Geophysical Journal International

Geophys. J. Int. (2026) **247**, 1–18
Advance Access publication 2026 August 17
Research Paper

https://doi.org/10.1093/gji/ggag326

Testing a multistation deep learning seismic phase picker trained on surface array data in a high frequency downhole setting

H.J. Burnett,[1,2] C.S.Y. Lim ,[1] M.J. Werner [1] and J.P. Verdon[1]

[1] *School of Earth Sciences, University of Bristol, Bristol BS8 1RJ, UK. E-mail: cindy.lim@bristol.ac.uk*
[2] *Scripps Institution of Oceanography, UC San Diego, La Jolla, CA 92037, USA*

Accepted 2026 August 13. Received 2026 July 28; in original form 2025 November 21

SUMMARY

Induced seismicity can damage infrastructure, cause injury and generate negative public perceptions for geo-energy operations. To monitor, forecast and understand induced earthquake physics, phase picking models must accurately detect and classify microearthquake signals. Deep learning (DL) algorithms can detect and classify seismic signals to the accuracy of a human analyst in real-time. Recent advances have introduced multistation DL algorithms which leverage 2-D spatial information to improve the detection of low signal-to-noise ratio arrivals. However, most such models are trained on surface networks, and applying them to borehole arrays which differ in sampling rate, station geometry and waveform characteristics, remains an open challenge. Here, we develop input adaptation strategies that allow a surface-trained multistation DL picker to be applied effectively to borehole data, without retraining. Our central finding is that rotating the input traces so that the observed *P*-and *S*-wave energies align with the input channels the model has learned during training substantially improves detection. We demonstrate this by applying CubeNet, a multistation DL model, to detect induced seismicity at the Preston New Road 2 (PNR2) hydraulic fracturing site, UK using a borehole array and an existing catalogue. Using station geometry information and rotating the borehole traces to mimic surface traces increases successful *P* and *S* picks by 9.5 per cent points and 13 per cent points, respectively, for $-1.5 \leq M_w \leq 1.2$. We also develop a workflow to explore how the increases in the number of picks affect the quality and number of phase associations and located earthquakes. By rotating the borehole traces, we show a 65 per cent increase in the number of earthquakes located in a 1-hr period (203 to 335 events). Moreover, CubeNet can operate 1.86 times faster than real-time data recording and our relatively computationally efficient workflow can reproduce previous location estimates. We recall 100 per cent of earthquakes above M_w -0.2, but only 18 per cent of the existing catalogue because just $\sim$16 per cent of CubeNet's original training data set consisted of events with $M_w < -0.2$. We conclude that these input adaptation strategies can be implemented to other surface-trained multistation DL pickers applied to borehole arrays. Our proposed methods offer a practical route to deploying existing models on new borehole geometries without retraining. While CubeNet proves useful for near real-time monitoring of larger events, detecting smaller ($M_w < -0.2$) events will require transfer learning on low signal-to-noise ratio events, or indeed the development of bespoke models for downhole data sets.

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

1 INTRODUCTION

Downhole microseismic monitoring can provide high resolution microseismic observations of industrial activities like shale gas exploration, carbon capture and storage (CCS), and enhanced geothermal. Microseismicity can be used to forecast incipient larger earthquakes (J.P. Verdon & J. Budge 2018; S. Mancini *et al.* 2021) and their hazards (e.g. G. Cremen & M.J. Werner 2020), can provide important insights into earthquake physics (e.g. T. Kettlety *et al.* 2019, 2020; N. Igonin *et al.* 2022; J.M. Holmgren *et al.* 2023) and can be used to track fluid propagation (N. Igonin *et al.* 2021). By deploying stations close to the source and sampling with high frequencies, one can obtain unaliased observations of microseismicity that are less affected by path and site effects than surface recordings. Common applications for downhole microseismic arrays have included monitoring hydraulic fracture propagation (e.g. J.T. Rutledge *et al.* 2004), monitoring caprock integrity at CCS sites (e.g. J.P. Verdon *et al.* 2011; B.D.E. Dando *et al.* 2021) and observing geomechanical deformation in mature conventional hydrocarbon reservoirs (e.g. A. Al-Anboori & J.M. Kendall 2010).

Sampling at high frequencies generates large volumes of data that needs to be processed in real-time to aid forecasting models and decision making frameworks. Deep learning methods have substantial potential for rapid processing of downhole microseismic data.

A branch of supervised classification algorithms, named convolutional neural networks (CNNs), can be trained to perform seismic phase detection and picking (e.g. T. Perol *et al.* 2018; Z.E. Ross *et al.* 2018; W. Zhu & G. C. Beroza 2019; S.M. Mousavi & G.C. Beroza 2020; S. Lapins *et al.* 2021).

CNNs require extensive training using hundreds of thousands of labelled waveforms: CNN performance can grow logarithmically with the volume of training data (C. Sun *et al.* 2017). Further, the training data should be diverse so the CNN can generalize when applied to unseen data (S. Lapins 2021). In seismology, one can increase the training data diversity by varying the earthquake source mechanism, magnitude, array geometry, recording device type and the signal frequency. However, downhole microseismic data is often acquired at great expense, often at commercially sensitive industrial sites, and hence most data is not openly available: there is a limited availability of downhole microseismic data sets for model training. Therefore it is important to understand whether models trained on surface data can be applied to borehole data sets. This will influence model selection and future training data sets.

Downhole stations are typically deployed in close proximity to one another as a geophone array or 'string' and so will record coherent *P* and *S* wavefronts from a single event. Widely used single-station phase pickers such as PhaseNet (W. Zhu & G. C. Beroza 2019) and EQTransformer (S.M. Mousavi & G.C. Beroza 2020) are computationally lighter but do not exploit this interstation coherence, which can help discriminate low-SNR arrivals. Without this, single-station detections are prone to false positives (C.S. Lim *et al.* 2025) that must be filtered using phase move-out patterns or during phase association. Hence, we may be able to enhance detection and picking by using a multistation approach which considers the spatial distribution of stations (G. Chen & J. Li 2022) and looks for coherent arrivals between stations. Multistation methods treat detection and picking as a 3-D problem. Some multistation machine learning pickers use Graph Neural Networks (GNNs) (T. Feng *et al.* 2022). While these can better account for spatial variations than CNNs, GNNs require re-training when applied to different data (M.P. van den Ende & J.P. Ampuero 2020; I.W. McBrearty & G.C. Beroza 2025). G. Chen & J. Li (2022) developed a CNN approach for multistation phase picking that does not require re-training for different network geometries. More recently, Neural Operators have also found success in multistation picking (H. Sun *et al.* 2023) and are flexible in the sense that they can adjust to new network configurations without re-training. However, these approaches have primarily been developed for surface seismic networks, whereas downhole monitoring poses additional challenges that require deep learning methods capable of generalizing across diverse array geometries.

Here, we propose different input adaptation strategies to optimize the application of a multistation DL models to detect and pick HF induced earthquakes using a downhole array at the Preston New Road-2 (PNR2) HF site, UK (H. Clarke *et al.* 2019; T. Kettlety *et al.* 2021). Multistation CNNs have shown promise and including wavefront coherence may be vital to detect low signal-to-noise ratio (SNR) arrivals on downhole arrays. However, these methods are novel and have not been tested to the same extent as is now the case for single station CNN-based pickers (C.S. Lim *et al.* 2025).

In this study, we first test CubeNet (G. Chen & J. Li 2022), a multistation DL algorithm using the standard east-north-vertical (ENZ) coordinate frame to determine its classification accuracy when applied off-the-shelf to downhole microseismic data. When monitoring an industrial site in real-time, no pre-existing catalogue will be available, and so algorithms must perform well without site-specific training. Particularly, we assess whether CubeNet can accurately pick phases on waveforms that have a much higher sampling frequency, 2000 Hz, than the training data, 100–500 Hz. Next, consider the differences between the training and application data set: we adapt the input data by varying the coordinate frame and spatial regularization and evaluate how this affects the number of phase picks and the picking accuracy produced by CubeNet. We then incorporate CubeNet into a four step workflow to detect and associate phases and locate microseismic events. This workflow is designed for a linear borehole array with a low azimuthal coverage. We apply this workflow to a subset of the PNR-2 data set and compare it to a pre-existing event detection method. We also examine the computational efficiency of our workflow. Finally, we discuss CubeNet's suitability for application to real-time processing of downhole microseismic data sets at industrial sites.

2 DATA

2.1 Preston new road 2 hydraulic fracturing site

The study uses time-series data recorded by a borehole array monitoring HF operations at the PNR2 Site, UK (T. Kettlety *et al.* 2021). Here continuous waveforms were recorded by twelve 3-component borehole instruments sampling at 2000 Hz for a deployment period of approximately 50 d (2019 August 13–October 2). The UK government requires that injected fluids do not escape a pre-defined region (Cuadrilla 2018), and so the primary objective of the downhole array was to map the growth of hydraulic fractures to ensure they remained within the permitted limits. A high sampling frequency of 2000 Hz was therefore used to detect small microseismic events that may reveal fracture growth and fluid pathways. The borehole array, consisting of eleven geophones and one accelerometer, was deployed in the build section of a neighbouring well (PNR1) and has a spatial aperture of 300 m (T. Kettlety *et al.* 2021). A processing contractor performed real-time analysis of the data, producing a catalogue of approximately 54 000 detected and located microseismic events (Fig. 1), which provides a catalogue for testing and comparison (T. Kettlety *et al.* 2021). The seismicity forms three clusters which extend parallel to the maximum horizontal stress direction, tracking the growth of hydraulic fractures away from the well.

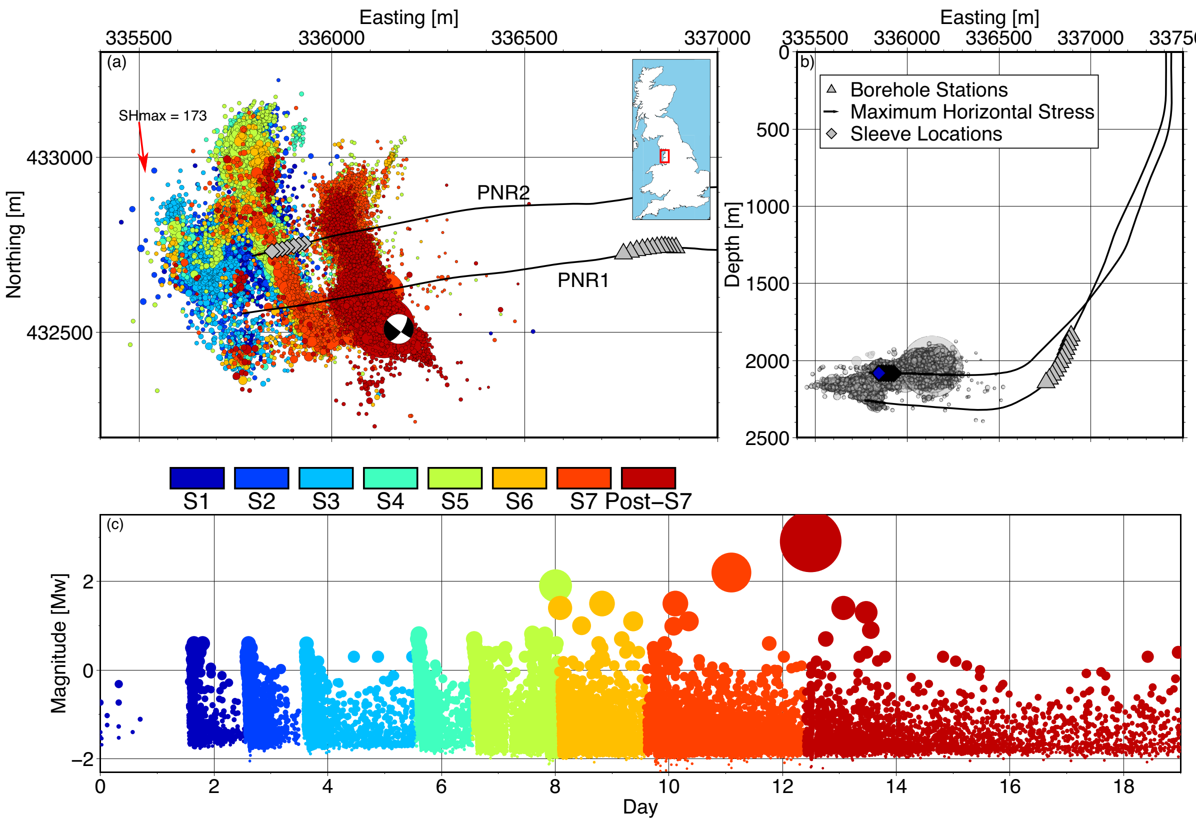

Figure 1. Spatiotemporal distribution of seismicity at PNR2. **(a)** Map view of seismicity. Circular markers denote events. Marker size scales with magnitude and colour indicates the injection stage when the event occurred: the locations of the seven injection stages are shown as grey diamonds. The positions of geophones in the monitoring array are marked by triangles. The focal mechanism solution for largest event and maximum horizontal stress direction (S_{Hmax}) from T. Kettlety *et al.* (2021) are shown. Black lines show the two wells (PNR1 and PNR2). A smaller box in the inset shows the site location in the UK. **(b)** Cross-section of seismicity and wells—in this panel the injection points are coloured according to injection stage. **(c)** Temporal evolution of magnitudes: colours show injection stage and marker size scales with magnitude. Figure inspired by T. Kettlety *et al.* (2021) and made using the Generic Mapping Tools (P. Wessel et al. 2019).

Seven injection stages were carried out before earthquake magnitudes became too high and operations were terminated. A M_{L} 2.9 strike slip event occurred approximately three days after the last injection stage on a reactivated fault plane (T. Kettlety *et al.* 2021). Fig. 2 shows example waveforms for two microseismic events recorded on the eleven geophones.

The processing contractor used the Coalescence Microseismic Mapping method (CMM) to produce the existing catalogue (J. Drew *et al.* 2013). CMM performs the tasks of detection and picking, association and location within a single algorithm and has three steps. Traveltime functions are established between a 3-D grid of nodes (in the subsurface) and each station in the array. Next, STA/LTA is used for detection and the STA/LTA function quantifies the error in the pick. For each cluster of picks, binned in time, the method backpropagates from each station to every node in the grid. The model outputs probability density functions (PDFs) in space and time which delineate the most probable hypocentre location and origin times. A supercomputer was required to run the CMM procedure in real-time. Therefore, CMM was able to detect a large number of small microseismic events, but it is computationally expensive.

2.2 Data pre-processing

The instruments were lowered into the well on a wireline. The tools are free to rotate during this process. As such, one axis of each 3-component geophone will be parallel to the wellbore axis, but the orientation of the other two components will be arbitrary, and different for each downhole geophone. Weight drop experiments at known locations were used to identify the orientation of each tool, enabling us to rotate the traces into the east-north-vertical (ENZ) coordinate frame. The rotation procedure was validated by applying a principal component analysis (PCA) to delineate particle motions. We observe that many traces of earthquakes with magnitude above $M_{\mathrm{w}} = 1.2$ were clipped. Station PNR008 is an accelerometer: we integrated the accelerometer traces using a cumulative trapezoid integration to convert to particle velocity but this produced unsuitable amplitude estimates. To avoid bias during multistation phase picking, where picking is dependent on neighbouring stations, we do not include traces from PNR008 in our analyses.

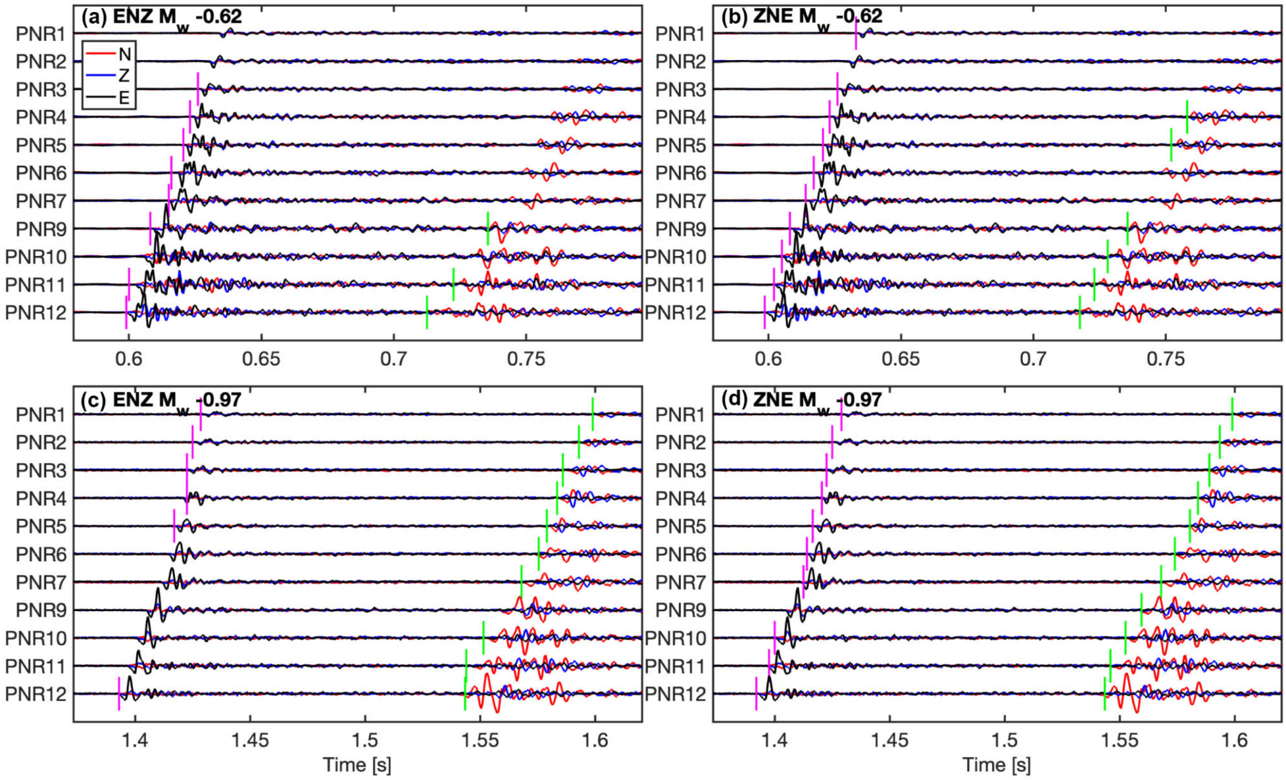

Figure 2. Waveforms for two microseismic events with magnitudes of M_w −0.62 (a,b) and M_w −0.97 (c,d). The CubeNet picks (magenta = *P*, green = *S*) are shown. In (a) and (c) we show the picks when the traces are input in the ENZ coordinate frame, in (b) and (d) we show the picks when traces are input in the ZNE frame.

3 CUBENET

3.1 CubeNet model architecture

We use CubeNet, a multistation DL algorithm, to detect and pick earthquake signals in both space and time (G. Chen & J. Li 2022). By searching for coherent arrivals across multiple waveforms from different stations, the algorithm aims to pick *P* and *S* arrivals with more certainty. The algorithm requires 3C traces and 2-D spatial data (latitude and longitude of stations). CNNs can only process data with a regular structure (G. Chen & J. Li 2022). Thus, the station locations are regularized, or spatially binned, into an eight-by-eight grid. To do this, the station coordinates are centred, rotated to maximize the number of occupied grid cells and then min-max normalized before being binned into the grid (G. Chen & J. Li 2022). When more than one station falls within the same grid cell, multiple data cubes are created so that each cube contains only one station per cell, and the output with the highest peak probability is retained for each station (G. Chen & J. Li 2022). During station regularization, when there are no stations assigned to a grid, that grid is then filled with zeros (G. Chen & J. Li 2022). CubeNet has a U-Net structure: The first half of the U-Net performs feature extraction and downsampling. During feature extraction, the algorithm uses 1-D convolutions to detect signals on each trace and uses 3-D convolutions to detect a wavefront across the array. The second half of the U-Net uses semantic segmentation to classify each data point in the waveform. The model outputs three probability density functions (PDFs) for each station with the same length as the input data. The PDFs quantify the probability of a *P* arrival, *S* arrival or noise for each data point in the input-waveform.

3.2 CubeNet training

To fully understand CubeNet's behaviour when picking the PNR-2 data, it is worth considering the differences between PNR2 and the training data set (Table 1). CubeNet was trained using approximately 151 000 earthquake traces from 4400 microseismic events and 93 500 noise traces (G. Chen & J. Li 2022). These data were recorded on 187 3C nodal stations divided into three polygonal arrays of varying aperture. 80 per cent of the training data were sampled at 100 Hz while the remaining 20 per cent were sampled randomly between 100 and 500 Hz. CubeNet was trained using HF induced seismicity with a magnitude range of the $-1 \leq M \leq 3$, which spans most of the upper magnitude range at the PNR2 site ($-2.5 \leq M \leq 2.9$), but does not include the smallest events $M < 1$.

Table 1. Summary of CubeNet's training data set and the PNR-2 data set.

Data set	CubeNet training data	PNR-2 data
Sampling rates	100–500 Hz	2000 Hz
Array type	Surface array	Linear borehole array
Context	Hydraulic-fracturing (shale gas)	Hydraulic-fracturing (shale gas)
Stations	187 (3C nodal)	11 (3C geophones)
Magnitude range	$-1 \leq M \leq 3$	$-2.5 \leq M \leq 2.9$
Events	4400	54 000

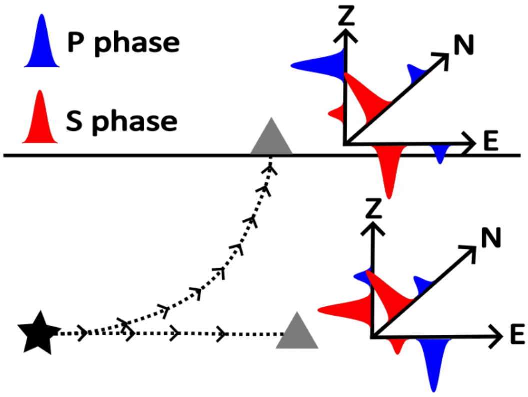

Figure 3. Schematic demonstrating the differences in *P*- and *S*-wave polarizations recorded on borehole and surface instruments. Near-surface ray-bending means that for surface stations the *P* wave will be polarized vertically while the *S* waves are horizontal. For a downhole station at a similar depth to the earthquake, the *P* wave will be polarized horizontally, while the *S*-wave will have vertical and horizontal components.

CubeNet was trained using surface array data (G. Chen & J. Li 2022). Near surface ray bending will strongly influence the incidence angles of seismic waves arriving at surface stations (P. Shearer 2019). In contrast, we are applying CubeNet to a linear borehole array which is at a similar depth to the seismicity. Fig. 3 illustrates how these differences will affect waveform polarization. A waveform recorded at a surface array will have *P* arrivals on the vertical components and S_{H} and S_{V} arrivals on the horizontal components. Given the locations of events to the west of the array, waveforms recorded at PNR2 will have *P* arrivals on the E–W component, S_{H} arrivals on the N–S component and S_{V} arrivals on the vertical component.

PNR2 has a much higher sampling frequency than the training data (2000 Hz versus 100–500 Hz). This is significant because CNNs do not consider the sampling interval between data points. However, CubeNet has been trained on data of varying sampling frequency which may allow it to generalize (G. Chen & J. Li 2022). Lastly, a linear near-vertical borehole array may be non-optimal for the CubeNet regularization, which was designed for a 2-D surface array.

4 MODEL CLASSIFICATION TESTS

4.1 Workflow to test CubeNet

Before applying CubeNet to continuous data, we test the model's performance on known events. We randomly select 25 events from the pre-existing catalogue, in each of three different magnitude bands: $0.5 \leq M_{\mathrm{w}} \leq 1.2$ (CAT1), $-0.5 \leq M_{\mathrm{w}} \leq 0.5$ (CAT2) and $-1.5 \leq M_{\mathrm{w}} \leq -0.5$ (CAT3). We use a maximum magnitude of M_{w} 1.2 because events above this magnitude are clipped.

We rotate the traces to the East-North-Vertical (ENZ) coordinate system and detrend the time-series. The waveforms are windowed from 0.6 s before the origin time to 2.4 s after. We manually check each trace and remove unsuitable records. We manually pick *P* and *S* arrivals for each trace of each earthquake (1650 phase picks for all sub-catalogues) which act as the ground truth. In all of the tests described below, these data were not filtered. We initially tested a 60 Hz highpass filter but found that the increase in classification accuracy was negligible.

After pre-processing and manual picking, we input the traces and station locations to CubeNet. We apply the model on the 2000 Hz data without re-sampling to lower sampling rates. The traces have a length of 6000 data points, which corresponds to 3 s of 2000 Hz data. To analyse the output PDFs, we follow G. Chen & J. Li (2022) and define a probability threshold of 0.3 for a positive detection and assume that the maximum probability corresponds to the best-fitting arrival time. We extract the time of the best pick and the maximum probability from each *P* and *S* PDF.

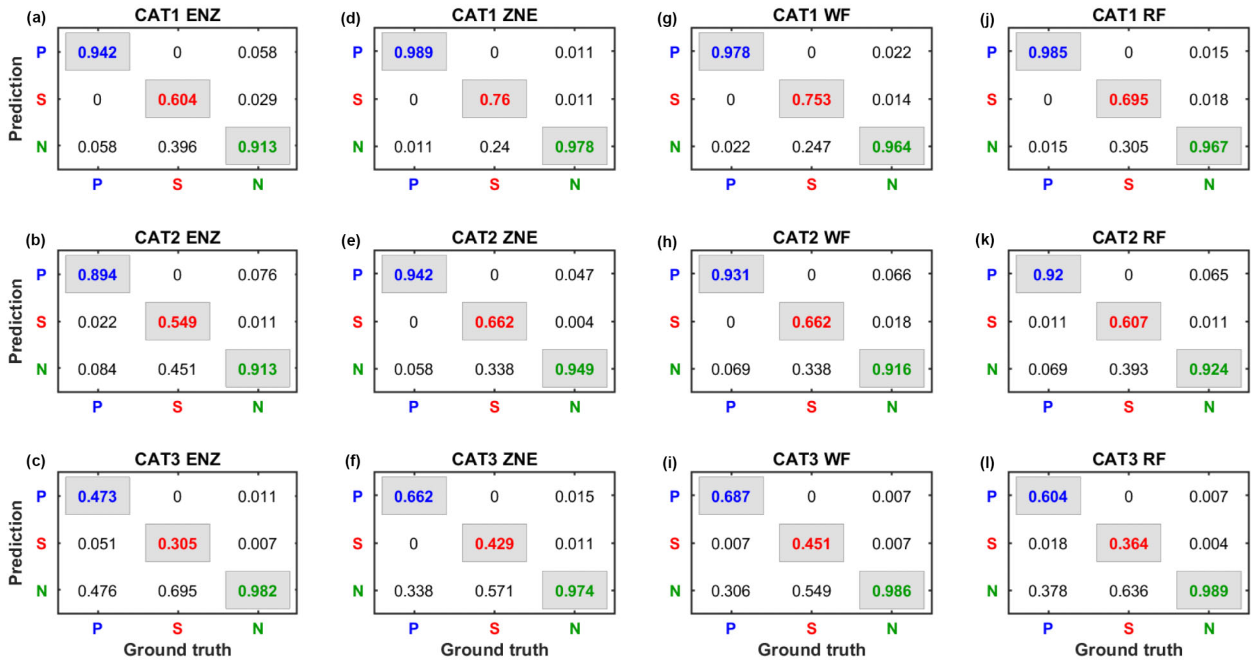

Figure 4. Confusion matrices for ENZ **(a–c)**, ZNE **(d–f)**, WF **(g–i)** and RF (**j–l**) rotation frames for CAT1, CAT2 and CAT3 magnitude events. Values are given as a proportion of the total number of phases for each test.

We compare the model and ground truth picks. We were able to manually pick *P* and *S* arrivals for every waveform. We require that a successful model pick must have an absolute misfit lower than 0.005 s with the ground truth pick. This value was chosen to span 10 data points which has been used in other studies for comparing residuals (e.g. W. Zhu & G. C. Beroza 2019, define a true positive pick with an absolute residual less than 0.1 s for 100 Hz data). The procedure uses a priority system whereby pick times are initially compared to the pick times of the same phase. However, if unsuccessful, picks are then compared to the opposite phase. This accounts for *P* and *S* phase misclassification. If the pick times do not match, then the model predicted noise where the ground truth pick should have been, or has incorrectly picked some part of the *P* or *S*-wave coda. Further, if no model pick was made, the model predicted noise. Parts of the time-series more than 0.005 s from the true phase arrival are classed as true noise for the purpose of these tests. The results are presented in a confusion matrix. In total, 25 events with *P* picks, *S* picks and Noise at 11 stations leads to 825 tests for each confusion matrix.

4.2 Model performance in the ENZ frame

Confusion matrices can be used to gauge the performance of supervised classification algorithms (Fig. 4). The columns of the matrix represent the ground truth phase picks and the rows represent the model predicted phase picks. The matrix values are given as a proportion of the total ground truth picks for each phase (*P*, *S* and *N*). A perfect model, which classifies all picks correctly, would have diagonal elements of one and non-diagonal elements of zero.

Results for the initial test (a–c in Fig. 4) demonstrate that the model can successfully pick some *P* and *S* phases within 0.005 s of the ground truth pick. At large magnitudes (CAT1) the model performs best and successfully classifies 94.2 per cent of *P* picks and 60.4 per cent of *S* picks. As magnitude decreases, the number of accurate *P* and *S* picks decreases and in CAT3 the model successfully classifies 47.3 per cent of *P* picks and 30.5 per cent of *S* picks. There are few instances of inaccurate picks where the model makes a pick which is labelled as noise in the ground truth. This happens in less than 8 per cent of cases for each phase in all three catalogues. Moreover, the model rarely misclassifies these data (labels a ground truth *P* pick as an *S* or vice versa). Misclassification does not occur at large magnitudes but 5.1 per cent of *P* picks in CAT3 were misclassified as *S* picks. For all magnitudes, the model successfully classifies more *P* picks than *S* picks. Moreover, more than 91 per cent of noise is successfully classified in all three catalogues. The higher noise recall for CAT3 (98.2 per cent) compared to CAT1 and CAT2 (both 91.3 per cent) is likely due to the model failing to pick weak arrivals of events with low SNR. This observation is consistent with the sharp drop in *P* and *S* classification accuracy in CAT3 and the corresponding increase in true *P* and *S* arrivals being labelled as noise. We also generated confusion matrices for 300 waveforms containing pure noise. These noise windows were selected from times in the recording where the CMM method detected no events, and manually checked to ensure that no microseismic events were present. The model has perfect classification accuracy and labels all data as noise.

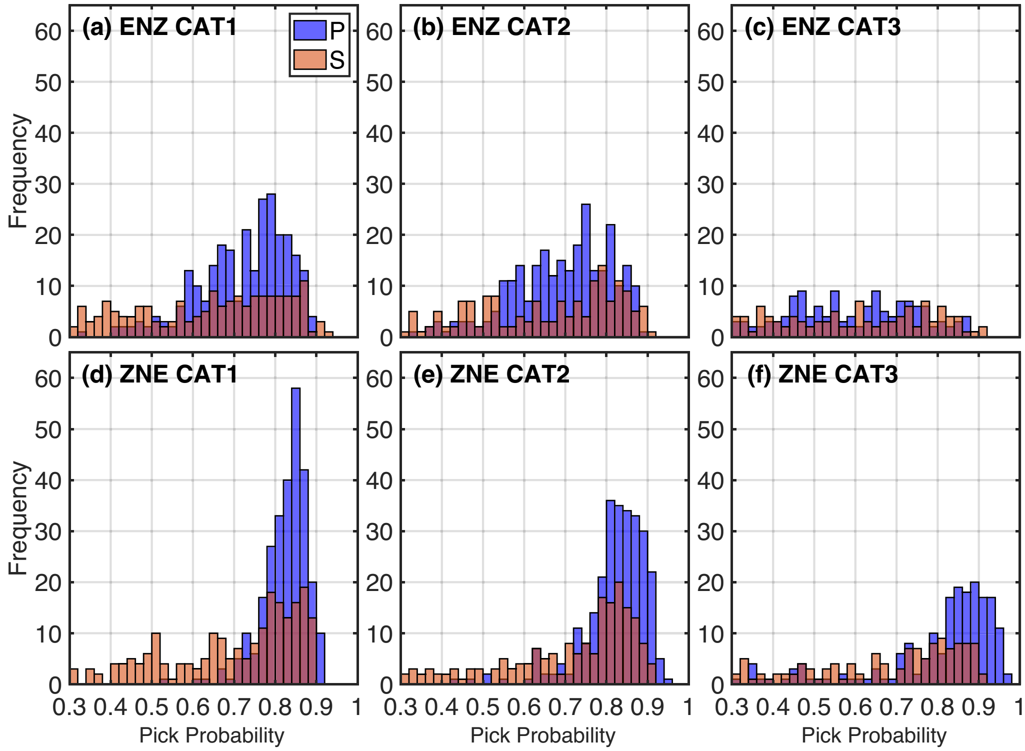

Figure 5. Distribution of probability values for signals classified as *P* or *S* by CubeNet. We show results for the ENZ **(a–c)** and ZNE **(d–f)** coordinate frames. Histograms for CAT1, CAT2 and CAT3 shown in left, middle and right column, respectively.

5 MODEL SENSITIVITY TO ROTATION FRAME AND REGULARIZATION

5.1 Rotation frames

To optimize performance for multistation models, we propose rotating the waveform data into three new coordinate frames to account for differences in the incidence angles of seismic waves arriving at surface and borehole stations. We apply the rotations, input the rotated traces into CubeNet and generate confusion matrices as described in Section 4.1. To reiterate, seismicity at PNR2 originates to the west of the array and at similar depths: the most energetic *P*-wave arrivals will be recorded on the E-W component and the most energetic *S*-wave arrivals will be recorded on the N-S and Z component (see Fig. 3). The goal of these rotations is to adjust the traces to more closely resemble the traces used in training in order to improve model performance. CubeNet expects the E, N and Z component to be input into channels 1, 2 and 3, respectively (defined as ENZ in this study) (G. Chen & J. Li 2022). First we test the ZNE frame whereby we input Z traces to channel 1, N-S traces to channel 2, and E-W traces to channel 3.

Next, we test the effect of rotating the waveform data about the position of the injection point so one component is radial and two components are transverse. We define this rotation frame as WF (well frac). We input the radial component to channel 3 of CubeNet and we input the two transverse components to channels 1 and 2. Finally, we rotate each trace to the ray frame. Here, one component is radial to the earthquake hypocentre (‘P’) and two components are transverse (‘S_H’ and ‘S_V’). We input the ‘P’ component trace to channel 3 whilst the ‘S_V’ and ‘S_H’ components are input to channel 1 and 2, respectively. We define this frame as RF. The rotation angles for the RF orientation are estimated from PCA of the *P*-wave particle motion (e.g. K. De Meersman *et al.* 2006). Note that the RF rotation is only possible when post-processing known events, since rotation angles to an event hypocentre cannot be known until an event is identified.

Results for the ZNE frame show a significant increase in the number of successful *P* and *S* classifications for all magnitudes when compared to the ENZ results (see Figs 4d–f). There is a mean increase in the number of successful *P* and *S* picks of 9.5 per cent points and 13 per cent points, respectively. The greatest increase in the number of *P* picks is seen in CAT3 with an 18.9 per cent point increase and the greatest increase in *S* arrival picking is seen in CAT1 with a 15.6 per cent point increase. There are no instances of *P* or *S* misclassification in any catalogues for the ZNE frame and there is a reduction in the number of inaccurate *P* and *S* picks. Moreover, the algorithm picks *P* and *S* arrivals with more certainty: Fig. 5 shows the distribution of pick probability values in the ENZ and ZNE frames. Probability values of the ZNE picks are skewed towards higher probabilities than the values of the ENZ picks, implying greater picking confidence. As an example of improved picking performance, Fig. 2 shows phase picks made on waveforms of two events in

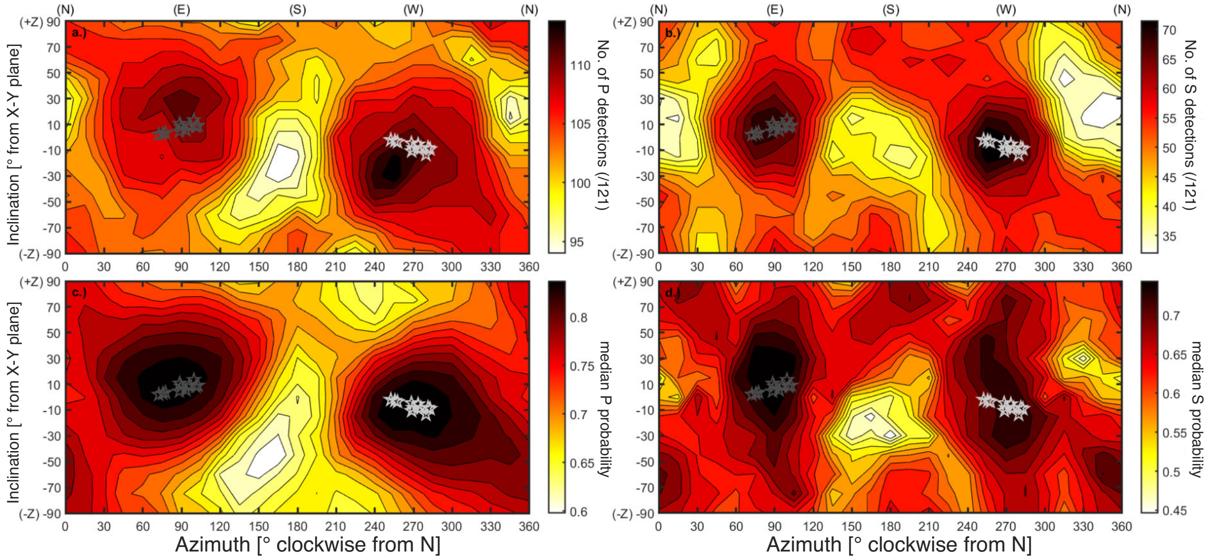

Figure 6. Model performance with arbitrary rotational frames. Azimuths and inclinations give the rotated orientation of the component entered to channel 3 of CubeNet (expects 'vertical' component data). Components in channel 1 and 2 are orthogonal, with one component always being horizontal. The contours in each panel show the numbers of detected *P* arrivals **(a)**, the numbers of detected *S* arrivals **(b)**, the median *P* pick probability **(c)** and the median *S* pick probability **(d)**, as a function of rotation azimuth and inclination. White stars show azimuths and inclinations from the centre of the array to each of eleven test events. Dull grey stars show azimuths and inclinations with a phase shift of 180 deg and inclination with reversed polarity about the X-Y plane.

the ENZ and ZNE rotation frames. For Event 1 (a,b), the ENZ frame results in eight *P* picks and three *S* picks whereas the ZNE frame results in ten *P* picks and six *S* picks. For Event 2 (c,d), the ENZ frame results in six *P* picks and ten *S* picks whereas the ZNE frame results in ten *P* picks and eleven *S* picks.

We observe similar results when using the WF and RF frame rotations (Fig. 4). Results in the WF frame show that this rotation scheme also significantly improves the classification accuracy of the algorithm compared to the ENZ frame (Figs 4g–i). Notably, we observe the largest number of successful *P* and *S* picks for CAT3 in the WF frame. Results in the RF also show an increase in the number of successful *P* and *S* picks when compared to ENZ (Figs 4j–l). However, the fractions are slightly smaller than in the ZNE and WF frames. Instances of misclassification are low in the WF and RF catalogues and there are few inaccurate picks. Pick-residual distributions for each rotational framework and magnitude band are provided in the Supplementary Material (Figs S1–S3), showing that ZNE, WF and RF produce more accurate *P*- and *S*-picks than ENZ, with the difference growing at smaller magnitudes for *S*-picks.

To explore the full parameter space of the model, we selected eleven events and perform model classification tests with the data rotated by a full range of azimuths and inclinations in 15-deg increments. Fig. 6 shows the results of these tests. The radial component for each azimuth and inclination is input to the vertical channel and the transverse components to the horizontal channels. To ensure the events include a range of magnitudes, we include four, four and three events from CAT3, CAT2 and CAT1, respectively. Fig. 6 shows that the largest values for the number of detections and median probability delineate two poles on each graph. One pole has an azimuth and inclination of 270 deg and −15 deg, respectively, whilst the other has an azimuth and inclination of 90 deg and 15 deg, respectively. These poles have an azimuth phase shift of approximately 180 deg and an opposite polarity of inclination. Further, the poles agree well with the actual azimuth and inclination from receivers to the events, based on existing locations (white stars) and test events with a 180-deg azimuth phase shift and opposite polarity inclination (grey stars).

The weights belonging to filter kernels in shallower layers of the DL model can inform about general decision making abilities of a CNN (S. Lapins *et al.* 2021). For the three input components (E,N,Z), we extract seven weights belonging to eight filter kernels in the input 1-D convolutions layer of CubeNet (G. Chen & J. Li 2022). In Figs 7(a), (b), (d) and (e), the filter weights that convolve with the horizontal components are approximately zero, whereas the vertical component weights are non-zero. The opposite can be seen in other kernels (Figs 7c, g, h). Weights in Fig. 7(f) are near-zero for the N and Z component but non-zero for the E component. These kernels suggest that the CNN has learnt to pick apart *P*- and *S*-wave arrivals on different components—specifically, that it expects to find *P*-wave arrivals on the Z component and *S*-wave arrivals on the E and N components.

When working with data for which this is not the case, for example with data recorded on downhole arrays where the seismic energy is expected to arrive near-horizontally, rotating the data such that it meets the expectations of the CNN can produce significant improvements in model performance.

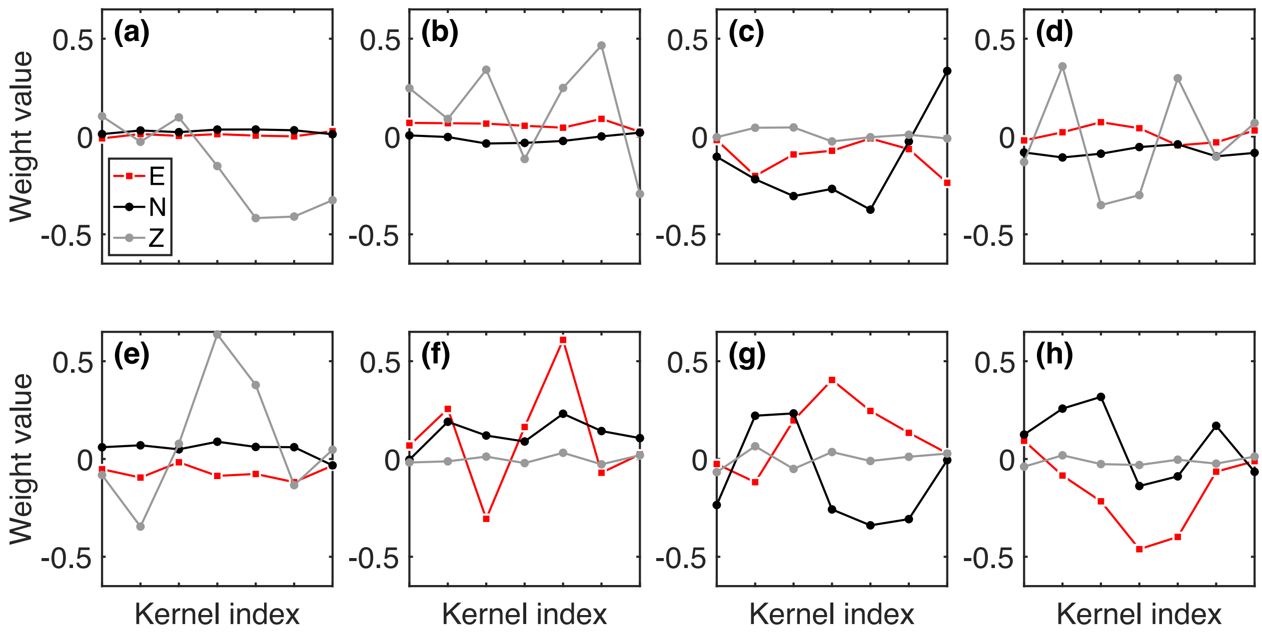

Figure 7. Filters kernels within the first 1-D convolutional layer of CubeNet. Each panel shows three kernels with seven weights. Line colours indicate the component of the trace: E is red, N is black and Z is grey.

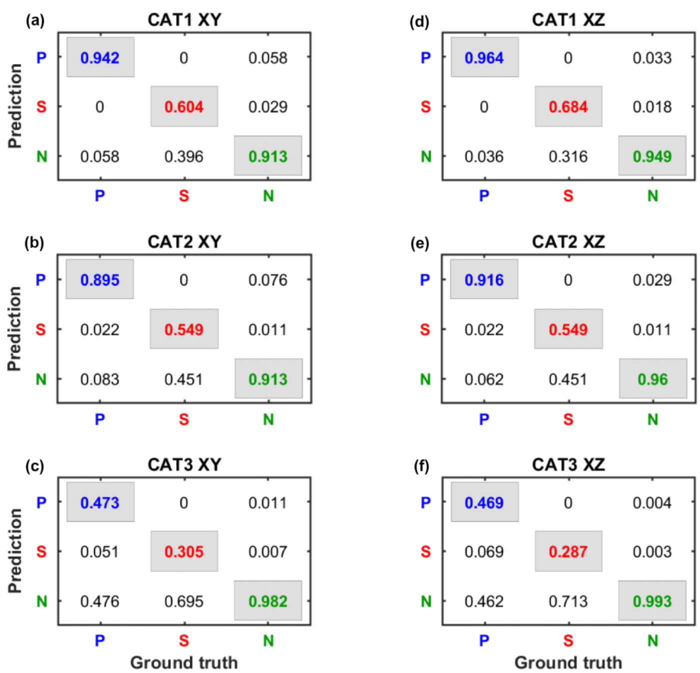

Figure 8. Confusion matrices when regularizing station locations according to station latitude and longitude (XY) **(a–c)** and station longitude and depth (XZ) **(d–f)**. Note the ENZ rotation frame is used for all tests.

5.2 Model performance in different regularizations

The PNR2 monitoring array has a small spatial extent with respect to longitude and latitude (XY). Instead, most of the variation in station position occurs with respect to depth (Z), as is seen in Fig. 1(b). We test and compare two different spatial regularization schemes to observe how these may alter model performance. The tests described above performed regularization in XY-coordinates, as done by G. Chen & J. Li (2022). Here, we also test the results when regularization is done across the XZ coordinates. Note that we revert back to the original ENZ rotation frame for these tests because this allows us to measure the impacts of varying the regularization alone without impacts from varying the rotation frame.

Results in Fig. 8 suggest that regularizing with respect to depth (XZ) has varying effects at different magnitudes. For CAT1, there is a 2 per cent increase in the number of successful *P* picks and an 8 per cent increase in the number of successful *S* picks when the regularization is switched from XY to XZ. However, there is little change in the number of successful picks for CAT2 and there is a decrease in the number of successful picks for CAT3 when switching from an XY to an XZ regularization. Finally, there is an approximately 50 per cent reduction in the number of misclassified or inaccurate (>0.005 s from manual pick) *P* picks when changing from the XY to XZ regularization. Hence, there is an increase in successful noise classification.

6 CONTINUOUS WORKFLOW TO DETECT, ASSOCIATE AND LOCATE NEW EARTHQUAKES

Having tested CubeNet's performance, we incorporate it into a workflow to detect, associate and locate earthquakes using continuous waveforms from PNR2. We choose a 60-min period, between 10:00 and 11:00 a.m. on 2019 September 16, during which seismicity rates were very high (Fig. 1). The operators injected fluid during this period and 1853 earthquakes were detected, associated and located using the CMM (J. Drew *et al.* 2013) method. We select this very active period because we can analyse how well CubeNet processes phase arrivals tightly clustered in time, and how well it processes arrivals from earthquakes of varying magnitude ($-2.2 \leq M_w \leq 0.6$).

6.1 Phase picking

We apply CubeNet iteratively using 3 s waveforms with a 0.5 s overlapping window to account for arrivals which may be at the very start or end of a trace (S. Lapins *et al.* 2021; G. Chen & J. Li 2022). We vary several parameters in CubeNet to observe how these affect the number of picks and the number of located earthquakes. We test three rotation frames, ENZ, ZNE and WF, and two regularization schemes, XY and XZ, leading to six catalogues in total. For each waveform from each station, we find peaks in the *P* and *S* probability density functions with a minimum threshold of 0.3 and a minimum separation of 50 samples. As before, we exclude the accelerometer PNR008 from this analysis.

6.2 Phase association, earthquake locations, origin times and magnitudes

We associate the picks using a two-step procedure. First, a density-based clustering algorithm is used to group the *P* and *S* picks separately (F. Pedregosa *et al.* 2011). We cluster the picks in 2-D according to the arrival time and station location. We enforce a minimum cluster size of 2 picks for each phase. For *P* picks, we assume a maximum connection fraction of 0.08 s and 333 m between stations. For *S* picks, we assume a maximum connection fraction of 0.13 s and 333 m. The pick times and relative station locations are weighted, before clustering, to scale these data. The maximum distance between two stations is 333 m, and 0.08 and 0.13 s represent the slowest possible propagation times of a *P* and *S* wave across the array, respectively. We use the lowest velocity, between depths of 1700 to 2300 m (about the borehole) from the velocity model to calculate the slowest propagation time (Cuadrilla 2019). Next, the *P* and *S* clusters are associated. For each *P* cluster, the algorithm searches for an *S* cluster occurring 0.05 to 0.55 s after the *P* cluster. If multiple *S* clusters are found in the window, the earliest cluster is selected. Finally, we check that each *P* and *S* cluster pair has a minimum of eight picks in total.

A quality check procedure is used to validate the association and eliminate signals from anthropogenic sources such as tube waves (A.M. Ionov & G.A. Maximov 1996). We generate two predicted moveout profiles for each phase, for the lower and upper wave velocity, and remove any phase pick outside this region. The profiles start at the earliest phase pick. A linear move out is suitable, in this case, because the size of the array is small compared to the hypocentral distances. We use a range of $4 \leq \mathrm{kms}^{-1} \leq 11$ for *P* picks and $3 \leq \mathrm{kms}^{-1} \leq 9$ for *S* picks. We obtain these values by analysing the PNR2 1-D velocity model. After the quality check, we remove any clusters with less than 8 picks and require at least 2 *P* picks and 2 *S* picks. We note that this quality check procedure depends on the quality of the first pick. Given these events have already been clustered in space and time we assume the picks are coherent, however, some good picks may be lost where a poor first pick is made.

We determine hypocentre locations by minimizing the residuals between modelled and picked traveltimes (G.A. Jones *et al.* 2014). We use a 1-D layered velocity model to generate traveltime look-up tables for each station using the Non-Lin-Loc Eikonal solver (A. Lomax *et al.* 2000). We perform a joint inversion for traveltime residuals and *P* wave particle motions to better constrain location estimates. We bound our hypocentre search within a $2 \times 2 \times 0.7$ km box centred on the median hypocentre location of the seismicity located by the CMM (occurring in our 1-hr analysis window).

We assign magnitudes by comparing origin times and locations to the pre-existing catalogue. We search for events in the pre-existing catalogue with an origin time within 1 s of our catalogue. If there is more than one pre-existing event, we find the event which minimizes the location difference. This assignation assumes that our workflow did not detect any new events missed by the CMM method. Based on the numbers of events detected, this is a reasonable assumption (see below).

6.3 Phase detection and picking results

CubeNet makes 4686 picks in the ENZ XY set-up (Table 2) of which 61 per cent are *P* picks and 39 per cent are *S* picks. regularizing with respect to depth (ENZ XZ) increases the number of picks by 187. Rotating into the ZNE and WF frame increases the number of

Table 2. Statistics for the continuous workflow. N(P+S) = combined picks made by the CubeNet. N(P) = total *P* picks. N(S) = total *S* picks. N(ev) = number of events after association N(qc)'ed = number of events after association and quality check. N(P)ev = total *P* picks for locatable events. N(S)ev = total *S* picks for locatable events.

Rotation	Regularization	N(P+S)	N(P)	N(S)	N(ev)	N(ev) QC'ed	N(P)ev	N(S)ev
ENZ	XY	4686	2859	1827	232	203	1903	1225
ZNE	XY	6780	4228	2552	314	300	2986	2099
Well-frack	XY	6905	4304	2601	321	307	3082	2180
ENZ	XZ	4873	2838	2035	241	221	1943	1333
ZNE	XZ	6944	4308	2636	349	323	3109	2152
Well-frack	XZ	7119	4400	2719	354	335	3182	2265

picks significantly. For example, the ZNE XY catalogue contains 2094 extra picks compared to ENZ XY. The relative proportions of *P* and *S* picks remains constant for all catalogues (57–62 per cent *P* picks and 38–43 per cent *S* picks). regularizing with respect to depth increases the number of picks by 150–250 for all catalogues. The WF XZ catalogue contains 7119 picks which is the largest number of picks of all the catalogues and represents a 52 per cent increase from the ENZ XY catalogue. Note that we do not produce a catalogue for the ray frame rotation because this requires locations to be known *a priori*—the ray frame implementation was out-performed by the WF and ZNE frames in our preliminary testing anyway (Fig. 4).

6.4 Phase association results

For the initial ENZ XY catalogue, 67 per cent of the *P* and *S* picks are successfully associated to form 203 earthquakes (Table 2). In contrast, 75 per cent and 76 per cent of picks are successfully associated in the ZNE XY and WF XY catalogues, respectively. When varying the rotation, a smaller proportion of events fail the quality check criteria. regularizing with respect to depth increases the total number of associated phase picks but does not change the proportion of phase picks that are successfully associated. The WF XZ catalogue is the largest with 335 events.

The ENZ XY catalogue is the smallest with 203 earthquakes. Switching to the ZNE or WF frames added 97 or 104 events to the catalogue, an improvement of 47.8 per cent and 51.2 per cent, respectively. regularizing with respect to depth rather than lateral position further increased the catalogue size by 5–10 per cent. However, these numbers are still small compared to the CMM catalogue that contains 1853 earthquakes during the same time period—the comparison with the CMM catalogue will be discussed further in the following sections.

6.5 Spatiotemporal analysis and comparisons with pre-existing catalogue

The earthquake locations are shown in Figs 9(a) and (b). The CubeNet locations are systematically shifted to the east in comparison to the CMM locations. This is not an issue with inaccurate picks, but is caused by differences in the velocity model used during real time processing by the contractor using CMM to which, we do not have access. The difference between the CubeNet and CMM locations show an approximately constant median offset vector of 132 m East, 56 m South for the ENZ XY implementation and 131 m E, 54 m South for WF XZ, corresponding to absolute offsets of 143 and 142 m, respectively.

Accounting for this artefact, the hypocentres match well with the CMM location estimates, forming two clusters both trending along the maximum horizontal stress direction (roughly north–south). When comparing the events detected by the ENZ XY versus ZNE XZ CubeNet implementations (red dots in Fig. 9), the extra events detected by the ZNE XZ method better delineate the main cluster, as well as starting to delineate the western-most cluster, which was barely identified at all by the ENZ XY catalogue. The clusters generated by the WF XZ implementation appear more linear than the clusters produced by the ENZ XY implementation.

Although picking residuals for the continuous test cannot be directly evaluated for PNR-2 as the phase picks are not available for the CMM catalogue, the increased linearity of the CubeNet cluster after rotation and regularization may indicate more accurate event locations. This is supported by the classification tests (Figs 4, 8 and Supplementary material Figs S1–S3), which show that changing rotational frameworks and regularizations produce more picks within 0.005 s of the analyst pick across the magnitude bands (CAT1 to CAT3).

Figs 9(c) and (d) compares the magnitude time distributions for the catalogues in this study and the CMM. At large magnitudes, the catalogues are well matched. Most of the extra events, from the ZNE rotation and XZ regularization, have a magnitude less than -0.5 M_w. The lowest magnitude event detected by this study is M_w -1.92 whereas the lowest magnitude event detected by the CMM is $M_\mathrm{w} - 2.1$.

The magnitude–frequency distributions of this study and the CMM (Fig. 10) agree well at the higher tail of the distribution. However, it is evident that the CMM detected far more earthquakes than CubeNet. The CMM magnitude distribution peaks at $M_\mathrm{w} = -1.4$ and the lowest magnitude is at -2.1. This distribution has a steep fall-off at low magnitudes. In contrast, the magnitude distributions recalled by CubeNet peak around -0.75 and few earthquakes below $M_\mathrm{w} = -1.5$ are detected. Our distributions fall-off gradually at

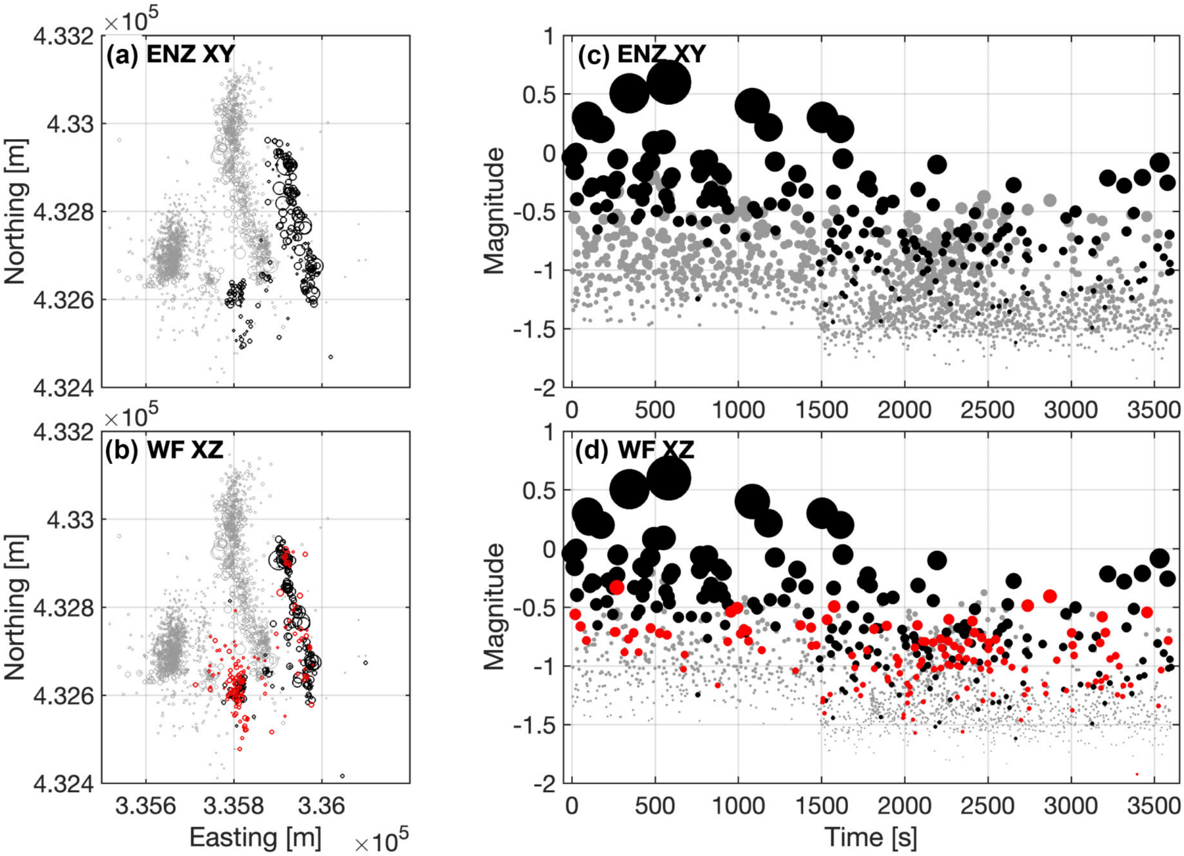

Figure 9. Spatiotemporal analysis of the event catalogues produced by the different CubeNet implementations. In **(a)** and **(b)** we show event location maps, and in **(b)** and **(c)** we show the magnitude evolution with time. Grey circles are events from the CMM catalogue (T. Kettlety *et al.* 2021). Black circles in **(a)** and **(c)** are events from the ENZ XY CubeNet implementation. Black circles in **(b)** and **(d)** are events detected by both the ENZ XY and WF XZ implementations, and red circles are events detected by the WF XZ implementation but not the ENZ XY implementation.

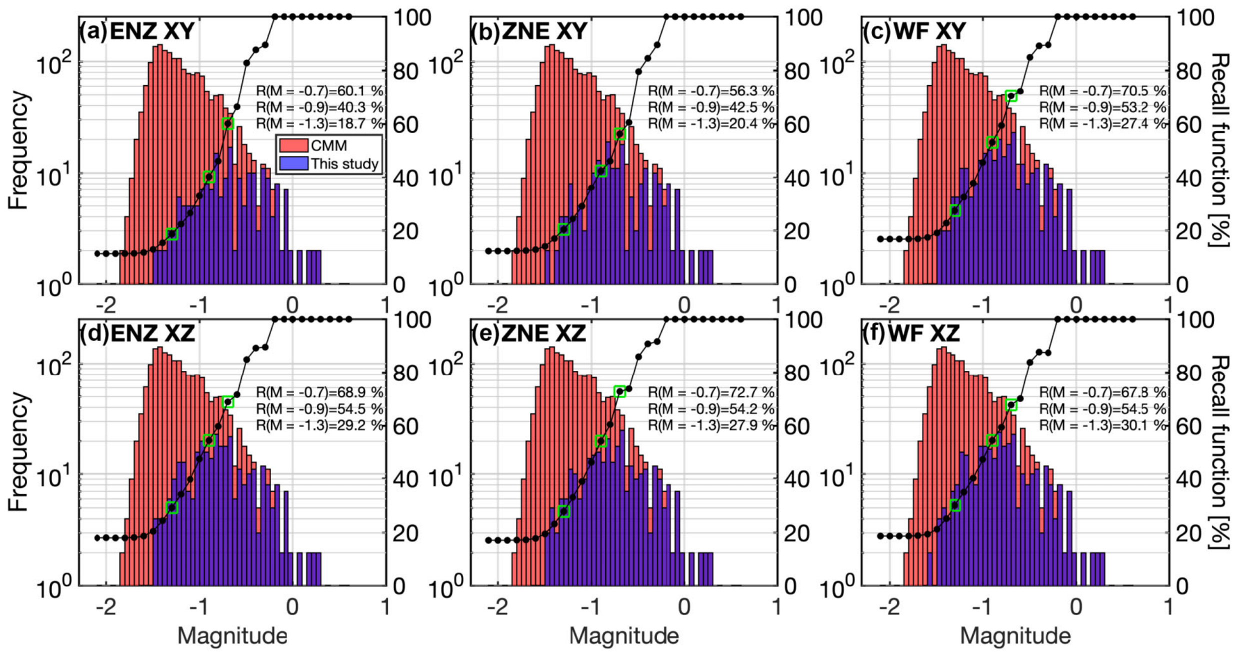

Figure 10. Comparison of the magnitude distributions for events detected by the CMM method and the workflow used in this study. Six catalogues are shown: ENZ frame and XY regularization **(a)**, ZNE frame and XY regularization **(b)**, WF frame and XY regularization **(c)**, ENZ frame and XZ regularization **(d)**, ZNE frame and XZ regularization **(e)**, and WF frame and XZ regularization **(f)**. The black lines with black dots shows recall function. For each magnitude M, the recall function is defined as the percentage of CMM events that are successfully detected in this study for all magnitudes greater than M. Green boxes highlight recall values at $M_w = -0.7$, $M_w = -0.9$ and $M_w = -1.3$.

Table 3. CubeNet run-time statistics on different processors. Run-time for 5 min of time-series data with a 0.5 s overlapping window and including pre-processing stages.

Processor	Processing time [mm:ss]	Data window length [mm:ss]	Data window length / Processing time
2.3 GHz Ryzen quad core CPU	30:22	05:01	0.165
2.3 GHz Intel Xeon CPU	07:54	05:01	0.635
NVIDIA V100 Tensor Core 16 GB GPU	03:03	05:01	1.645
NVIDIA A100 Tensor Core 80 GB GPU	02:42	05:01	1.858

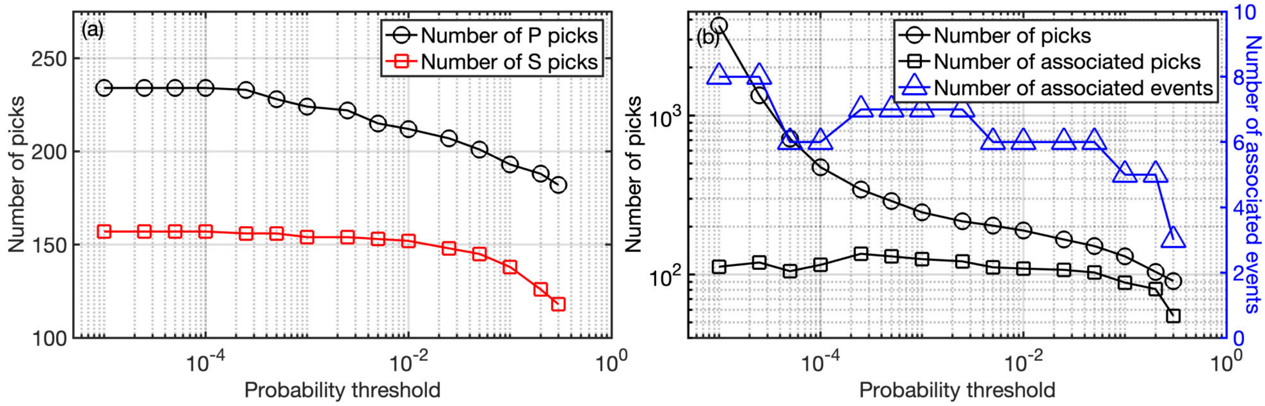

Figure 11. Influence of the pick probability threshold on the classification tests (a), showing the numbers of successfully classified *P* (black circles) and *S* (red squares) picks as a function of the probability threshold. Results are for CAT3 magnitude events. In (b) we test the influence of the probability threshold on the continous workflow, showing the number of detected (black circles) and associated (black square) picks for 1 min of continuous data, and the number of associated events (blue triangles). Note that the CMM catalogue has 18 earthquakes during this period.

lower magnitudes. All catalogues recall all events with $M_w \geq -0.2$. We observe that varying the rotation frameworks and regularization successfully increases the recall at lower magnitudes. For example, CubeNet successfully recalls 40.5 per cent of the CMM catalogue for $M_w \geq -0.9$ in the ENZ XY implementation, whereas in the best catalogue (WF XZ), CubeNet successfully recalls 55.1 per cent at $M_w \geq -0.9$.

6.6 Runtime and computational efficiency

To assess CubeNet's speed, we processed 5 min of continuous waveforms on different processors. We note that the run-time also includes pre-processing stages and that a 0.5 s overlap was used between each three second trace. The results are listed in Table 3. The 2.3 GHz quad core CPU is slow and runs CubeNet six times slower than real time speed. Next, we test one CPU and two Graphics Processing Units (GPUs) (M. Garland 2011). While the Google Colab CPU operates slower than real-time, both the V100 and A100 GPUs operate at least 50 per cent faster than real-time (M. Garland 2011). The A100 GPU performs best and processes these data at nearly twice real-time speed. These results highlight that using state-of-the-art GPUs can lead to much improved gains in processing speed.

We consider the CubeNet implementation as relatively more efficient than the CMM method, which we use as our benchmark. While CubeNet is not as computationally efficient as widely used single-station phase pickers (e.g. PhaseNet; W. Zhu & G. C. Beroza 2019), this is expected because it is a multistation approach and so CubeNet should be compared against other multistation techniques such as beamforming-based methods for a like-for-like assessment. In contrast with the CubeNet catalogue, the CMM catalogue required a supercomputer to generate the PNR2 catalogue in real-time. In this respect, CubeNet represents a more cost efficient solution as it is able to process 2000 Hz data at speeds faster than real-time with the use of a single GPU.

6.7 Probability thresholds

In Fig. 11 we explore how varying the picking probability threshold affects the classification tests and continuous workflow. We performed these tests in the ZNE domain to maximize results. In the classification test results, we note that as the probability threshold decreases, the number of phase picks increase. The number of picks begins to level-off at a threshold of 0.01 for *S* picks and 0.001 for *P* picks and converges towards 234 and 157 picks for *P* and *S* picks, respectively. The previous classification tests in this study, at a threshold at 0.3, resulted in 182 and 118 picks, respectively. The continuous workflow tests show that as the probability threshold decreases, CubeNet makes exponentially more picks. However, the proportion of these which cannot be associated also increases dramatically. The presence of large numbers of unassociated picks suggests that at these low thresholds, a significant portion of the picks are actually false positives, since picks occurring at random will not be associated given our association criteria that requires picks to occur within a range of pre-defined move-outs. CubeNet shows the greatest performance for probability thresholds between 0.00025 and 0.0025. Here, we successfully associate seven events, all of which matched those from the previous catalogue, whilst re-training a high proportion of the initial phase picks.

7 DISCUSSION

The test results show that CubeNet can generalize to unseen data such as from borehole settings because it can successfully detect and pick seismic phases on 2000 Hz borehole microseismic data without the need for re-training. This is a promising result because other multistation models proposed thus far employ GNNs and require re-training for each new network or array-geometry before they can be applied to unseen data. Models used in real-time monitoring for downhole arrays installed to monitor subsurface industrial activities must generalize well to unseen data because an existing catalogue, for training and fine-tuning, will not exist *a priori*.

However, the CubeNet workflow misses many of the events detected by CMM. Notably, the model misses 25 per cent of *S* picks, even at larger magnitudes, with the classification accuracy further decreasing with decreasing magnitude. We hypothesize that CubeNet may have learnt to pick *S* waves according to changes in frequency content of the signal but that it struggles to do this with 2000 Hz data. In our classification tests, most of the model picks are less than 0.005 s from the ground truth pick and the high noise classification accuracy suggests the model is impervious to impulsive noise. However, as in Fig. 2, the model is picking some phases early: this again may be related to the high sampling frequency. We note that our pick time corresponds to the maximum probability and there may be other probability peaks, with a lower peak probability, that give a better fit to the true phase pick.

7.1 Improving model performance: a consideration of incidence angle

Optimizing the input data by accounting for incidence angle and source-station geometry when preparing input data is a straightforward way to adapt surface-trained multistation models to borehole settings without retraining. We show that adjusting the target waveforms to better resemble the training data can significantly improve model performance. The tests in Fig. 6 show the two rotational poles with the highest detections and probabilities are well aligned with the known location of the seismicity, since this rotation places the *P* wave energy onto the Z component. Hence, these tests demonstrate that the model performance is maximized when we input radial component data to the vertical channel of CubeNet and transverse component data to the horizontal channels of CubeNet. These observations highlight the learnt features of CubeNet and are supported by the filter kernels which separate the vertical and horizontal components in the shallowest layers of the model (the first 'decision' the model makes).

In our case, the alternative rotation frames (ZNE and WF) increased the number of successful picks, increased the pick probability and eliminated instances of misclassification. Small increases in the number of *P* and *S* picks, across the entire continuous data set, could significantly increase the numbers of detected events, as well as reducing location uncertainties with greater numbers of more accurate picks. Our results emphasize that care should be taken when applying DL pickers, trained on surface data, to borehole data. Referring to Table 2, when our data were not orientated favourably (i.e. using the ENZ rotation), there was an up to a 36 per cent reduction in *P* detections and up to a 33 per cent reduction in *S* detections.

While the RF rotation framework is theoretically optimal, there is a growing performance gap of CubeNet's phase classifications between the other frameworks (ZNE and WF) and the RF framework at lower magnitudes. For CAT3, the PCA-derived particle motion rotation angle estimates can become less reliable due to low SNR. Low SNR can therefore result in inaccurate rotation angles that fail to correctly maximize the *P*-wave energy onto the expected channel. For larger events (CAT1), we still observe that RF *P*-wave classification performance is comparable to ZNE and WF. RF *S*-wave classification could be underperforming due to structural heterogeneity in the shale along the ray path, causing *S*-wave polarization to deviate from the perpendicular orientation assumed by the RF rotation.

Importantly, all three rotation frameworks (ZNE, WF and RF) improve classification performance by up to 21 percentage points (for small events) for *P*-wave classification relative to ENZ. CubeNet's performance improvement confirms that component orientation correction is beneficial when we use directional source information to rotate the waveforms to match the polarization learned by the model.

We use the continuous workflow to highlight how changes in the number of picks, associated with varying regularization and rotation implementations, affect an earthquake catalogue. Not only did the switch from the ENZ to the WF or ZNE rotation frames provide a significant increase in the numbers of picks, but the more accurate picking meant there was an increase in the proportion of successfully associated events. Given these new phase picks are well-clustered with existing picks, they are likely good quality phase picks from true events rather than from noise.

For the PNR2 data set, the ZNE and WF rotations were easy to implement in a continuous workflow and have no effect on the computational expense. A single rotation was effective because all of the seismicity occurred in relatively tight clusters, and the hypocentral distances were larger than the aperture of the array, meaning that all of the events would have had similar incidence angles. However, this may not apply in other settings. For example, there may be multiple injection wells at different azimuths to the array, or hydraulic conduits may allow seismicity to migrate to unexpected positions (e.g. N. Igonin *et al.* 2021). At CCS sites, the pressure plume created by injection will be extensive, and it may be difficult to predict in advance where microseismicity might nucleate.

To account for this, traces could be rotated in real-time to focus on the most recent seismicity. After each pick, the azimuth and inclination can be measured from the waveforms (K. De Meersman *et al.* 2006) and the traces can be rotated to 'anticipate' the next seismicity. However, this could introduce a detection bias.

Lastly, re-training, fine-tuning or transfer-learning CubeNet using the PNR2 downhole data set would likely enhance model performance, especially for the smallest events. Even after implementing the low-cost input adaptation strategies, we find that CubeNet struggles to detect events under M_w -0.2. Only $\sim$16 per cent of CubeNet's training data set containing events with $M_\mathrm{w} < -0.2$ (G. Chen & J. Li 2022), which likely explains the missed detections. However, the aim of this study is to evaluate different strategies to optimize detection for borehole arrays using an out-of-the-box multistation DL model (CubeNet) trained solely on surface recordings, as this mirrors the reality where labelled downhole data sets are scarce or entirely unavailable due to proprietary constraints. Furthermore, the PNR2 data set is one of the very few high-quality downhole data sets that are publicly accessible, and retraining exclusively on this site would risk overfitting to its local geological and operational site characteristics. Future work could explore retraining or applying transfer-learning approaches using multiple available downhole data sets from diverse sites and geological settings to enhance model performance while preserving generalizability.

7.2 Detection thresholds for CubeNet versus CMM

The magnitude recall results showed our changes in rotation frame and regularization allowed us to detect lower magnitude events. If arrival picking were solely related to SNR, we would expect a sharp falloff in the number of events at lower magnitudes (as with the CMM catalogue at $M_\mathrm{w} \leq -1.5$ in Fig. 10). However, we observe a gradual fall-off. This suggests that SNR is not the only property controlling successful picking by CubeNet. For example, the frequency content of the signal may affect whether a pick is made. We find that our changes in regularization and rotation improved the recall at lower magnitudes. However, the completeness limit (lowest magnitude where recall is 100 per cent) remains at $M_\mathrm{w} = -0.2$ for all the CubeNet catalogues, implying this limit is independent of the rotation and regularization.

Fig. 10 emphasizes the large disparity in the number of events detected by our workflow compared to CMM. In addition to the different frequency contents of the training data versus the application, one important factor may be that CubeNet expects 64 3C traces input on an 8-by-8 grid. However, at PNR2 we have only 11 3C stations. As such, in our analysis the majority of input channels to CubeNet are always empty, irrespective of the choice of regularization. This could have had a significant impact on the model's performance.

The CMM method involves a stacking procedure which serves to boost the SNR, and this may allow it to detect events with relatively low SNRs on individual traces. The improvements in detection of small-magnitude events produced by methods that stack microseismic data across arrays are well established: for example J.P. Verdon *et al.* (2017) showed that stacking, even across relatively small arrays of up to 12 stations, enabled detection of events where the SNR was close to 1 on each individual trace.

The events that comprise the training data set for CubeNet were detected using a Kurtosis-based method (C.D. Saragiotis *et al.* 2002). Our own experience suggests that Kurtosis-based methods require a relatively high SNR for successful detection. This means that the training data set for CubeNet would have been based on events that had high SNRs on at least a proportion of the recorded traces. The significant out-performance by the CMM method may stem from the CMM method's ability to detect events where SNRs are close to 1 on all the individual traces, whereas CubeNet has not been trained with such events. Further improvements in CubeNet's detection performance may therefore be achieved by re-training using a population of events that have been detected using beamforming and stacking-based methods (which would therefore have significantly lower SNRs).

7.3 On the choice of probability threshold

Throughout this study, we use a probability threshold of 0.3 to distinguish between successful and unsuccessful phase picks. Given the differences between the training and application data set, a lower threshold might be better. We indeed observe increases in performance in both the classification test and continuous workflow. While CubeNet did not detect as many events as CMM, the algorithm can still detect high quality signals (which passed our association and quality check) when using a probability threshold of 0.00025. This threshold is very low and suggests that probability estimates from CNNs cannot be treated like standard probability values (Y. Park *et al.* 2023). Suitable probability thresholds may vary between different data sets and for different models.

8 CONCLUSIONS

We have demonstrated that considering different input adaptation methods can optimize the application of an off-the-shelf multistation DL algorithm like CubeNet to detect and pick body wave phases recorded on a borehole array sampling at 2000 Hz. The proposed methods, including a rotational framework and spatial regularization informed by source direction and station geometry, are simple, implementable strategies for applying surface-trained multistation DL models to future borehole deployments. This is a key result because the number of publicly available borehole data sets is limited, as most are proprietary, making the adequate training of bespoke borehole DL models from scratch impractical. For the same reason, this study is based on a single array and site and so testing on further borehole data sets, as they become more available would help establish how broadly the approach generalizes.

On a catalogue of known and analyst-picked waveforms, CubeNet has a high classification accuracy for earthquakes with magnitude greater than M_w 0.5 and successfully picks 94 per cent of *P* picks and 60 per cent of *S* picks. However, CubeNet misses many picks of magnitude less than −0.5 M_w, only picking 47 per cent of *P* picks and 30 per cent of *S* picks. We were able to increase the number of successful picks by accounting for differences in the training and application data set. We rotated traces to account for different incidence angles at surface versus borehole stations. The rotation increased *P* and *S* recall by 9 per cent points and 14 per cent points, respectively, for our known events, and produced a 65 per cent increase in the number of earthquakes detected, associated and located using a continuous workflow. These results provide new insights about the classification abilities of DL algorithms in seismology. We also find that accounting for the depth variation of borehole stations when performing the regularization can increase the number of picks made by CubeNet. While these changes improve model performance, the model still misses 32 per cent of *P* picks and 55 per cent of *S* picks of lower magnitude earthquakes compared to CMM. When it does pick, CubeNet is very accurate and the majority of the picks were within 0.005 s of a human analyst pick.

We incorporate CubeNet into a continuous workflow to detect, associate and locate earthquakes from a borehole array with low azimuthal coverage. The earthquake locations are well matched with those from pre-existing catalogues and provide further evidence of CubeNet's high picking accuracy. However, CubeNet misses most of the lower magnitude events from the existing catalogue, which make up 82 per cent of the catalogue. This is because a smaller proportion (only ∼16 per cent) of CubeNet's training data set comprizes events with $M_w < -0.2$. Lastly, we test the computational expense of CubeNet and find that it can process data at nearly twice real-time speed using a single GPU. Another aim of this study was to investigate CubeNet's applicability for real-time monitoring. We conclude that CubeNet can monitor HF induced seismicity on borehole arrays, however CubeNet is not currently suited to detect smaller magnitude seismicity when applied off the shelf. CubeNet performs best for events within the magnitude range that is well-represented in its original training data (recalling 100 per cent of events above M_w −0.2). Additional training data would likely be required to improve performance for smaller or lower SNR events, which were under-represented during training. Despite this, CubeNet's ability to generalize to a very different application data set and its high picking accuracy makes it a very promising base model. Multistation DL methods are in the early stages of development, and further development, re- or transfer-learning with borehole data, or with low SNR events detected using stacking-based methods, may substantially improve microseismic monitoring at industrial sites.

ACKNOWLEDGMENTS

We thank Joanna Holmgren and Sacha Lapins for helpful discussions which improved this work. We also thank the editors and reviewers Jannes Münchmeyer, Peidong Shi, Zahra Zali and two anonymous reviewers for their constructive feedback. This work was supported by the U.K. Engineering and Physical Sciences Research Council (grant no. EP/Y020960/1), DarkSeis (EP/Y021126/1), the Natural Environment Research Council (NERC) (grant no. NE/R017956/1) and the NERC SeisGreen Grant (grant no. NE/W009293/1). CSYL was also supported by the NERC Centre for the Observation and Modelling of Earthquakes, Volcanoes and Tectonics (COMET), grant no. GA/13M/031, a partnership between UK Universities and the British Geological Survey (BGS) University Funding Initiative (BUFI) (S468), grant no. GA/20M/004. JPV has acted as an independent consultant for a variety of organizations including hydrocarbon operating companies and governmental organizations on issues pertaining to the regulation and management of induced seismicity. This work was partially funded by the Bristol and Oxford Passive Seismic (BOPS) project, which is funded by a range of oilfield operating and service companies. None of these organizations had any input into the conception, development, analysis or conclusions of this study.

SUPPLEMENTARY MATERIAL

Supplementary materials are available at *GJIRAS* online.

Figure S1 Pick residuals for the *P* and *S*-wave phase picks between CubeNet and the human analyst for CAT1 in the XY regularization.

Figure S2 Pick residuals for the *P* and *S*-wave phase picks between CubeNet and the human analyst for CAT2 in the XY regularization.

Figure S3 Pick residuals for the *P* and *S*-wave phase picks between CubeNet and the human analyst for CAT3 in the XY regularization.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

DATA AVAILABILITY

All seismic data, velocity models, station locations and the benchmark catalogue for PNR-2 operations are publicly available through the North Sea Transition Authority (https://www.nstauthority.co.uk/regulatory-information/exploration-and-production/onshore/onshore-reports-and-data/preston-new-road-well-pnr2-data-studies/, last accessed, 2026 July 15). The CubeNet model is provided by G. Chen & J. Li (2022) at https://github.com/Billy-Chen0327/CubeNet.git (last accessed, 2026 July 15).

REFERENCES

Al-Anboori, A. & Kendall, J.M., 2010. Passive seismic monitoring of reservoirs: a case study from Oman, in Johnston, D.H. ed., *Methods and Applications in Reservoir Geophysics,* Society of Exploration Geophysicists, pp. 441–450.

Chen, G. & Li, J., 2022. CubeNet: Array-based seismic phase picking with deep learning, *Seismol. Res. Lett.,* **93**(5), 2554–2569.

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, *Seismol. Res. Lett.,* **90,** 1902–1915.

Cremen, G. & Werner, M.J., 2020. A novel approach to assessing nuizance risk from seismicity induced by UK shale gas development, with implications for future policy design, *Natural Haz. Earth Syst. Sci.,* **20**(10), 2701–2719.

Cuadrilla, 2018. *Hydraulic Fracture Plan PNR 1/1Z,* pp. 1–24. Available from: https://www.nstauthority.co.uk/media/5845/pnr-1z-hfp-report.pdf (accessed 25 August 2026).

Cuadrilla, 2019. *Hydraulic Fracture Plan PNR-2,* pp. 1–8. *Available from*: https://consult.environment-agency.gov.uk/onshore-oil-and-gas/information-on-cuadrillas-preston-new-road-site/user_uploads/pnr2-hfp.pdf (accessed 25 August 2026).

Dando, B.D.E., Goertz-Allmann, B.P., Kuhn, D., Langet, N., Dichiarante, A.M. & Oye, V., 2021. Relocating microseismicity from downhole monitoring of the Decatur CCS site using a modified double-difference algorithm, *Geophys. J. Int.,* **227**(2), 1094–1122.

De Meersman, K., van der Baan, M. & Kendall, J.M., 2006. Signal extraction and automated polarization analysis of multicomponent array data, *Bull. seism. Soc. Am.,* **96**(6), 2415–2430.

Drew, J., White, R.S., Tilmann, F. & Tarasewicz, J., 2013. Coalescence microseismic mapping, *Geophys. J. Int.,* **195**(3), 1773–1785.

Feng, T., Mohanna, S. & Meng, L., 2022. Edgephase: A deep learning model for multistation seismic phase pickin, *Geochem. Geophys. Geosyst.,*

Garland, M., 2011. NVIDIA GPU, in Padua, D., ed., *Encyclopedia of Parallel Computing,* Springer.

Holmgren, J.M., Kwiatek, G. & Werner, M.J., 2023. Nonsystematic rupture directivity of geothermal energy induced microseismicity in helsinki, finland, *J. geophys. Res.: Solid Earth,* **128**(3), e2022JB025226.

Igonin, N., Verdon, J.P., Kendall, J.M. & Eaton, D.W., 2021. Large-scale fracture systems are permeable pathways for fault activation during hydraulic fracturing, *J. geophys. Res.,* **126,** e2020JB020311.

Igonin, N., Verdon, J.P. & Eaton, D.W., 2022. Seismic anisotropy reveals stress changes around a fault as it is activated by hydraulic fracturing, *Seismol. Res. Lett.,* **93**(3), 1737–1752.

Ionov, A.M. & Maximov, G.A., 1996. Propagation of tube waves generated by an external source in layered permeable rocks, *Geophys. J. Int.,* **124**(3), 888–906.

Jones, G.A., Kendall, J.M., Bastow, I.D. & Raymer, D.G., 2014. Locating microseismic events using borehole data, *Geophys. Prospect.,* **62**(1), 34–49.

Kettlety, T., Verdon, J.P., Werner, M., Kendall, J.M. & Budge, J., 2019. Investigating the role of elastostatic stress transfer during hydraulic fracturing-induced fault activation, *Geophys. J. Int.,* **217,** 1200–1216.

Kettlety, T., Verdon, J.P., Werner, M. & Kendall, J.M., 2020. Stress transfer from opening hydraulic fractures controls the distribution of induced seismicity, *J. geophys. Res.,* **125,** e2019JB018794.

Kettlety, T., Verdon, J.P., Butcher, A., Hampson, M. & Craddock, L., 2021. High-resolution imaging of the M_L 2.9 August 2019 earthquake in Lancashire, United Kingdom, induced by hydraulic fracturing during Preston New Road PNR-2 operations, *Seismol. Res. Lett.,* **92**(1), 151–169.

Lapins, S., 2021. *Detecting and characterising seismicity associated with volcanic and magmatic processes through deep learning and the continuous wavelet transform,* Ph.D thesis, University of Bristol.

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, *J. geophys. Res.: Solid Earth,* **126,** e2021JB021910.

Lim, C.S., Lapins, S., Segou, M. & Werner, M.J., 2025. Deep learning phase pickers: how well can existing models detect hydraulic-fracturing induced microseismicity from a borehole array?, *Geophys. J. Int.,* **240**(1), 535–549.

Lomax, A., Virieux, J., Volant, P. & Berge-Thierry, C., 2000. Probabilistic earthquake location in 3D and layered models , in Thurber , C. H. & Rabinowitz , N., eds, *Advances in Seismic Event Location. Modern Approaches in Geophysics,* vol. **18,** Springer.

Mancini, S., Werner, M.J., Segou, M. & Baptie, B., 2021. Probabilistic forecasting of hydraulic fracturing-induced seismicity using an injection-rate driven ETAS model, *Seismol. Res. Lett.,* **92**(6), 3471–3481.

McBrearty, I.W. & Beroza, G.C., 2025. Double difference earthquake location with graph neural networks: I. mcbrearty et al., *Earth Planets Space,* **77**(1), 127.

Mousavi, S.M. & Beroza, G.C., 2020. A machine-learning approach for earthquake magnitude estimation, *Geophys. Res. Lett.,* **47,** e2019GL085976.

Park, Y., Beroza, G.C. & Ellsworth, W.L., 2023. A mitigation strategy for the prediction inconsistency of neural phase pickers, *Seismol. Res. Lett.,* **94**(3), 1603–1612.

Pedregosa, F. *et al.*, 2011. Scikit-learn: Machine learning in python, *J. Mach. Learn. Res.,* **12,** 2825–2830.

Perol, T., Gharbi, M. & Denolle, M., 2018. Convolutional neural network for earthquake detection and location, *Sci. Adv.,* **4**.

Ross, Z.E., Meier, M.A., Hauksson, E. & Heaton, T.H., 2018. Generalized seismic phase detection with deep learning, *Bull. seism. Soc. Am.,* **108**(5A), 2894–2901.

Rutledge, J.T., Phillips, W.S. & Mayerhofer, M.J., 2004. Faulting induced by forced fluid injection and fluid flow forced by faulting: An interpretation of hydraulic-fracture microseismicity, Carthage Cotton Valley Gas Field, Texas, *Bull. seism. Soc. Am.,* **94**(5), 1817–1830.

Saragiotis, C.D., Hadjileontiadis, L.J. & Panas, S.M., 2002. PAI–S/K: a robust automatic seismic *P* phase arrival identification scheme, *IEEE Trans. Geosci. Remote Sens.,* **40**(6), 1395–1404.

Shearer, P., 2019. *Introduction to Seismology,* 3rd edn., Cambridge Univ. Press.

Sun, C., Shrivastava, A., Singh, S. & Gupta, A., 2017. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. Institute of Electrical and Electronics Engineers Inc., vol. 2017-October, pp. 843–852.

Sun, H., Ross, Z.E., Zhu, W. & Azizzadenesheli, K., 2023. Phase neural operator for multistation picking of seismic arrivals, *Geophys. Res. Lett.,* **50**(24), e2023GL106434.

van den Ende, M.P. & Ampuero, J.P., 2020. Automated seismic source characterization using deep graph neural networks, *Geophys. Res. Lett.,* **47,** e2020GL088690.

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

Verdon, J.P., Kendall, J.M., White, D.J. & Angus, D.A., 2011. Linking microseismic event observations with geomechanical models to minimize the risks of storing CO_2 in geological formations, *Earth. planet. Sci. Lett.,* **305**(1–2), 143–152.

Verdon, J.P., abd Stephen P Hicks, J.M.K. & Hill, P., 2017. Using beamforming to maximise the detection capability of small, sparse seismometer arrays deployed to monitor oil field activities, *Geophys. Prospect.,* **65**(6), 1582–1596.

Wessel, P., Luis, J.F., Uieda, L., Scharroo, R., Wobbe, F., Smith, W.H. & Tian, D., 2019. The generic mapping tools version 6, *Geochem. Geophys. Geosyst.,* **20**(11), 5556–5564.

Zhu, W. & Beroza, G.C., 2019. Phasenet: a deep-neural-network-based seismic arrival-time picking method, *Geophys. J. Int.,* **216,** 261–273.