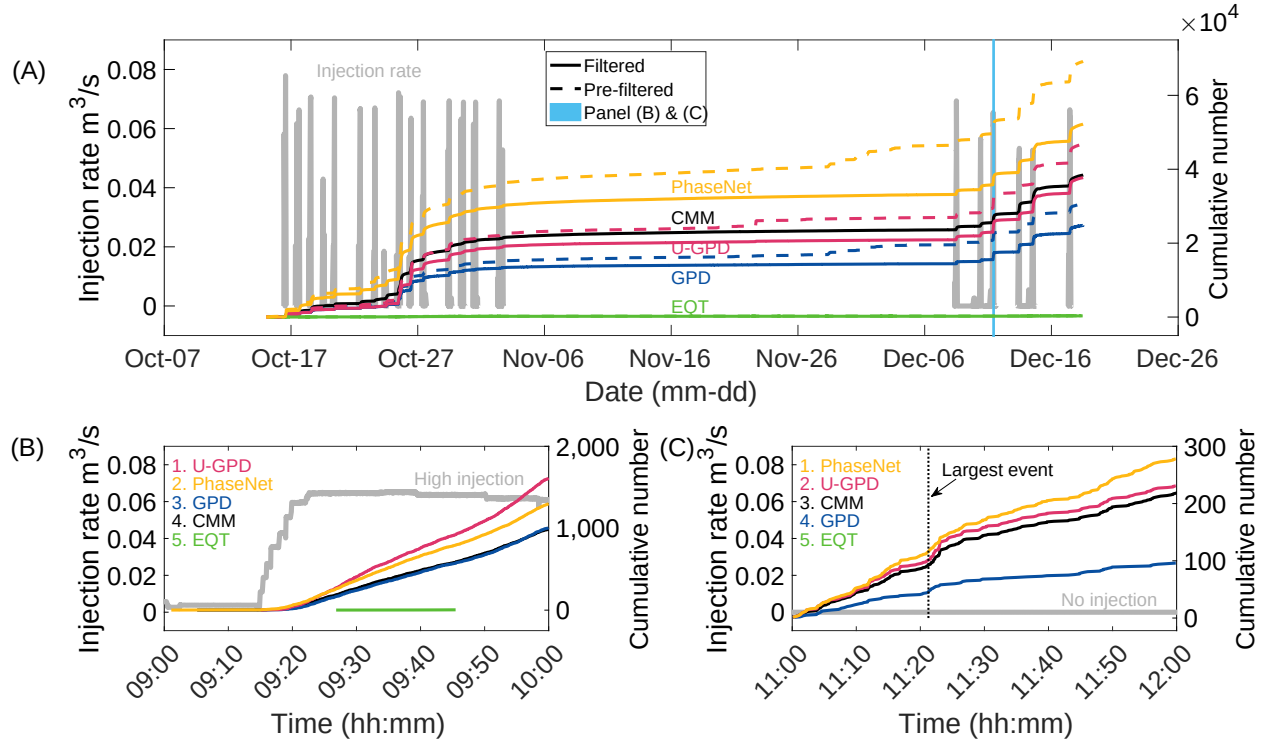


Figure 9. Temporal plots of the cumulative number of model detections (right y-axis) (A) across the PNR-1z continuous dataset with injection rates overlaid (in grey, left y-axis) for PhaseNet (orange), U-GPD (red), CMM (black), GPD (blue) and EQT (green). The solid lines represent the cumulative number of events in the LinMEF-filtered catalogue whereas the dashed lines show the cumulative number of the pre-filtered catalogue. Cumulative number for each model (B) during high injection rate and high event rate, and (C) during no injection and when the largest PNR-1z event occurred. The inset list of models in (B) and (C) represent the performance ranking of the DL models and CMM method.

Table 2. Model runtime statistics for GPD, U-GPD, EQTransformer and PhaseNet on Google Colab, using an NVIDIA Tesla K80 12 GB GPU. These processing times are for one hour of PNR-1z data (2000 Hz), which is equivalent to 20 hours of 100 Hz data. We also include results (recall rate and number of new events) for the selected hour during high injection.

Models	Time taken (24 stations)	$\frac{\text{Data window length}}{\text{Time taken}}$	Recall rate (n events)	Number of new events
GPD	12 min 8.53 s	4.94	87.5% (863)	143 (+14.5%)
U-GPD	4 min 42.37 s	12.75	99.5% (981)	632 (+64.1%)
EQT	20 min 50.19 s	2.88	0.41% (4)	0
PhaseNet	16 min 5.88 s	3.72	91.6% (903)	396 (+40.2%)

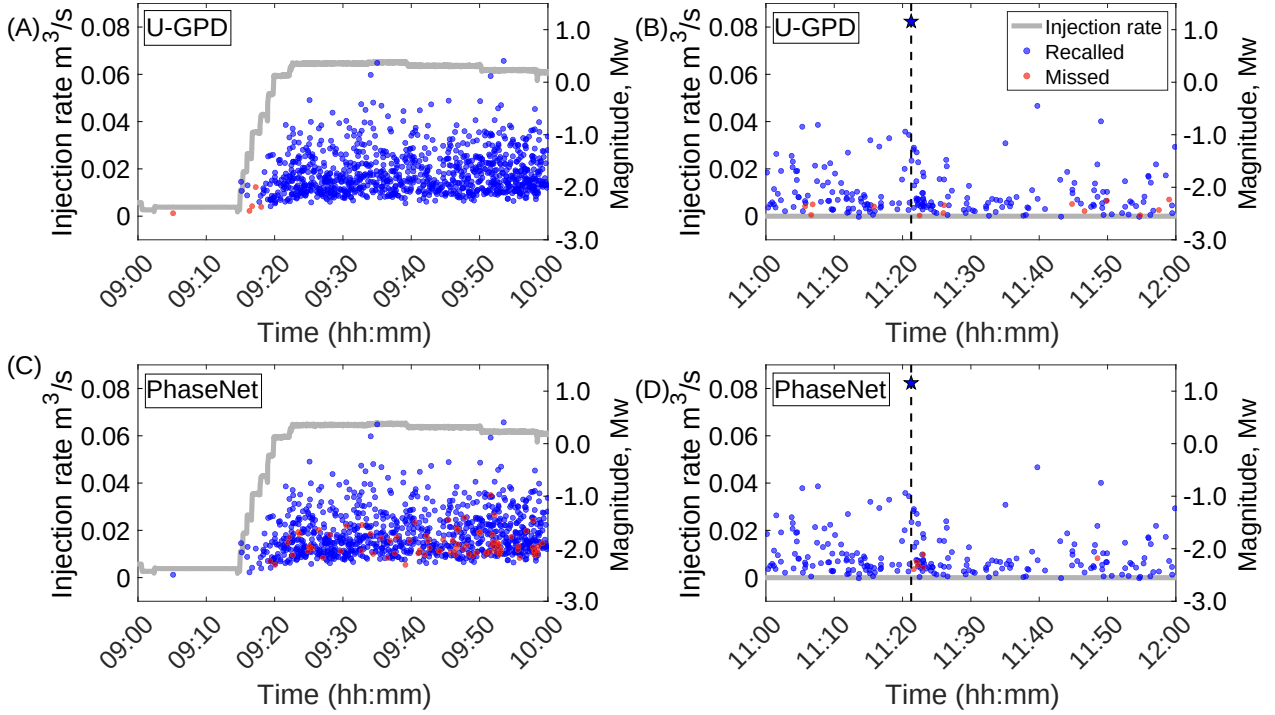


Figure 10. Time-magnitude plots of the high injection rate hour (A, C) and the quieter period (B, D) with injection rates (black line, left axis) for the U-GPD (top row) and PhaseNet (bottom row) models. Missed (red) and recalled (blue) events are plotted with their respective moment magnitudes, M_w (right axis). The largest earthquake (M_w 1.1) is plotted as a star and is denoted with a dashed line.

6 DISCUSSION

The results demonstrate the applicability of GPD, U-GPD and PhaseNet to high frequency borehole data, although to varying degrees. Our study shows that, in particular, PhaseNet outperforms the CMM method, as it recalls up to 95% of the catalogue and detected approximately 15,800 new events. In total, PhaseNet detects over 52,000 target events. Although U-GPD and GPD detect fewer earthquakes than the CMM method, both DL models still identify a significant number of events within the dataset (over 37,800 and over 24,900 respectively). All models had no issue picking phases on amplitude-clipped events.

In contrast, EQT might encounter challenges with its attention mechanism in effectively detecting events within the high frequency borehole data. This issue arises from the data representations learned by the deeper, more complex model and that may not generalise well to higher frequency data. The attention mechanism in EQT, which predicts the arrival of seismic phases after determin-

ing the presence of an earthquake, may face difficulties when the P and S phase arrive at different positions and appear differently in the 2000 Hz data compared to the 100 Hz training data. Consequently, the complex transformer-based EQT model struggles to interpret the data, resulting in its inability to detect events effectively.

The precision of phase picking heavily depends on the network architecture. Models like U-GPD and PhaseNet, which utilise fully convolutional networks, provide more precise picks by leveraging higher resolution continuous probability phase traces. The pick precision of the GPD model is influenced by its combination of convolutional and fully connected neural networks, which output a single probability value per class for each data-window shift. Therefore, in GPD's case, pick precision is significantly dependent on the size of the window shift across the data.

PhaseNet outperforms other models in terms of precision, recall and F1-score for microseismic phase classification. It exhibits high metric scores for both P and S arrivals, while other models show inconsistencies in classifying P and S phases. GPD, U-GPD and EQT have better metric scores for classifying P phases compared to S phases due to S waves characteristically having lower signal-to-noise ratio.

PhaseNet's success on this dataset may be attributed to its exposure to different types of instrument data during training. The model's training incorporated seismic data from several frequency bands and instruments (Zhu & Beroza, 2019), resulting in an apparently improved generalisation for phase picking in the HFIS setting. Furthermore, the relatively small model size of PhaseNet (10 layers, 269,675 parameters) combined with a large training dataset reduces the risk of overfitting to its initial training set (623,054 training samples). Although U-GPD has a similar architecture and model size to PhaseNet (9 layers, 672,419 parameters), it displays signs of overfitting to the small volcanic dataset (in 100 Hz) used to fine-tune the model (Lapins et al., 2021). Overfitting can be mitigated through stronger data augmentation, such as resampling the training data to lower and higher sample rates to vary the number of samples between a P and S phase and expanding

the training dataset to include more diverse events with different P to S time differences. Lastly, as the deepest of all models tested (56 layers, 376,935 parameters), EQT might be susceptible to overfitting because of its number of layers and the lack of data diversity of its original training set (1.3 million samples). The deep transformer-based model might have learned more complex and abstract features that is more tailored to the 100 Hz training data as the number of layers increase (Goodfellow et al., 2016) and therefore, EQT might struggle to recognise any extracted features in higher frequency data.

As observed in Fig. 10, U-GPD performs better at picking events during high seismic rates due to its smaller window size (400 samples) compared to PhaseNet (3000 samples). This enables U-GPD to handle the presence of multiple events within the same temporal data window more effectively. Although PhaseNet struggles (relative to U-GPD) during higher event rates, it still detects more small events ($M_w \leq -2$) than U-GPD overall (Fig. 10).

The limitations of the CMM method become apparent when multiple events overlap within the same user-defined time window (see Fig. S3, Supporting Information). When there is more than one event within this window, CMM detects only the largest event based on energy maxima (Drew et al., 2013), potentially missing smaller events. Reducing the time step window in the CMM method could enhance its performance, but this would increase its computational intensity and overall cost, thereby limiting its feasibility.

In contrast, DL models have the potential to pick out overlapping events (see Fig. S3, Supporting Information). However, our current approach to phase association and event grouping lacks the sophistication needed to differentiate multiple closely arriving events within in a fixed time window. Future research should consider more advanced phase associators such as REAL (Zhang et al., 2019), an optimised grid-search algorithm, GaMMA (Zhu et al., 2022), which considers arrival time moveout and amplitude decay with distance or PyOcto (Münchmeyer, 2023), a 4D space time partitioning algorithm to separate overlapping events in time. These more advanced

techniques could leverage DL-detected phases arriving in close temporal proximity, thereby significantly enhancing event detection.

GPD, U-GPD and PhaseNet might benefit from additional training or exposure through fine-tuning with microseismic datasets ($-3 \leq M_w \leq 0$) to improve their ability to recognise smaller events recorded at higher sampling rates. However, PhaseNet already exhibits promising performance and can be deployed without any modification. The events missed by PhaseNet are primarily small events ($M_w \leq -0.5$) that occur during high seismic rates, where larger events may overshadow the smaller earthquakes. Shortening PhaseNet’s input window length could potentially isolate smaller events for detection, but there would be a trade-off with increased computational time due to the greater number of windows to be processed. Since these models can be employed in near real-time, finding a balance between efficiency and effectiveness is crucial. The efficiency of DL models makes them suitable for automated detection methods in real-time monitoring.

7 CONCLUSIONS

In this study, we compare the ability of four DL models (GPD, U-GPD, EQT and PhaseNet) to detect hydraulic fracturing induced seismicity through phase picking in high frequency (2000 Hz) downhole data. We find that PhaseNet is most suitable for microseismic monitoring applications, when comparing the phase classification abilities of the models and their resulting event detections. PhaseNet shows the best phase classification results and the most precise picking compared to the other models.

Based on our evaluation, we conclude that while additional transfer learning might further enhance the performance of PhaseNet, it can still be readily applied ‘off-the-shelf’ (recall rate of 94.6%) for downhole monitoring of induced seismicity with moment magnitudes M_w greater than -0.5. Models such as GPD and U-GPD (recall rate of 59.5% and 76.6% respectively) require transfer learning to improve their microseismic detection, to detect events with M_w below -0.5. EQT,

on the other hand, currently requires retraining to be applied to high frequency downhole data.

We infer that model architecture and exposure to different training data influences microseismic phase detection and picking precision, which are important for event location and moment magnitude estimation. Network architectures such as fully convolutional U-Nets with a small model size that provide continuous high-resolution probability traces offer more precise phase picking and better generalisation to pick events in high frequency data compared to more complex transformer models (e.g., EQT) or models comprising fully-connected layers (e.g., GPD).

In addition to the over 38,000 catalogued earthquakes already detected on site, PhaseNet identified about 15,800 new events occurring within the PNR-1z dataset. The identification of thousands of additional events could lead to improved observations of the spatio-temporal characteristics of HFIS within the context of various driving mechanisms (e.g., Kettlety & Verdon, 2021; Herrmann et al., 2022) and facilitate the real-time tracking of seismicity during well stimulations.

ACKNOWLEDGMENTS

We thank James Verdon, Joanna Holmgren, Tom Kettlety and Alan Baird for helpful discussions about the PNR-1z operations. This work was carried out using the computational and data storage facilities of the Advanced Computing Research Centre, University of Bristol -<http://www.bristol.ac.uk/acrc/>. C.S.Y. Lim was supported by the U.K. Natural Environment Research Council (NERC) Centre for the Observation and Modelling of Earthquakes, Volcanoes, and Tectonics (COMET), grant number GA/13M/031, a partnership between UK Universities and the British Geological Survey (BGS) University Funding Initiative (BUFI) (S468), grant number GA/20M/004. S. Lapins was funded by a Leverhulme Trust Early Career Fellowship (<https://www.leverhulme.ac.uk/>). M.J. Werner was supported by the Natural Environment Research Council (grant number NE/R017956/1) and the Engineering and Physical Sciences Research Council (grant number EP/Y020960/1). This work is also supported by the Bristol University Microseismicity ProjectS (BUMPS).

DATA AVAILABILITY

All seismic data and microseismic catalogue for PNR-1z operations are publicly available through the North Sea Transition Authority (<https://www.nstauthority.co.uk/exploration-production/onshore/onshore-reports-and-data/preston-new-road-pnr-1z-hydraulic-fracturing-operations-data/>). All code and data to run the classification tests are available at https://github.com/cindylimsy/PNR_run. The model earthquake catalogues for PNR-1z and velocity models are also available in the same repository.

References

- Atkinson, G. M., Eaton, D. W., & Igonin, N., 2020. Developments in understanding seismicity triggered by hydraulic fracturing, *Nature Reviews Earth & Environment*, **1**(5), 264–277.
- Beroza, G. C., Segou, M., & Mostafa Mousavi, S., 2021. Machine learning and earthquake forecasting—next steps, *Nature communications*, **12**(1), 4761.
- Chen, X., Nakata, N., Pennington, C., Haffener, J., Chang, J. C., He, X., Zhan, Z., Ni, S., & Walter, J. I., 2017. The pawnee earthquake as a result of the interplay among injection, faults and foreshocks, *Scientific Reports*, **7**(1), 4945.
- Clarke, H., Verdon, J. P., Kettlety, T., Baird, A. F., & Kendall, J.-M., 2019. Real-time imaging, forecasting, and management of human-induced seismicity at preston new road, lancashire, england, *Seismological Research Letters*, **90**(5), 1902–1915.
- Drew, J., White, R. S., Tilmann, F., & Tarasewicz, J., 2013. Coalescence microseismic mapping, *Geophysical Journal International*, **195**(3), 1773–1785.
- Eaton, D. W., Igonin, N., Poulin, A., Weir, R., Zhang, H., Pellegrino, S., & Rodriguez, G., 2018. Induced seismicity characterization during hydraulic-fracture monitoring with a shallow-wellbore geophone array and broadband sensors, *Seismological Research Letters*, **89**(5), 1641–1651.
- Eyre, T. S., Eaton, D. W., Garagash, D. I., Zecevic, M., Venieri, M., Weir, R., & Lawton, D. C., 2019. The role of aseismic slip in hydraulic fracturing–induced seismicity, *Science advances*, **5**(8), eaav7172.

- Frohlich, C. & Brunt, M., 2013. Two-year survey of earthquakes and injection/production wells in the eagle ford shale, texas, prior to the mw4. 8 20 october 2011 earthquake, *Earth and Planetary Science Letters*, **379**, 56–63.
- Gadallah, M. R. & Fisher, R., 2008. *Exploration geophysics*, Springer Science & Business Media.
- 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.
- Goodfellow, I., Bengio, Y., & Courville, A., 2016. *Deep learning*, MIT press.
- Grigoli, F., Cesca, S., Rinaldi, A. P., Manconi, A., Lopez-Comino, J. A., Clinton, J., Westaway, R., Cauzzi, C., Dahm, T., & Wiemer, S., 2018. The november 2017 m w 5.5 pohang earthquake: A possible case of induced seismicity in south korea, *Science*, **360**(6392), 1003–1006.
- Herrmann, M., Piegari, E., & Marzocchi, W., 2022. Revealing the spatiotemporal complexity of the magnitude distribution and b-value during an earthquake sequence, *Nature Communications*, **13**(1), 5087.
- Holmgren, J. M., Kwiatak, G., & Werner, M., 2023. Nonsystematic rupture directivity of geothermal energy induced microseismicity in helsinki, finland, *Journal of Geophysical Research: Solid Earth*, **128**(3), e2022JB025226.
- Kao, H., Eaton, D., Atkinson, G., Maxwell, S., & Mahani, A. B., 2016. Technical meeting on the traffic light protocols (tlp) for induced seismicity: Summary and recommendations, *Geological Survey of Canada, Open File*, **8075**(20), 15.
- Kettlety, T. & Verdon, J. P., 2021. Fault triggering mechanisms for hydraulic fracturing-induced seismicity from the preston new road, uk case study, *Frontiers in Earth Science*, **9**, 382.
- Klinger, A. G. & Werner, M. J., 2022. Stress drops of hydraulic fracturing induced microseismicity in the horn river basin: Challenges at high frequencies recorded by borehole geophones, *Geophysical Journal International*, **228**(3), 2018–2037.
- 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), e2021JB021910.

- Lei, X., Huang, D., Su, J., Jiang, G., Wang, X., Wang, H., Guo, X., & Fu, H., 2017. Fault reactivation and earthquakes with magnitudes of up to $m_w 4.7$ induced by shale-gas hydraulic fracturing in sichuan basin, china, *Scientific reports*, **7**(1), 1–12.
- Li, L., Tan, J., Wood, D. A., Zhao, Z., Becker, D., Lyu, Q., Shu, B., & Chen, H., 2019. A review of the current status of induced seismicity monitoring for hydraulic fracturing in unconventional tight oil and gas reservoirs, *Fuel*, **242**, 195–210.
- Lomax, A., Virieux, J., Volant, P., & Berge-Thierry, C., 2000. Probabilistic earthquake location in 3d and layered models, in *Advances in seismic event location*, pp. 101–134, Springer.
- Mancini, S., Segou, M., Werner, M. J., Parsons, T., Beroza, G., & Chiaraluce, L., 2022. On the use of high-resolution and deep-learning seismic catalogs for short-term earthquake forecasts: Potential benefits and current limitations, *Journal of Geophysical Research: Solid Earth*, **127**(11), e2022JB025202.
- 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), 1–12.
- Münchmeyer, J., 2023. Pyocto: A high-throughput seismic phase associator, *arXiv preprint arXiv:2310.11157*.
- Münchmeyer, J., Woollam, J., Rietbrock, A., Tilmann, F., Lange, D., Bornstein, T., Diehl, T., Giunchi, C., Haslinger, F., Jozinović, D., 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), e2021JB023499.
- Perol, T., Gharbi, M., & Denolle, M., 2018. Convolutional neural network for earthquake detection and location, *Science Advances*, **4**(2), e1700578.
- Ross, Z. E. & Cochran, E. S., 2021. Evidence for latent crustal fluid injection transients in southern california from long-duration earthquake swarms, *Geophysical Research Letters*, **48**(12), e2021GL092465.
- Ross, Z. E., Meier, M.-A., Hauksson, E., & Heaton, T. H., 2018. Generalized seismic phase detection with deep learning, *Bulletin of the Seismological Society of America*, **108**(5A), 2894–

2901.

Schultz, R., Wang, R., Gu, Y. J., Haug, K., & Atkinson, G., 2017. A seismological overview of the induced earthquakes in the duvernay play near fox creek, alberta, *Journal of Geophysical Research: Solid Earth*, **122**(1), 492–505.

Schultz, R., Skoumal, R. J., Brudzinski, M. R., Eaton, D., Baptie, B., & Ellsworth, W., 2020. Hydraulic fracturing-induced seismicity, *Reviews of Geophysics*, **58**(3), e2019RG000695.

Sheriff, R. E., 2002. *Encyclopedic dictionary of applied geophysics*, Society of exploration geophysicists.

Shi, P., Grigoli, F., Lanza, F., & Wiemer, S., 2022. Malmi: towards combining machine learning and waveform migration for fully automated earthquake detection and location, in *EGU General Assembly Conference Abstracts*, pp. EGU22–6087.

Smith, E., Smith, A., White, R., Brisbourne, A., & Pritchard, H., 2015. Mapping the ice-bed interface characteristics of rutford ice stream, west antarctica, using microseismicity, *Journal of Geophysical Research: Earth Surface*, **120**(9), 1881–1894.

Tan, Y. J., Waldhauser, F., Ellsworth, W. L., Zhang, M., Zhu, W., Michele, M., Chiaraluce, L., Beroza, G. C., & Segou, M., 2021. Machine-learning-based high-resolution earthquake catalog reveals how complex fault structures were activated during the 2016–2017 central italy sequence, *The Seismic Record*, **1**(1), 11–19.

Verdon, J. P. & Budge, J., 2018. Examining the capability of statistical models to mitigate induced seismicity during hydraulic fracturing of shale gas reservoirs, *Bulletin of the Seismological Society of America*, **108**(2), 690–701.

Verdon, J. P. & Stork, A. L., 2016. Carbon capture and storage, geomechanics and induced seismic activity, *Journal of Rock Mechanics and Geotechnical Engineering*, **8**(6), 928–935.

Wong, I., Nemser, E., Bott, J., & Dober, M., 2015. Induced seismicity and traffic light systems as related to hydraulic fracturing in ohio.

Zhang, M., Ellsworth, W. L., & Beroza, G. C., 2019. Rapid earthquake association and location, *Seismological Research Letters*, **90**(6), 2276–2284.

Zhu, W. & Beroza, G. C., 2019. Phasenet: a deep-neural-network-based seismic arrival-time

picking method, *Geophysical Journal International*, **216**(1), 261–273.

Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., & Beroza, G. C., 2022. Earthquake phase association using a bayesian gaussian mixture model, *Journal of Geophysical Research: Solid Earth*, **127**(5), e2021JB023249.

8 FOR EDITORS

This paper has been produced using the Blackwell Scientific Publications GJI L^AT_EX2e class file.