



**This electronic thesis or dissertation has been
downloaded from Explore Bristol Research,
<http://research-information.bristol.ac.uk>**

Author:

Lim Shin Yee, Cindy

Title:

Using deep learning for phase detection and event location on hydraulic fracturing-induced seismicity

General rights

Access to the thesis is subject to the Creative Commons Attribution - NonCommercial-No Derivatives 4.0 International Public License. A copy of this may be found at <https://creativecommons.org/licenses/by-nc-nd/4.0/legalcode>. This license sets out your rights and the restrictions that apply to your access to the thesis so it is important you read this before proceeding.

Take down policy

Some pages of this thesis may have been removed for copyright restrictions prior to having it been deposited in Explore Bristol Research. However, if you have discovered material within the thesis that you consider to be unlawful e.g. breaches of copyright (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please contact collections-metadata@bristol.ac.uk and include the following information in your message:

- Your contact details
- Bibliographic details for the item, including a URL
- An outline nature of the complaint

Your claim will be investigated and, where appropriate, the item in question will be removed from public view as soon as possible.



School of Earth Sciences

Using Deep Learning for Phase Detection and Event Location on Hydraulic Fracturing-Induced Seismicity

CINDY LIM SHIN YEE

**A dissertation submitted to the University of Bristol in accordance with the
requirement for award of the degree of Masters by Research in the School of
Earth Sciences, Faculty of Science.**

NOVEMBER 2020

Supervisor: Maximilian Werner
Word count: 17,652

Abstract

Induced seismicity is a significant concern during fluid injection projects such as hydraulic fracturing for shale gas, enhanced geothermal systems and wastewater injection. With downhole microseismic monitoring, operators can obtain large seismic datasets to detect hydraulic fracturing induced seismicity (HFIS). Deep learning models like convolutional neural networks (CNNs) can offer rapid event detection in these large datasets. Rapid event detection can be useful for risk management strategies. CNNs have already displayed success in detecting regional earthquakes. Here, we examined whether a CNN pre-trained on regional earthquakes can also detect HFIS within high frequency continuous downhole data. We used data from a shale gas site at Preston New Road, UK, to assess the CNN model. The catalogue of the site, which contains over 23,000 events ($-2.839 \leq M_w \leq 1.155$), was generated using the coalescence microseismic mapping (CMM) method. Using confusion matrices, we evaluated the model's ability to pick P and S-phases on single stations. To assess multi-station performance, we compared event catalogues and locations determined by the model and CMM method. We found that model performance declines with decreasing M_w . The model often misses small $M_w < -2$ events but detects new events not previously catalogued (230 new events within one hour). The model detected many new events during periods of high seismicity during injection. We infer that the CMM catalogue is more complete during less seismically active periods as the CNN model did not detect many new events. This study indicates that the pre-trained CNN offers the potential of detecting most events that the CMM detects (87.7%) in addition to more events during very active periods. The CNN produces these results more efficiently so it is promising, however, it requires further retraining with a dataset that represents the HFIS to improve phase detection and accurate picking.

Dedication and acknowledgements

I would like to give great thanks to everyone who supported me during my Masters by Research project. The first important person to thank is my supervisor Max Werner for this project, his incredibly helpful and inspiring advice, patience and scientific insight. Big thanks to Sacha for guiding and effectively teaching me deep learning- I'm really grateful for all your help! Thank you to Alan, Antony, JV, Tom K, Simone, Adam and Joanna for all the help with the PNR dataset- without you guys, this thesis would not be possible. Thanks also to Men-Andrin Meier for the interesting and helpful discussion on the GPD model!

To my office-mates Stanley, Robbie, Tan, JP, Jessica, Alex and Sirawich thank you for making G10a a chill and productive place to be in (also a fun place to complain about needing coffee). Thanks to the Earth Science department and the Bristol Geophysics Group for being so accommodating, friendly and cool (the geophysics coffee breaks are such a good call). I am very grateful to my former supervisor Mike Kendall and Max (again) for encouraging me to pursue further academia (either inadvertently or deliberately). Also thanks to James Wookey, Nick Teanby and Rich Pancost for support during my APM and throughout my undergraduate!

My fellow Bruneians- Brandon, Doreen, June and Orlando. I enjoy our many inside jokes, amazing dinners and genuine company!

To my friends, Aimee, Beth, Jenny, Jo and Georgia- I appreciate all the supportive dog pictures and fun games.

Also to Nikos and Nazaal!! Thank you for bountiful big-brained discussions, dank memes and much support.

I dedicate this thesis to the coolest and most loving parents who financially funded my academic endeavours as well as my brother and sister, Li Min and Shin Hui who emotionally carried me whilst back at home.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the requirements of the University's Regulations and Code of Practice for Research Degree Programmes and that it has not been submitted for any other academic award. Except where indicated by specific reference in the text, the work is the candidate's own work. Work done in collaboration with, or with the assistance of, others, is indicated as such. Any views expressed in the dissertation are those of the author.

SIGNED: DATE:.....

Students must sign the examination copies but should only print their name in the final version that they electronically submit so that no personal signifiers are shown in the online release of the dissertation.

Table of Contents

Abstract	i
Dedication and acknowledgements	ii
Author's declaration	iii
List of Figures	vii
List of Tables	xii
	Page
1 Introduction	1
2 Background and objectives	4
2.1 Induced Seismicity	4
2.2 Hydraulic Fracturing Induced Seismicity (HFIS)	5
2.3 Mechanisms of Induced Seismicity	7
2.4 Site: Preston New Road, UK	9
2.5 Methods of Earthquake Phase Detection	12
2.5.1 STA/LTA	12
2.5.2 Coalescence Microseismic Mapping (CMM)	13
2.5.3 Template Matching (Waveform Cross-correlation / Match Filtering) and Subspace Waveform Detection	15
2.5.4 Autocorrelation	16
2.5.5 Fingerprint and Similarity Thresholding (FAST)	17
2.6 Deep learning in Earthquake Phase Detection	18
2.6.1 Convolutional Neural Networks (CNN)	19
2.6.2 Transfer Learning	21

2.6.3	The Generalised Seismic Phase Detection (GPD) model	21
2.7	Objectives	22
3	Data	25
3.1	Preston New Road-1z (PNR-1z) continuous downhole data	25
3.2	Preston New Road-1z (PNR-1z) Event Catalogue, Velocity Model and Injection Data	27
3.3	Preston New Road Event Waveforms	31
4	Methods	34
4.1	GPD model	34
4.2	Data Pre-processing	37
4.3	Single Station Classification test	37
4.3.1	Test Dataset for Single Station Classification	39
4.3.2	Confusion matrix, classification metrics and probability thresholds	39
4.3.3	Classification Performance Review on Different Magnitude Bands	41
4.4	Multi-station comparison	42
4.4.1	PNR-1z Continuous Downhole Data Model Tests during Periods of Interest	42
4.4.2	Event Detection Workflow: Phase Grouping, Phase Association and Event Location	43
4.4.3	NonLinLoc Event Locations	47
4.4.4	Event Catalogue Comparison	48
5	Results	49
5.1	Model Classification Performance	49
5.1.1	Classification test results	49
5.1.2	Classification Test on Different Event Magnitude Bands	52
5.2	Multi-station Event Detection Workflow	53
5.2.1	Phase Picks and Phase Grouping	53
5.2.2	Results of the Phase Association Methods	54
5.3	Multi-station Catalogue Comparison	58
5.3.1	High Magnitude Period	58
5.3.2	Event Onset Period During High Injection Rates	59
5.3.3	Event Location Comparison	61
5.3.4	Computational Runtime Comparison	65

TABLE OF CONTENTS

6	Discussion	72
6.1	Single Station Classification and Multi-station Detection Performances .	72
6.2	Event Detection Workflow: Phase Picks, Phase Associations and Event Locations	74
6.3	Periods of Interest	76
6.4	Comparison of the GPD model with Other Earthquake Detection Methods	78
6.4.1	Computational Runtime	78
6.4.2	Event Detection and New Events	78
6.5	The GPD model, Other Deep Learning Models and HFIS	79
7	Conclusions	81
	Appendix A	83
	Confusion Matrices from the Classification Test on Different Event Magnitude Bands	83
	References	86

List of Figures

2.1	Adapted from Herwanger et al. (2013). (A) The three principal stresses: the vertical stress component (S_v), the maximum horizontal component (S_{Hmax}) and the minimum horizontal component (S_{Hmin}) acting in the subsurface where the stress components $S_v > S_{Hmax} > S_{Hmin}$. (B) Illustration of a horizontal well drilled in the direction of the S_{Hmin} with hydraulic fractures growing along the S_{Hmax} (perpendicular to the well bore).	5
2.2	Illustration of three proposed hypotheses to explain the mechanisms that cause HFIS. (A) The pore pressure diffusion model, (B) the poroelastic model and (C) the aseismic slip model (adapted from Eyre et al. (2019)).	8
2.3	A Mohr diagram showing that the Mohr stress circle shifts laterally towards the failure criterion during injection (adapted from Kim & Hosseini (2017)). The increase in pore pressure results in a decrease in effective stress.	8
2.4	The Preston New Road (PNR) hydraulic fracturing site at the North West of England with coordinates from the United Kingdom (UK) Grid Reference (from Clarke, Verdon, Kettleby, Baird & Kendall (2019)). The map shows major roads, nearby villages and the positions of the PNR-1z and PNR-2 wells, the PNR well pad and the surface stations in the area.	10
2.5	The geological cross section at the Preston New Road site showing the wells (PNR-1 and the horizontal well PNR-1z), lithology and the structural features on site (adapted from Cuadrilla Bowland Ltd. (2018)).	11
2.6	Schematic from Drew et al. (2013) of the 3D subsurface travelttime grid with the blue dots as subsurface nodes and the green triangles as surface stations. The CMM method uses the velocity model and travelttime look-up tables to backmigrate onset functions from the seismic data (all 3 components) on each node from each station for each time step. The onset functions are then stacked at each node point.	15

LIST OF FIGURES

2.7	The GPD model successfully detecting the 2016 M_w 7.0 Kumamoto earthquake in Japan (from Ross, Meier, Hauksson & Heaton (2018)). This displays the applicability of the GPD model to detect earthquakes that are outside the region and moment magnitude range of the training dataset.	23
3.1	Spatial maps of all the catalogued events, the geophone array, well stages and wellpaths of PNR-1z (magenta) and PNR2 (black). (A) Plan view and (B) 3D plot. The x, y and z-axis show the Easting, Northing and depth in metres, respectively.	27
3.2	Amplitude clipping on the 3-component seismogram for the $M_w = -0.237$ event on station 24.	28
3.3	Amplitude clipping on the 3-component waveforms of the largest magnitude event, $M_w = 1.155$ on station 2.	29
3.4	Velocity model of the Preston New Road site. Velocity of the P wave and the S wave are in blue and orange respectively. The lithology of the site has also been plotted on the figure.	30
3.5	CMM catalogued events from 8 October to 17 December 2018. Injection rate in black and the events with their respective magnitude in orange. The injection hiatus took place from 3 November to 7 December 2018. The dashed blue lines separate the red ($M \geq 0.5$), amber ($0 \leq M < 0.5$) and green ($M < 0$) light stages of the UK traffic light system.	31
3.6	The smallest event, $M_w = -2.839$, recorded on 3-component seismograms of multiple stations (stations 2, 10, and 23). The station near the toe of the well (stations 23) displays the event more clearly because it is closer to the events. Stations near the heel (stations 2 and 10) do not seem to display clear phase arrivals.	33
4.1	(A) The workflow of the GPD model from input waveforms to output class probabilities. (B) The top panel shows continuous seismic traces of 3 components (E, N and Z) and the bottom panel shows the probability traces of the P (red) and S (blue) phases through time (Ross, Meier, Hauksson & Heaton 2018).	35
4.2	A horizontal section plot showing the moveout of the largest magnitude event ($M_w = 1.155$) across all 24 stations. The three component traces are plotted over each other at each station (total of 72 traces). The y-axis shows the offset, which is the distance from the source to each receiver/station. Each individual phase is within 0.2 seconds.	36

4.3	Continuous trace seismogram (top panel) with the respective spectrogram (bottom panel) for the E, N and Z components on station 17 (A) before and (B) after filtering with a 50 Hz Butterworth highpass filter.	38
4.4	Temporal plots of injection data and events from the CMM catalogue on the periods of interest: (A) high magnitude period and (B) the event onset period during high injection rates. The yellow star in (A) represents the largest event in the CMM catalogue.	44
5.1	(A) Overall F1-scores and (B) F1-scores of individual classes (P, S, Noise) as a function of the probability detection threshold.	51
5.2	Precision (in shades of blue) and Recall (in shades of orange) as a function of probability detection threshold for the P (filled circles) and S phase classes (empty triangles).	51
5.3	(A) Precision and (B) Recall values of all classes (P, S, noise, and the overall value) as a function of moment magnitude bands ($M_w > 0$, $0 > M_w \geq -1$, $-1 > M_w \geq -2$ and $M_w < -2$).	52
5.4	Accurate model picks for the P (red) and S (green) phases. Subplots from top to bottom show the waveforms with normalised amplitude for the E, N and Z components on station 11. The shaded areas around the pick show the assigned time error (P = 0.0025 s, S = 0.0025 s) around the pick according to pick probability (P = 0.9995, S = 0.9734).	54
5.5	An early P phase pick (in red) and a late S phase pick (in green) on normalised waveforms for the E, N and Z component on station 8. The shaded areas represent the time error around each phase pick according to pick probability. The dotted lines are our manual phase picks. Time errors for the P and S phases are 0.025 and 0.0025 seconds, respectively (for probabilities P = 0.5 and S = 0.95).	55
5.6	Empirical Cumulative Distribution Function (ECDF) of the RMS errors for the PS (blue), fixed-difference SP (orange) and fixed-difference PS (green) association methods with a zoomed-in window to show the slight difference in the methods.	56
5.7	(A) ECDF of the horizontal location errors (ERH) and (B) ECDF of the vertical/depth location errors (ERZ) for the PS (blue), fixed-difference SP (orange) and fixed-difference PS (green) association methods.	56
5.8	CDF as a function of (A) absolute x, (B) absolute y and (C) absolute z location differences the three phase association methods.	57

LIST OF FIGURES

5.9	The finalised GPD event detection workflow from multi-station phase picks to an event origin time catalogue with locations.	58
5.10	A temporal plot of the GPD model events (new events as vertical orange lines, missed events in red and identified events in blue) for the high magnitude period on 11 December 2018 (11am to 12pm). The new events are represented by orange lines because the respective magnitudes are not known. The star at 11:21:15.62 represents the largest magnitude event in the CMM catalogue.	59
5.11	E, N and Z waveforms of the new aftershock (at 11:21:26.02) after the high magnitude event on station 24. The CMM and GPD model did not detect smaller events after this first event.	60
5.12	E, N and Z waveforms of a newly detected aftershock (at 11:21:16.67) within the coda of the high magnitude M_w 1.155 event on station 24. This new model event was classified as a false positive event because the phases of the event are not clear.	61
5.13	A temporal plot of the GPD model events (new events as orange lines, missed events in red and identified events in blue) for the Event Onset Period on 11 December 2018 (9am to 10am).	62
5.14	E, N and Z waveforms of the newly detected event followed by a catalogued event (detected by both the CMM and GPD model) during high injection rates on station 24.	63
5.15	Location of the PNR-1z (magenta line) and PNR2 (black line) wells with geophones (cyan triangles) and stages (green circles) overlaid with events during the 3 hour period on 11 December 2018 (9am to 12pm). Left panels (A, C, E) display the missed (red) and same/identified (blue) event locations using the CMM catalogued locations whereas the right panels (B, D, F) use model locations. GPD model locations for new events (orange) were used in all panels. The first row shows the map view of the event locations and the subsequent two rows are cross-sections along the easting and northing, respectively. Stage 38 is where injection took place during the 3 hour period (yellow star).	67
5.16	Seismic features in the cross-section along the easting on the locations of (A) the model events that were also catalogued by the CMM (blue) and (B) the new events only detected by the GPD model (orange). Magenta and black lines are the PNR-1z and PNR2 well, respectively. Red lines represent the subsurface grid set in the NonLinLoc location algorithm. Stage 38 is the representative well location (yellow star).	68

5.17	Histogram of the RMS pick residuals of the located events in milliseconds. Four events with $\text{RMS} \geq 100$ ms are off the graph. The blue, orange and red lines are the median (6.75 ms), mean (7.43 ms) and 95 th percentile (12 ms) of the event picks. The green line represents the RMS error at 6 ms.	68
5.18	Histograms of the (A) horizontal (ERH) and (B) vertical error estimates (ERZ) on the model event locations. In both graphs, the blue, orange and red lines are the median, mean and 95 th percentile.	69
5.19	Error probability density functions (red points) for the locations (blue) of (A, C, E) the largest event in the event catalogue and (B, D, F) a M_w -1.5 event during the high injection period. The red curves on the x and y axis on all subplots are the density curves for the PDFs.	70
5.20	CDF of the event location difference to the well in the (A) easting (blue) and northing (orange), and along (B) depth (purple). In both panels, the darker lines represent the locations from the GPD model and the lighter shaded lines denote the CMM locations.	71
6.1	From Drew et al. (2013), the panels show the event clusters on a cross-section along the strike of a dyke intrusion in Iceland located by (a) the CMM method and (b) after manual refinement of traveltimes, Hypoinverse locations (Klein 2002) and relocations using double-difference (HypoDD).	77

List of Tables

4.1	Layout of the confusion matrix table where TP_{xx} is a true positive of a phase and ε_{xy} is a misclassification for the phases P, S and noise. The subscripts on the labels show the phase the model predicts followed by the actual true phase (e.g., for ε_{ps} , the model predicts a P phase arrival when the actual phase arrival is an S phase- this makes it a misclassification).	40
4.2	Pick probability bands assigned to an error window around the phase pick in samples (chosen for 2000 Hz data) and the corresponding time error in seconds.	48
5.1	The GPD model confusion matrix for 750 400-sample phase windows filtered with a 50 Hz highpass Butterworth filter.	49
5.2	Computational runtimes of the available event detection methods. Autocorrelation, FAST and ConvNetQuake runtime results obtained from Yoon et al. (2015) and Perol et al. (2018).	65

Chapter 1

Introduction

Earthquakes play a crucial role in understanding the geological structures of the Earth. They can tell us about faults, but are also the product of faults. Earthquakes are also key to imaging the Earth's subsurface (seismic tomography). Seismic wave imaging can reveal the existence of geological structures within the Earth such as fault zones, folds and seismic stratigraphies (Romanowicz 2008). Seismic travel-time data from earthquakes aid in event location and velocity model building (Spencer & Gubbins 1980, Zelt & Smith 1992).

Earthquake detection is important in understanding the physics of earthquake clusters and triggering. The more events we are able to detect, the more observational data we have to develop new ideas, test hypotheses and construct models after physical mechanisms that explain induced earthquakes. An improved earthquake detection method can help us study the characteristics of induced earthquakes, interactions between seismic events and potentially aid earthquake arrival prediction. We need to first detect earthquakes to conduct focal mechanism studies (Nakamura 2002) and determine source parameters (Dziewonski et al. 1981). Enhanced earthquake detection can also better constrain the rates of seismicity, which in turn might improve seismic hazard maps and risk management strategies.

A recent article in the *Reviews of Geophysics* (Schultz et al. 2020) showed that hydraulic fracturing induced seismicity (HFIS) is common in many industrial energy-related projects such as in deep enhanced geothermal energy systems (Baisch et al. 2006, Grigoli et al. 2018), hydraulic fracturing for shale gas exploration (Lei et al. 2017, Clarke et al. 2014) and in wastewater disposal (Skoumal et al. 2018, Savvaidis et al.

2020). These induced earthquakes pose seismic hazards to industrial activity, people and infrastructure. Therefore, the implementation of microseismic monitoring is required in these projects so we can observe seismic activity during the project and better understand the physical mechanisms that govern induced seismicity. As the technology for microseismic monitoring becomes more developed, instruments in large arrays can now simultaneously record and collect large volumes of seismic data. Currently, microseismic monitoring instruments can be deployed in boreholes and record at high sampling frequencies (e.g. 2000 Hz). As a result, downhole microseismic monitoring produces high resolution continuous datasets available for analyses.

Improved microseismic event detection could be used to optimise the injection parameters during hydraulic fracturing to ensure that operations are well within Health, Safety, Security and Environment (HSSE) guidelines, improve real-time risk mitigation operations and to maximise reservoir production. Additional observations during hydraulic fracturing operations could potentially contribute to our bigger understanding of the Earth's structure, the mechanics of how earthquakes are triggered, and our understanding of inter-earthquake interactions.

Along with microseismic monitoring, the traffic light system (TLS) has been imposed by regulators to manage the risk of induced seismicity by adjusting the fluid injection rate in real-time during operations (Li et al. 2019, Grigoli et al. 2017). In the UK, Green et al. (2012) reviewed and recommended a TLS with a three stage action plan where each stage is defined by the magnitude range of induced events. During green light ($M < 0$ in the UK) injection operations can proceed as planned, during amber light ($0 \leq M < 0.5$) operations need to proceed cautiously with advised lower injection rates and finally, the red light ($M \geq 0.5$) is when injection operations have to be suspended. During red light, ground velocity and seismicity also have to be closely monitored before possibly resuming (Clarke, Verdon, Kettlety, Baird & Kendall 2019, Grigoli et al. 2017, Mignan et al. 2017). To implement this, the TLS requires processing big data in real-time. Therefore, computationally efficient algorithms are required for automated and fast data analysis.

Artificial intelligence refers to machines programmed to mimic or simulate human intelligence in order to complete tasks or perform analyses. Deep learning models are artificial intelligence technology that are composed of multiple neural layers to learn

general representations of input data with multiple levels of abstraction (LeCun et al. 2015). The implementation of artificial intelligence via deep learning models might be a fast and robust solution for microseismic earthquake detection on large continuous datasets. Deep learning models have been shown to efficiently process large datasets and detect seismic phases for regional earthquakes (Perol et al. 2018, Ross, Meier, Hauksson & Heaton 2018, Woollam et al. 2019) and induced seismicity from underground mining (Huang et al. 2018, Wilkins et al. 2020). Mousavi et al. (2019) and Yang et al. (2021) have trained deep neural networks to detect HFIS recorded on surface stations. However, we currently do not know how well a pre-trained deep learning model will be able to detect seismic phases of HFIS on continuous downhole data sampled at high frequency. With this in mind, this study aims to consider and test the performance of a deep learning model for the improved (and more efficient) detection of HFIS.

In the following chapters of this thesis, I introduce HFIS, the hydraulic fracturing site of interest, current available methods for earthquake detection, as well as deep learning (Chapter 2). In Chapter 3, I study the data collected from the hydraulic fracturing site. For my study, I examine the single-station performance of a pre-trained deep learning model to correctly classify microseismic phases in continuous downhole data from Preston New Road, UK: a hydraulic fracturing shale gas site. Additionally, I present results from a workflow that uses model phase detections from multiple stations for phase association, event detection and location (Chapters 4 and 5). Subsequently, I conduct a multi-station catalogue comparison of two event detection methods: the deep learning model and the coalescence microseismic mapping (CMM) method used for catalogue creation during hydraulic fracturing operations in Preston New Road on different periods of interest (defined in subsection 2.7 in the Objectives section and subsection 4.4.1 in the Methods section). I also assess the computational runtimes of the model and compare the model event locations against the CMM method. I further discuss these findings and its implications in Chapter 6.

Chapter 2

Background and objectives

2.1 Induced Seismicity

Induced seismicity is defined by McGarr et al. (2002) as “earthquakes resulting from anthropogenic activities (e.g. fluid injections into the subsurface) that cause a change in stress similar in magnitude to the ambient shear stress acting on a fault causing the slip to occur”. They pose business risks to the operations but more importantly lethal risks to people and buildings. Induced seismicity is common in fluid injection and geoenergy projects such as hydraulic fracturing for unconventional shale gas (Bao & Eaton 2016, Lei et al. 2017), large geothermal systems (Majer et al. 2007, Mignan et al. 2015, Grigoli et al. 2018) and carbon capture and storage (Hitzman et al. 2012, Zoback & Gorelick 2012, Verdon & Stork 2016). Induced seismicity and its relation to the aforementioned projects is an increasingly controversial topic that draws public concern (Ellsworth 2013). As fluid injection operations have the possibility of inducing large magnitude earthquakes, residents near the injection site are understandably concerned about possible risks, damages and the impact that geo-energy exploitation will have on their communities. Examples of damaging human induced earthquakes include the wastewater-induced 2011 M_w 5.7 Pawnee earthquake sequence in Oklahoma, USA (Keranen et al. 2013), the 2017 M_w 5.5 earthquake in Pohang, South Korea, at an enhanced geothermal system site (Grigoli et al. 2018, Kim et al. 2018), and the 2019 M_w 5.7 hydraulically fractured induced earthquake in the South Sichuan basin, China (Liu & Zahradník 2020). These geoenergy projects were either halted or disrupted as a result of these induced earthquakes.

2.2 Hydraulic Fracturing Induced Seismicity (HFIS)

Hydraulic fracturing is a well stimulation technique used to fracture rocks and thereby increase rock permeability by injecting high-pressure fracturing fluid- typically a water-based agent with sand grains that act as a ‘proppant’ (Hubbert & Willis 1972, Gregory et al. 2011). The shale gas boom in the 2000s started with the ability to horizontally drill along unconventional tight shale gas formations (Brown & Yücel 2013, Wang & Krupnick 2015). This technique is used in unconventional shale formations and where production is highly profitable (e.g. Canada, US). Hydraulic fracturing uses fracturing fluid to produce fractures perpendicular to the minimum horizontal stress, S_{Hmin} (i.e. the minimum principal stress component acting on the rock formation)- in the direction of the maximum horizontal stress, S_{Hmax} (Hubbert & Willis 1972, Ellsworth 2013) (Figure 2.1). The aim of this is to create a connecting fracture network that increases permeability in the rock formation.

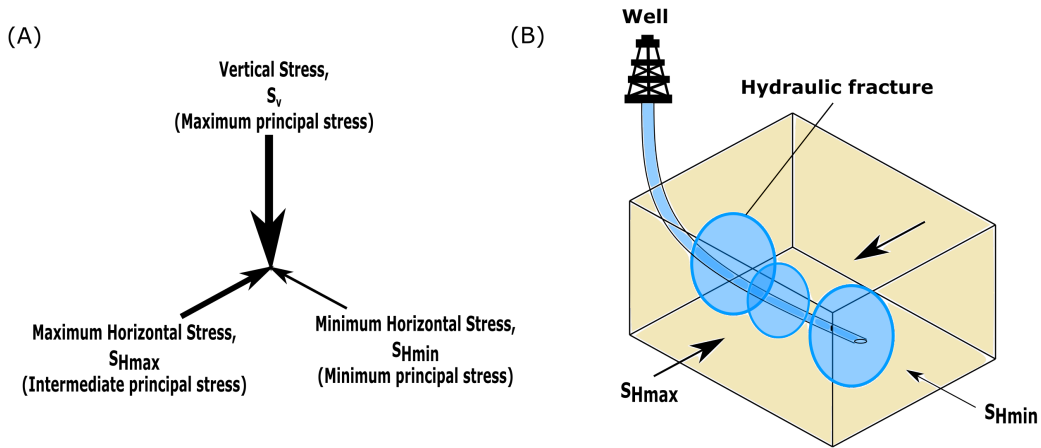


Figure 2.1: Adapted from Herwanger et al. (2013). (A) The three principal stresses: the vertical stress component (S_v), the maximum horizontal component (S_{Hmax}) and the minimum horizontal component (S_{Hmin}) acting in the subsurface where the stress components $S_v > S_{Hmax} > S_{Hmin}$. (B) Illustration of a horizontal well drilled in the direction of the S_{Hmin} with hydraulic fractures growing along the S_{Hmax} (perpendicular to the well bore).

Typically, hydraulic fracturing produces seismicity within the moment magnitude range, $-3 < M_w < 0$ (Van Der Baan et al. 2013, Warpinski et al. 2013). However, Bao

& Eaton (2016), Lei et al. (2017), Grigoli et al. (2018) and Liu & Zahradník (2020) have shown that hydraulic fracturing can induce earthquakes with moment magnitude up to M_w 3.9, M_w 4.7, M_w 5.5 and M_w 5.7 respectively. An important example that illustrates the risks of induced earthquakes is the seismicity induced at a conventional gas field in Groningen, Netherlands- the largest gas field in Europe (Whaley 2009). In 2003, seismicity increased rapidly and in 2012, the largest earthquake M_L 3.6 earthquake significantly damaged buildings, infrastructure and prompted public protests (van Thienen-Visser & Breunese 2015). Consequently, the Groningen gas field had to reduce production and forewent large economic income from expected revenues. The Netherlands Government (2014) also offered a total of 1.2 billion euros over 5 years to compensate residents affected by the earthquakes, causing a negative impact on the national budget. Following that, the Dutch cabinet announced reduced gas extraction activities in 2022 and a complete shut-down of all field operations by 2030 (Vlek 2019). This case example shows that induced earthquakes can shut down large oil and gas production as well as associated economic benefits that come with this project.

In some countries, regulators have required the implementation of microseismic monitoring on geoenery projects from the beginning of operations. Partially due to the increasing surge of induced seismicity from fluid injection projects, microseismic monitoring implementation is very important in tracking small induced earthquakes and imaging the subsurface during production operations (Warpinski et al. 2010). Using surface seismometers or borehole geophones, microseismic monitoring can be a powerful tool to obtain a picture of various geomechanical processes ongoing during fluid injection, including: the tracking of fluid flow (Mukuhira et al. 2020), measuring changes in the mechanical properties of the reservoir (Sone & Zoback 2013), inferring changes in the state of stress (Kettlety et al. 2020), mapping hydraulic fractures (Maxwell et al. 2010), and identifying fault reactivation or failures in the source rock (Rutqvist et al. 2013).

Microseismicity is expected to cluster in the direction of S_{Hmax} and along any pre-existing faults at the hydraulic fracturing site during injection (Chen et al. 2018). Hydraulic fracturing induced events may delineate pre-existing faults (if any) in the hydraulic fracturing site. Thus, microseismic monitoring can be utilised to track the opening of fractures (Pytharouli et al. 2011) and fluid migration pathways in the crust. Learning about the pre-existing geological structures at the site and tracking the fracturing fluid flow is essential in any risk mitigation strategies the operator might consider.

To mitigate the risks posed by HFIS, regulators have imposed traffic light schemes (TLSs). TLSs manage the risk by having action plans for the operator in multiple stages—often divided by the magnitude of induced events. In the UK, these stages are green, amber and red and are respectively defined from a review by Green et al. (2012) as $M < 0$, $0 \leq M < 0.5$ and $M \geq 0.5$ (UK Government, Department for Business, Energy & Industrial Strategy 2013). Operations are able to proceed during the green light stage but when induced events exceed the green-amber magnitude threshold, it is advisable to lower injection rates and proceed with caution. Injection operations should stop or pause if the largest induced event exceeds the magnitude set by the amber-red stage threshold. The implementation of a TLS with real-time monitoring helps to make informed decisions and enact effective action plans to manage risks during fluid injection operations. The microseismic monitoring of hydraulic fracturing-induced earthquakes is therefore invaluable for illuminating subsurface processes that could aid operators to optimise hydrocarbon recovery (Zou 2017, Le Calvez et al. 2005, 2007), identify seismic risks, and enable risk mitigation strategies of induced seismicity (Bommer et al. 2015).

2.3 Mechanisms of Induced Seismicity

There are three main hypotheses that have been proposed to explain the mechanisms that cause fluid injection induced seismicity. One of the proposed hypotheses involves direct hydrological communication with the stimulated fractures and a pre-existing fault (Raleigh et al. 1976, Bao & Eaton 2016). Azad et al. (2017) suggested that this hydrological link between the stimulated fractures and the pre-existing fault plane have to be directly intersecting (e.g. Figure 2.2(A)), or connected naturally via a fracture network (Eyre et al. 2019). Fluid presence in the pre-existing fault increases the pore pressure within the fault, causing effective stress to decrease. Figure 2.3 shows that the decrease in effective stress shifts the Mohr stress circle laterally towards the failure envelope. This makes it more possible for the Mohr circle to intersect the failure criterion and thus brings the fault closer to failure (Kamei et al. 2015, Rutter & Hackston 2017).

However, the pore pressure diffusion model struggles to explain documented HFIS that nucleate at distances away from the stimulating well (Shirzaei et al. 2016, Kozłowska et al. 2018). Segall & Lu (2015) use the concept of poroelastic stress transfer to explain

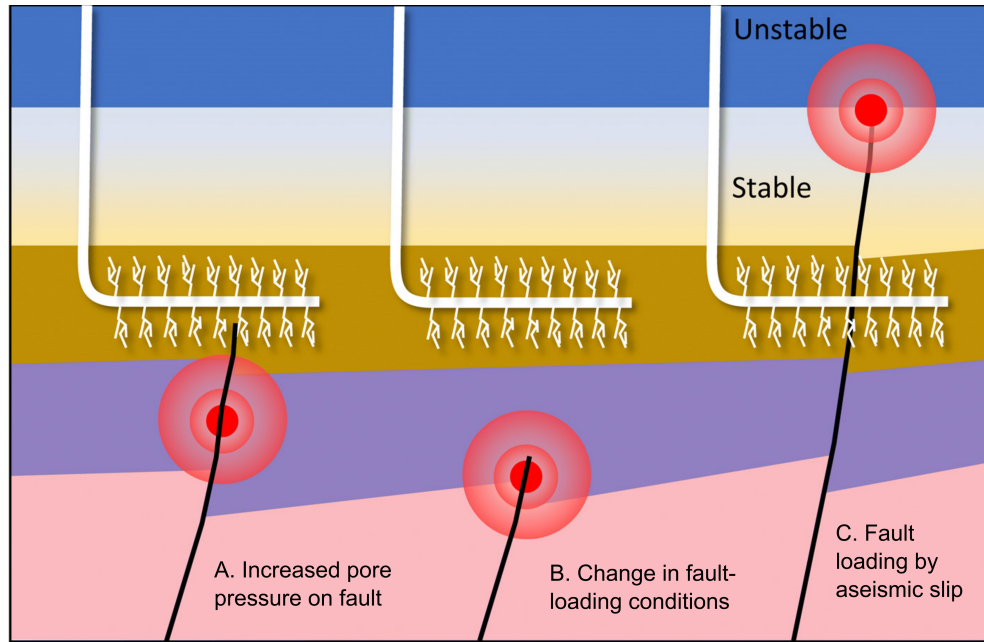


Figure 2.2: Illustration of three proposed hypotheses to explain the mechanisms that cause HFIS. (A) The pore pressure diffusion model, (B) the poroelastic model and (C) the aseismic slip model (adapted from Eyre et al. (2019)).

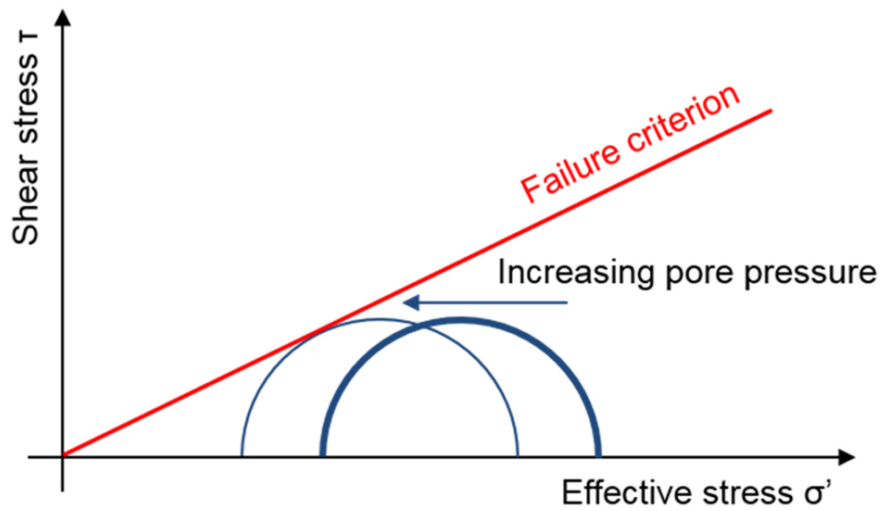


Figure 2.3: A Mohr diagram showing that the Mohr stress circle shifts laterally towards the failure criterion during injection (adapted from Kim & Hosseini (2017)). The increase in pore pressure results in a decrease in effective stress.

the distal induced seismicity. They propose that there is poroelastic coupling of the rock matrix and the fluid injected hydraulic fractures so it is not necessary for the pre-existing

fault to intersect with the stimulated fractures (Schultz et al. 2020). When fluids are injected into the hydraulic fractures, the change in pore volume in rock mass will alter the loading conditions on a fault- this can cause a critically stressed nearby fault to fail (Figure 2.2(B)). Therefore, poroelasticity means that seismicity can be induced at distances away from the injecting well and it has been suggested as the mechanism of the induced seismicity occurring hundreds of meters from wastewater disposal wells (Goebel & Brodsky 2018, Zhai et al. 2019).

Alternatively, Eyre et al. (2019) suggest that aseismic slip driven by pore pressure is the triggering mechanism for HFIS. Several laboratory experiments and in-situ observations have shown inconsistencies with both the pore pressure diffusion model and the poroelastic hypothesis- these studies show that presence of fluids (e.g. water or fracturing fluid) in host rock that is clay-rich and contains high total organic content (i.e. shales) will encourage stable sliding on shale-bearing faults (Kohli & Zoback 2013, Scuderi & Collettini 2018, Cappa et al. 2019). The aseismic slip model is consistent with rate-and-state friction theory and the velocity-strengthening behaviour in shale reservoir rocks. For this model, Eyre et al. (2019) explain that the influence of the pore pressure diffusion from hydrologic connections between the fault and the stimulated fractures only cause nearby aseismic slip that propagates along the fault. Aseismic slip has been observed to outpace the pore pressure migration front (Viesca 2015, Bhattacharya & Viesca 2019) until it reaches more distant seismogenic zones where the earthquakes are induced (Figure 2.2(C)).

Earthquake detection provides key insights and the observations needed to test hypotheses about the mechanisms that induce seismicity. An improved understanding on the characteristics and behaviour of the seismicity (e.g. clustering and event migration) via enhanced event detection can give indications of what physical mechanisms are in play during fluid injection operations. Knowing which mechanisms are causing the induced seismicity may aid hazard assessment and risk mitigation of these projects.

2.4 Site: Preston New Road, UK

The first hydraulic fracturing well in the UK, Preese Hall (PH-1) was drilled near Blackpool in April 2011. The injection activities in PH-1 led to induced seismicity up to M_L 2.3

(Clarke et al. 2014). The felt seismicity resulted in the suspension of hydraulic fracturing activities at Preese Hall. Following this, the UK Department of Energy and Climate Change (DECC) requires operators to implement the TLS as a risk mitigation strategy for induced seismicity in 2013 (Schultz et al. 2020).

The Preston New Road (PNR) site is a hydraulic fracturing site at Preston New Road, Lancashire, in the North West of England (Figure 2.4). The PNR site targets the Carboniferous Bowland Shale units (Clarke, Verdon, Kettlety, Baird & Kendall 2019). Two horizontal wells were drilled at this location; the Preston New Road 1-z well (PNR-1z) and the Preston New Road-2 well (PNR-2), both operated by Cuadrilla Resources Ltd. (Cuadrilla) (Clarke, Soroush, Wood et al. 2019). The surface coordinates of the site given by the PNR-1z Hydraulic Fracture Plan are approximately 432,749.50 m North and 337,433.54 m East (using the British National Grid- Ordnance Survey Great Britain, 1936) or Latitude 53° 47' 14.2827" N, Longitude 02° 57' 04.0278" W (Cuadrilla Bowland Ltd. 2018).

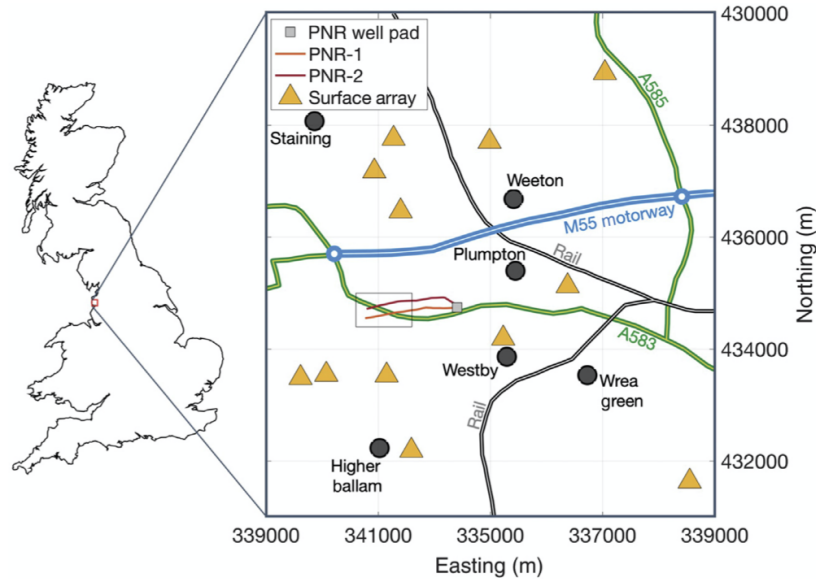


Figure 2.4: The Preston New Road (PNR) hydraulic fracturing site at the North West of England with coordinates from the United Kingdom (UK) Grid Reference (from Clarke, Verdon, Kettlety, Baird & Kendall (2019)). The map shows major roads, nearby villages and the positions of the PNR-1z and PNR-2 wells, the PNR well pad and the surface stations in the area.

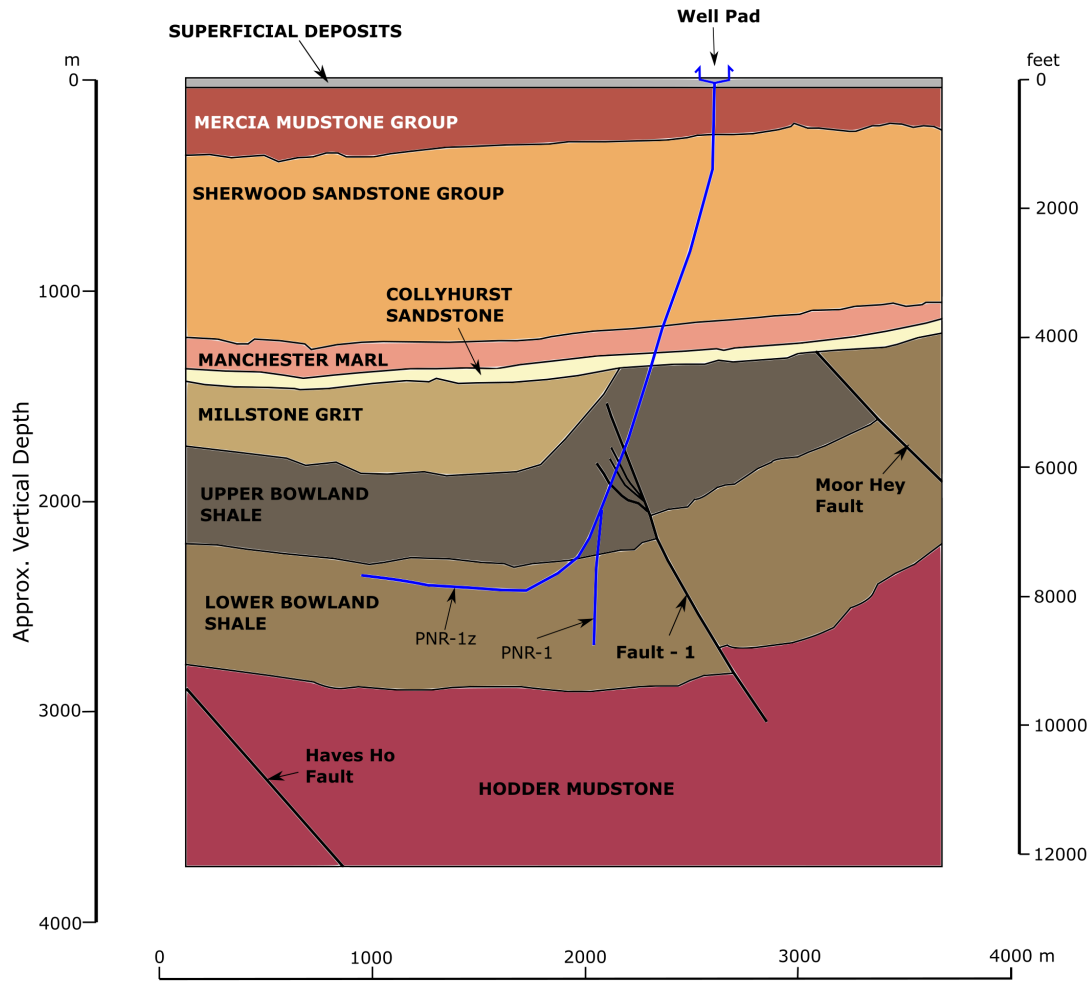


Figure 2.5: The geological cross section at the Preston New Road site showing the wells (PNR-1 and the horizontal well PNR-1z), lithology and the structural features on site (adapted from Cuadrilla Bowland Ltd. (2018)).

Figure 2.5 shows the lithology at the Preston New Road site. Within the Bowland Basin, the Lower and Upper Bowland Shale units are deposited on the Carboniferous Hodder Mudstone. The Bowland Shale units are overlaid partly by the Millstone Grit Group, Collyhurst Sandstone and capped by the Manchester Marl Formation (Anderson & Underhill 2020). The Manchester Marl formation is then overlaid by the Sherwood Sandstone group, Mercia Mudstone group and superficial deposits. Figure 2.5 also display pre-Permian regional reverse faults (the Haves Ho Fault and the Moor Hey Fault) intersecting the Bowland Shale units. At the PNR site, Cuadrilla Bowland Ltd. (2018) mapped a south-east dipping local reverse fault (Fault - 1) that intersects the Hodder

Mudstone and both the Bowland Shale units. As the horizontal well bore is anticipated to encounter a small number of faults, Cuadrilla Bowland Ltd. (2018) implemented downhole microseismic monitoring and the TLS as risk mitigation strategies for Preston New Road.

2.5 Methods of Earthquake Phase Detection

As microseismic monitoring spans across a number of important industrial geo-energy projects, it is important to have a robust method of seismic event and phase detections. In the next few sections, I review the traditional (2.5.1) and more recent methods of earthquake detection (2.5.2, 2.5.3, 2.5.4, 2.5.5), introduce the use of deep learning for this purpose (2.6.1) and describe the pre-trained deep learning model used in this thesis (2.6.3).

2.5.1 STA/LTA

Traditional automated methods such as the short-term-average/long-term-average (STA/LTA) trigger have been used widely in microseismic monitoring (Allen 1978, Liu & Zhang 2014). The STA/LTA algorithm was developed to replicate human analyst performance in P wave arrival detections on single seismometer traces (Allen 1978). The STA/LTA uses the characteristic function (CF) of a single seismic trace (Allen 1978, 1982). The most widely used characteristic function is the absolute value function defined as

$$CF(i) = |Y(i)| \quad (2.1)$$

where Y is the amplitude of the seismic trace and i is the i^{th} sample of the waveform (Allen 1982).

To make the algorithm more sensitive to sudden changes in amplitude and frequency, Allen (1978) updated the function to include the squared difference in amplitude of the i^{th} sample and the previous sample, multiplied by a weighing constant. They define this CF as

$$CF(i) = Y(i)^2 + K(Y(i) - Y(i - 1))^2 \quad (2.2)$$

where Y is the amplitude of the trace, i is the i^{th} sample of the waveform and K is a weighting constant that varies with sample rate and station noise characteristics.

The CF is calculated for the STA and the LTA window through time. The durations of both the STA and LTA windows are user defined. The STA window is sensitive to seismic events whereas the LTA window provides a measure of the background noise level (e.g. STA of 1 second and LTA of 2 seconds for microseismic applications) at that time (Trnkoczy 2009). An event is declared when the STA/LTA ratio exceeds a threshold specified by the user.

The STA/LTA method is highly valued for earthquake event detection because of its general applicability, efficiency and the ability of the trigger to run on low computational power. However, the STA/LTA algorithm may result in a high number of false positives (false triggers) or false negatives (missed events) due to non-impulsive events, low signal-to-noise ratio (SNR), overlapping events, or incorrectly specified trigger parameters (i.e. trigger threshold, STA or LTA window duration) (Trnkoczy 2009, Yoon et al. 2015, Bergen & Beroza 2018, Mousavi et al. 2019).

2.5.2 Coalescence Microseismic Mapping (CMM)

The coalescence microseismic mapping (CMM) method is a multi-station simultaneous detection and location method that can be very effective for microseismic monitoring (Tarasewicz et al. 2012, Bradford et al. 2013, Smith et al. 2015). In a review, Li et al. (2020) explain that waveform-based location methods can be described in four different categories; partial waveform stacking (Kao & Shan 2004, Grigoli et al. 2013), time reverse imaging (Gajewski & Tessmer 2005, Larmat et al. 2008), wavefront tomography (Duveneck 2004, Bauer et al. 2017) and full waveform inversion (Michel & Tsvankin 2014, Igonin & Innanen 2018). The CMM method is an example of a partial waveform stacking technique. This method has been applied successfully in various applications: to monitor volcanic activity (Tarasewicz et al. 2012) and hydraulic fracturing operations (Bradford et al. 2013). Schlumberger produced a catalogue of event origin times from the

hydraulic fracturing activities at the PNR site using code based on the CMM method.

The CMM method is a form of continuous grid-search event detection and location method using STA/LTA ratios, stacking the migrated signals across multiple stations and using a Gaussian approximation on each pick as an estimate for arrival time uncertainty (Drew et al. 2005, 2013, Pugh et al. 2016). The CMM method applies a user-defined time window to search across the data continuously. This method uses both traveltime inversion and imaging for event locations from continuous seismic waveforms. Drew et al. (2013) explain that the CMM method first constructs a travel times look-up table from each gridpoint to each receiver station using a pre-defined velocity model. Figure 2.6 illustrates that the method then uses the traveltimes look up table to migrate onset functions back to each grid node within a 3-D subsurface grid at each user-defined time step.

Subsequently, continuous seismic waveforms across multiple stations are transformed into STA/LTA signals for each time window. Thus, the performance of this algorithm strongly depends on the selection of STA/LTA parameters (i.e. the STA and LTA window duration) and the user-defined window duration. Inappropriate parameter settings may cause the algorithm to pick more false positives or miss real events. Within the time window, the CMM method takes the maximum coalescence value at each time step. If the user selects a long time window, the method can potentially miss the energy arrivals of smaller events within this time step. If a short time window is used, executing the algorithm will be more computationally intensive.

After migration, Drew et al. (2013) then use a Gaussian approximation to transform the STA/LTA signals into probability density functions of the arrival times on each node in the grid. As the STA/LTA signals are proportional to the SNR of the arrival times (Drew et al. 2013), these signals should coalesce on grid nodes where events are located. This exhaustive grid-search results in a catalogue of origin times along with their respective earthquake locations.

Overall, the CMM method is a robust method for microseismic earthquake detection with the added advantage of event hypocentre locations. It is of interest to see if we can use deep learning neural networks to replicate or even elevate this level of robustness for hydraulic fracturing induced microseismicity in a shorter computational runtime. A

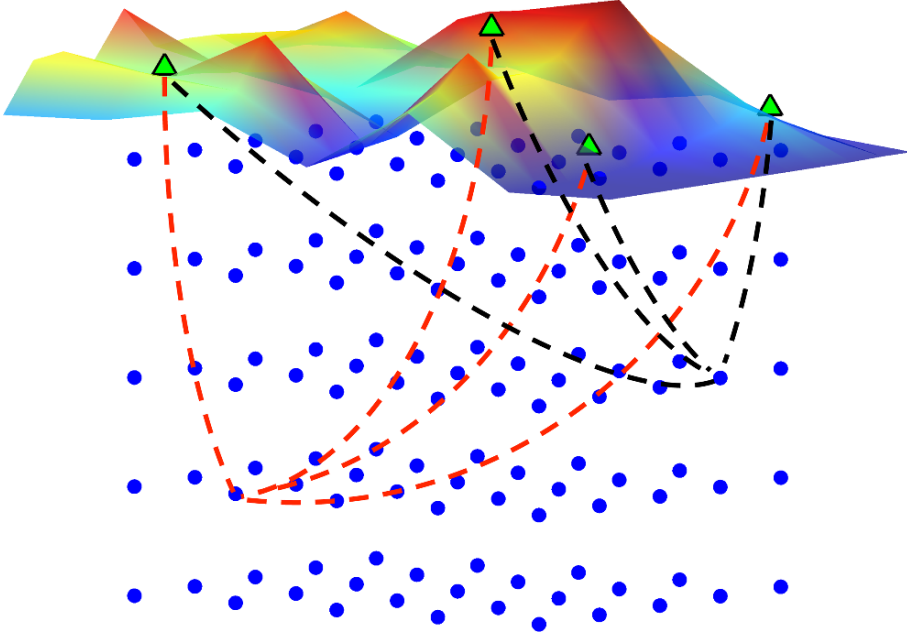


Figure 2.6: Schematic from Drew et al. (2013) of the 3D subsurface traveltime grid with the blue dots as subsurface nodes and the green triangles as surface stations. The CMM method uses the velocity model and traveltime look-up tables to backmigrate onset functions from the seismic data (all 3 components) on each node from each station for each time step. The onset functions are then stacked at each node point.

significant disadvantage of the CMM method is that it is a computationally intensive program. Deep learning neural networks have the potential to overcome this barrier and provide faster event detections than a grid-search method like the CMM method.

2.5.3 Template Matching (Waveform Cross-correlation / Match Filtering) and Subspace Waveform Detection

Template matching can be a robust technique for induced earthquake detection- this is exemplified by detecting wastewater disposal-induced seismicity (Skoumal et al. 2014) and hydraulic fracturing-induced seismicity (Song et al. 2010). Template matching (also known as match filtering) is a waveform cross-correlation method with known examples of earthquake events (Gibbons & Ringdal 2006). Yoon et al. (2015) describes this method as a ‘one-to-many’ search comparison method that cross-correlates windows of the continuous waveform with a template waveform to obtain normalised correlation coefficients. Events are detected when a waveform has high similarity with the template

waveform. Cross-correlated events that occur close together (within hundreds of meters) are thought to have similar focal mechanisms (Schaff & Waldhauser 2005).

However, template matching requires a prior seismic catalogue and templates near the observation site. Locations that are less seismically active may not have the required templates. Even if the area of interest is seismically active, a catalogue may not exist, or only the very largest events are captured by regional or global networks. The method also suffers from a detection bias, picking similar events and missing those with different focal mechanisms or locations.

Subspace detection is a generalised template matching method that uses Singular Value Decomposition (SVD) to capture the general features of waveforms efficiently (Harris 2006, Barrett & Beroza 2014). SVD represents a design set of waveforms (or a template) from an existing event catalogue as left singular vectors that contain hypocentres and possess improved SNRs compared to the actual template (Barrett & Beroza 2014, Bestmann 2020). Bestmann (2020) found that the subspace detector detects more events than classical methods like the STA/LTA. Subspace detectors are also able to detect non-repeating, more variable earthquakes than in template matching (Harris 2006). The drawbacks to this method are similar to those in template matching- the subspace detector still requires a prior knowledge of the earthquake events; i.e. an event catalogue to make their templates in vector space. Consequently, earthquake events that have no similar templates will be completely missed by this method.

2.5.4 Autocorrelation

Another earthquake detection method that uses waveform correlation is the autocorrelation method (Brown et al. 2008). Instead of requiring pre-existing templates, the autocorrelation method searches for clusters of similar waveforms and can be considered an unsupervised template-matching method. Setting the search window across continuous seismic data to be as short as the duration of an earthquake event, the continuous waveforms are divided into overlapping windows. Cross-correlation is performed on all possible pairs of windows and the windows with high correlation coefficients are grouped together to form a cluster (Yoon et al. 2015).

Compared to the STA/LTA, autocorrelation detects more events and, unlike template matching, autocorrelation can detect earthquakes with different focal mechanisms (Bergen & Beroza 2018, Yoon et al. 2015). However, this method is very computationally intensive and inefficient, making it unsuitable for most real-time earthquake monitoring endeavours or detection over large continuous datasets.

2.5.5 Fingerprint and Similarity Thresholding (FAST)

FAST is an adapted data mining algorithm inspired by data mining techniques used for audio recognition, namely, the Waveprint wavelet-based audio fingerprinting algorithm (Baluja & Covell 2008, Yoon et al. 2015). Instead of comparing waveforms, the FAST algorithm first undergoes feature extraction: it splits a single continuous waveform into windows of features in the time and frequency domains (or spectral images). For each window, it applies a 2D Haar wavelet transform and is subjected to data compression. Subsequently, binary fingerprints are generated for a large database and is used in a similarity search (Yoon et al. 2015). The search uses ‘Locality Sensitive Hashing’, which groups similar high-probability waveforms into hash buckets and allows efficient search in large datasets (Slaney & Casey 2008, Yoon et al. 2015). Correlations are then performed exclusively on similar events to reduce the runtime of the algorithm (Skoumal et al. 2016).

The advantage of the FAST algorithm is that it is significantly more computationally efficient than autocorrelation and template matching. Additionally, this algorithm does not need existing event templates and can detect earthquakes with different focal mechanisms. Yoon et al. (2017) explored the method’s trade-off between speed and accuracy and showed that the algorithm can detect more false positives (non-earthquake signals) or miss more events than detected using template matching. This is because information is lost through feature extraction, when waveforms are transformed into binary fingerprints (Yoon et al. 2017). FAST is a single-station event detector. Bergen & Beroza (2018) designed a robust multi-station application of the FAST algorithm in an effort to improve the algorithm’s sensitivity and limit the number of false positives.

2.6 Deep learning in Earthquake Phase Detection

Deep learning is a sub-field of machine learning that allows computational models called deep neural networks to learn generalised representations of input data (which could be vast datasets) with multiple levels of abstraction (LeCun et al. 2015). Deep learning is especially good at tasks involving images (e.g. the detection and recognition of visual objects). Most notably, the success of AlexNet (Krizhevsky et al. 2017) in the 2012 ImageNet competition led the renaissance of deep learning models into the mainstream technology we are seeing today. Within the context of supervised learning, Artificial Neural Networks (ANNs) can be used as the function that relates inputs to outputs. ANNs are typically a composition of linear and nonlinear functions. ANNs tend to have many parameters (or weights), sometimes in their millions, and these parameters must be learnt from the data (LeCun et al. 2015). This is done by first specifying a loss function (also known as a cost/error function) to estimate the errors of the model so the weights may be tuned to minimise these errors when mapping from input to output, and then using gradient based optimisation to obtain a set of parameters that reduce this loss function (Patterson & Gibson 2017). One of the cornerstones of this field is the backpropagation algorithm which allows us to efficiently compute this gradient (Chauvin & Rumelhart 1995). LeCun et al. (2015) explains that deep learning algorithms can ‘learn’ representations of the data by using backpropagation to adjust the parameters on each layer (a set of nodes/neurons) of a neural network. Deep learning has found much success in applications such as speech recognition (Deng et al. 2013), medical image analysis (Shen et al. 2017) and natural language processing (Deng & Liu 2018).

Seismology has also experienced major advances as a result of deep learning methods. In a 2019 special issue of the *Seismological Research Letters* journal, Kong et al. (2019) reviewed the current status and potential of artificial intelligence for applications in earthquake detection (tectonic and regional) and seismic imaging of the Earth. This review documented substantial advances in event detection and phase picking that deep learning methods have enabled. They note that convolutional neural networks have been developed and trained to recognise generalised representations of seismic phases (Ross, Meier, Hauksson & Heaton 2018, Zhu & Beroza 2019), pick P-wave arrivals and determine its first-motion polarity (Ross, Meier & Hauksson 2018). Recognition of seismic phases helps create an earthquake catalogue for areas where there is no existing catalogue or template for event detection methods such as template matching

or autocorrelation. Improved picking of P-wave arrivals and the determination of first motion polarity increases the number of detailed earthquake focal mechanisms in an area.

2.6.1 Convolutional Neural Networks (CNN)

Bergen et al. (2019) argued in a widely cited review paper that the class of deep learning models, convolutional neural networks (CNNs), in particular, can empower observational seismology to efficiently detect events in a vastly larger dataset than is achievable with classical methods. CNNs are ANNs that utilise the convolution operation within the function. CNNs typically have two main components; convolutional layers for feature extraction and fully connected neural networks for image classification. Convolutional layers involve convolving the data using filters that slide across the data in order to extract important features within the input data (feature extraction). The output from the convolutional layers is then flattened into a 1-D feature vector and will be the input for the fully-connected neural network (FCNN). In the training step, weights (a set of learnable nonlinear filters) are fine-tuned and computationally optimised by repeated convolution and pooling operations of the input data to extract features in the training dataset (waveforms) that will aid seismic classification (Woollam et al. 2019). The FCNN then uses the trained weights from the training phase to classify features found by the convolutional layers.

CNNs have been developed and trained to learn generalised representations of seismic phases and pick phase arrival time from continuous data (Ross, Meier, Hauksson & Heaton 2018, Perol et al. 2018, Wu et al. 2018, Wang et al. 2019, Woollam et al. 2019). CNNs treat seismic phase detection as a supervised object classification and 2D image recognition task. To be able to classify seismic phases, the neural network only requires training with large datasets of labelled seismic phases. In seismology, large labelled datasets are readily available from decades of supervised earthquake monitoring, automatic and analyst-labelled phase arrival picks.

The advantages of using CNNs for automated phase detection are their relatively fast speed in processing large datasets in comparison with state-of-the-art detection algorithms, and the sensitivity of a CNN phase picker in detecting more earthquake events than the classical STA/LTA trigger (e.g. Ross, Meier, Hauksson & Heaton (2018),

Woollam et al. (2019)). Compared to similarity based algorithms, CNN phase detectors are more computationally efficient and have less bias in event detection as they do not rely on waveform similarity to set templates. The CNN model should also be able to return results faster than grid-based earthquake event search algorithms. Perol et al. (2018) showed that their CNN produced faster results than both the autocorrelation and FAST algorithm.

The potential of speeding up event detection in real-time processing with deep learning phase detection models gives the advantage of adding more time to react to human-induced seismicity and thus improving the real-time management of risk and hazard. In addition, if the CNN can detect more induced events, the CNN will provide a more complete catalogue that may fill gaps in the spatio-temporal evolution of induced seismicity, and may even provide unprecedented views into geological fault zone structures. For these reasons, I have chosen to utilise CNNs for the detection of hydraulic-fracturing induced seismicity.

There are certain challenges in processing and detecting microseismic events- one of these challenges is the low SNR (Sabbione & Velis 2013). Sabbione & Velis (2013) highlight that accurate picking of phase arrivals is difficult because microseismic events have low SNR and thus these microseismic phases could easily be buried within the noise. Even so, Zhang et al. (2018) and Huang et al. (2018) have both shown success in detecting microseismicity using CNNs trained with microseismic event datasets.

A drawback of CNNs is that they require a large training dataset (e.g. thousands or millions of labelled seismic samples) to achieve good generalised phase picking. To sidestep this limitation, I have decided to utilise a pre-existing phase detecting CNN that has already been trained with millions of labelled seismic phases and test its general applicability to detect hydraulic fracturing-induced microseismicity. I will be observing the robustness of a deep learning seismic phase detection model against the CMM method that was used to collect the catalogue in the PNR site. Additionally, I will compare computational runtimes of the CNN with other earthquake detection methods (i.e. autocorrelation, FAST). I will also evaluate whether a deep learning algorithm that was pre-trained on regional catalogued earthquakes collected from different tectonic settings could perform just as well for microseismic monitoring. I will check if the CNN needs retraining, whether it performs well off-the-shelf, or if it would be better utilised

alongside current detection methods to make initial quick detections for real-time monitoring.

2.6.2 Transfer Learning

Transfer learning is taking the learnable filters (or weights) and network architecture from a pre-existing CNN already trained on some dataset and using them as the base for the retraining of the CNN (Shin et al. 2016). In transfer learning, the parameters in the CNN layers are fine-tuned for better performance. Rather than starting with random weights from scratch, transfer learning builds on the optimised weights from initial training (Gopalakrishnan et al. 2017).

Transfer learning of CNNs is commonly used in the medical field for recognising objects in medical images. For example, CNNs are in the detection of breast cancer on mammograms (Huynh et al. 2016) as well as in the automatic detection of COVID-19 on x-ray images (Apostolopoulos & Mpesiana 2020). Tajbakhsh et al. (2016) found that models that undergo transfer learning using a large dataset have shown to either outperform or perform as well as the models trained from scratch and off-the-shelf models. Huang et al. (2017) also suggest that fine-tuned models perform better than off-the-shelf models even with a limited (smaller) dataset. Huang et al. (2017) recommend strong consideration of whether the features learnt from the off-the-shelf model are adequate or if fine-tuning of the parameters is needed for good performance.

If the off-the-shelf model does not perform well or if it can not detect hydraulic fracturing-induced seismicity at all, transfer learning will be required to improve model performance. In this thesis, I will explore whether transfer learning is necessary for microseismic event detection on a downhole geophone dataset or if the results from the off-the-shelf CNN are adequate.

2.6.3 The Generalised Seismic Phase Detection (GPD) model

The Generalised Seismic Phase Detection (GPD) model developed by Ross, Meier, Hauks-son & Heaton (2018) was chosen for this project because it is the most extensively trained CNN on earthquake phases. Other seismic phase-detecting CNNs by Zhu &

Beroza (2019), Perol et al. (2018) and Woollam et al. (2019) also exist but are not as extensively trained as the GPD model. The GPD model was trained and validated on 4.5 million 100 Hz three-component seismic traces that were recorded by the Southern California Seismic Network (SCSN) (Ross, Meier, Hauksson & Heaton 2018). More specifically, Ross, Meier, Hauksson & Heaton (2018) used 1.5 million seismograms each for the P wave picks, S wave picks and noise windows. Each of these seismograms were 400 samples long (i.e. 4 seconds for 100 Hz data). The training set of the model had the P waves and S waves at the centre of their respective pick and the noise windows were chosen 5 seconds before each P wave pick.

The training and validation dataset of the GPD model were exclusively taken from regional earthquakes in Southern California. Ross, Meier, Hauksson & Heaton (2018) developed the model to detect local and regional earthquakes ($M > 0$) with magnitudes relatively larger than most of the microseismic events catalogued at the PNR site. They trained this model with a magnitude range of $-0.81 < M < 5.7$ events using seismograms from regional surface stations. Ross, Meier, Hauksson & Heaton (2018) displayed the adaptability of the GPD model by applying it to new datasets: during the 2016 Bombay Beach, California swarm and the 2016 M_w 7.0 Kumamoto earthquake in Japan. They showed that the GPD model can detect events of different magnitudes and different locations that are not represented by the training data (Figure 2.7). The GPD model working successfully on data not represented during training suggests the possibility of the model detecting smaller events (induced seismicity) in the downhole dataset obtained from hydraulic fracturing activities in PNR.

2.7 Objectives

The primary objectives of this thesis are:

Objective 1: Model classification performance

- To assess the single-station performance of an off-the-shelf (OTS) deep learning phase detection model that was extensively trained on regional earthquakes to detect hydraulic fracturing induced earthquakes (microseismicity) on the continuous high-frequency downhole data from the Preston New Road site.

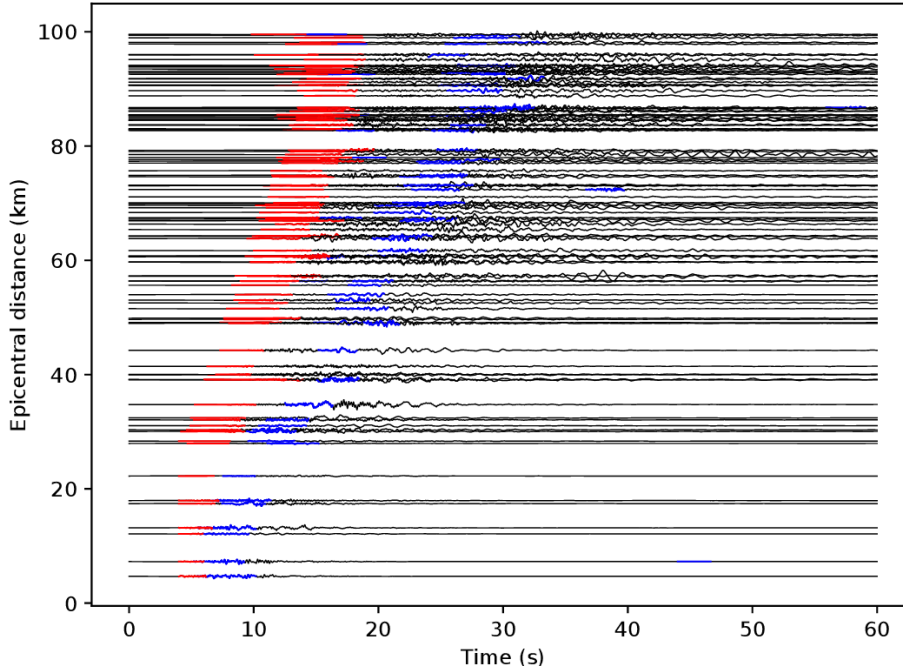


Figure 2.7: The GPD model successfully detecting the 2016 M_w 7.0 Kumamoto earthquake in Japan (from Ross, Meier, Hauksson & Heaton (2018)). This displays the applicability of the GPD model to detect earthquakes that are outside the region and moment magnitude range of the training dataset.

- Evaluate the ability of the OTS model to detect and classify phase arrivals (P/S) and noise on single stations under various conditions, including large magnitude ($M_w > 0$) events, small magnitude events ($M_w < -2$), mid-range magnitude events ($-2 < M_w < 0$) and events when there is high seismicity during injection.

Objective 2: Multi-station event detection workflow

- Construct an event detection workflow that associates phases and locates events using the phase picks from the GPD model.

Objective 3: Multi-station event catalogue comparison

- Conduct a multi-station method comparison of the GPD model catalogue with the event catalogue of the previous detection and location method (i.e. the CMM method).

- Assess whether the OTS model can detect more events during periods of interest; event onset (when events occur) during high injection rates and during high magnitude induced events ($M_w > 0$).
- Periods that include larger magnitude events ($M_w > 0$) are of interest to check if the model is able detect more aftershock events or if the model can detect any type of precursory microseismic activity.
- Periods of an event onset during high injection rates were selected to see if the model can detect as many events or find additional events during high seismicity.
- Determine if the detection runtime of the GPD model is faster and computationally more efficient than the previous detection method (i.e. the CMM method).

Chapter 3

Data

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

Injection operations on the PNR-1z well were monitored by Cuadrilla Bowland Ltd. to obtain the PNR-1z continuous downhole dataset. Cuadrilla monitored the hydraulic stimulations that took place from 15 October to 17 December 2018 (Clarke, Verdon, Kettlety, Baird & Kendall 2019). Figure 3.1 shows the location of the two wells at the PNR site. At the Preston New Road site, the PNR-1z well has 41 'stages' along the well where fracturing fluid is pumped through to create the hydraulic fractures. Cuadrilla planned to stimulate each of the 41 stages from the toe to heel on PNR-1z whilst PNR-2 was equipped with a downhole array of geophones, and was used as an observation well during injection for microseismic monitoring. However, only 17 stages were injected because Cuadrilla had to skip stages where hydraulic fracturing would intersect with a pre-existing North-East orienting fault within the Bowland shale formation (Clarke, Verdon, Kettlety, Baird & Kendall 2019). Cuadrilla proceeded to skip the stages up to stage 37 and resumed injection operations on stages 37-41 (closer to the heel of the well). For our study, we use data collected from the downhole array of geophones on PNR-2 during hydraulic stimulation operations on the PNR-1z well.

The continuous dataset was recorded with a sampling frequency of 2000 Hz by 24 down-the-borehole Avalon Geochain Slimline 15 Hz geophones in the observation well, PNR-2 from 8 October to 18 December 2018. The downhole array had an average spacing of 30 meters between each geophone. Between those dates, two periods of hydraulic stim-

ulation operations were carried out: from 15 October to 2 November and subsequently from 8 to 17 December 2018. A hiatus in injection operations took place from 3 November to 7 December 2018.

The data files containing the PNR-1z continuous seismic traces are all in SEG-Y format. One SEG-Y file contains the 3 component traces of each station, resulting in a total of 72 traces per SEG-Y file. Each SEG-Y file is 9.2 MB and contains 16 seconds of recorded seismic data. Thus, this continuous dataset has a total size of 3.92 TB. As the continuous seismic data are separated into multiple 16-second SEG-Y files, continuous data can be fed into the model by combining these SEG-Y files. An hour of continuous seismic data is made by stitching each file into a $24 \times 7,200,000 \times 3$ matrix that combines 24 stations, 7,200,000 data points (as number of data points = sampling frequency \times duration of continuous data, i.e., 2000 Hz \times 3600 seconds for one hour) and the 3 components. The resulting matrix of continuous data can then be used as input for the GPD model.

Figure 3.1 shows that the stations were set up from the shallow part of the well (the heel) down towards deeper part (the toe) of the well (numbered station 1 to 24). From visually inspecting the continuous downhole waveforms, we observe that most events above moment magnitude magnitude M_w -0.237 display amplitude clipping on at least one station in the seismograms (Figure 3.2). Amplitude clipping occurs especially in stations closer to the toe of the well (e.g., stations 22, 23, 24) because they are closer to the event locations and thus have larger amplitudes. The waveforms with large amplitudes were strongly affected by clipping. The amplitudes of the larger events were clipped because the gain of the recording instruments was set high in order to detect more microseismic events during real-time monitoring of the hydraulic fracturing project. Figure 3.3 clearly shows event clipping on the largest catalogued event ($M_w = 1.155$). Assuming that all events down to $M_w = -0.237$ have been affected by amplitude clipping, up to 133 events have clipped amplitudes on at least one station. This could be a potential challenge to event detection using the trained GPD deep learning model. However, Ross, Meier, Hauksson & Heaton (2018) have shown that their network primarily identifies the onset of coherent phase arrivals so this should not be an issue for the deep learning phase detector even if part of the amplitude has been clipped.

3.2. PRESTON NEW ROAD-1Z (PNR-1Z) EVENT CATALOGUE, VELOCITY MODEL AND INJECTION DATA

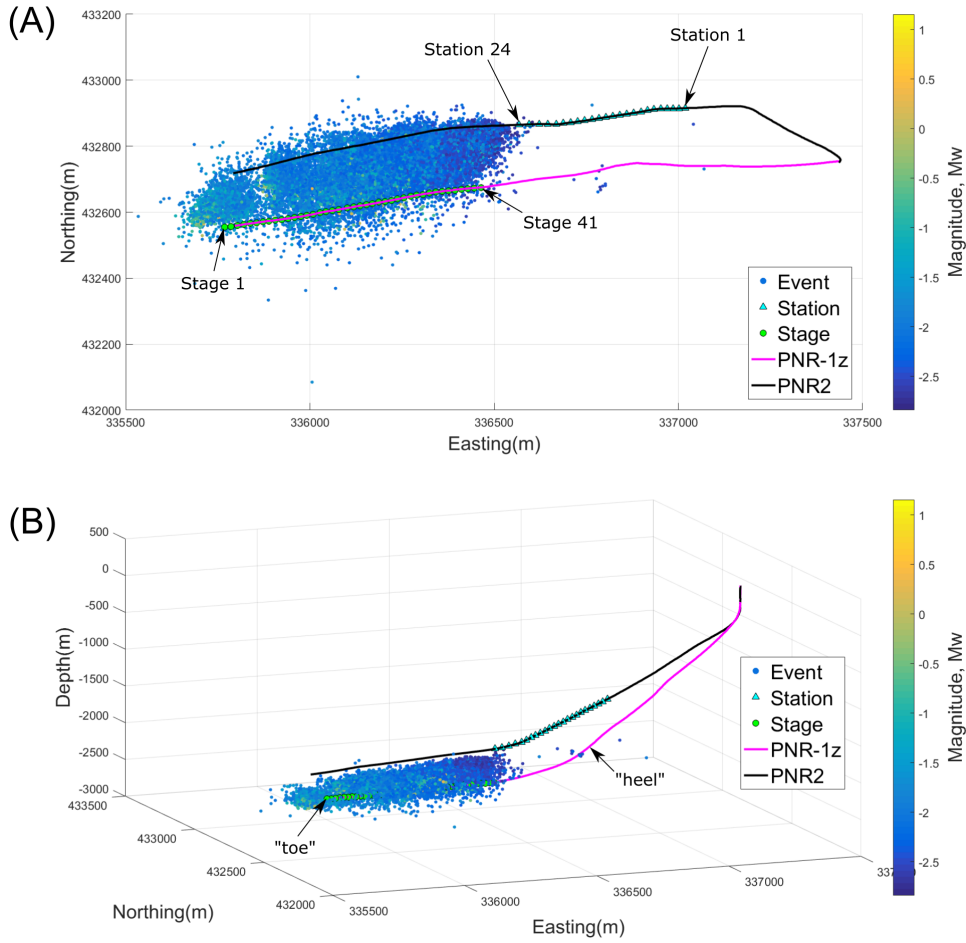


Figure 3.1: Spatial maps of all the catalogued events, the geophone array, well stages and wellpaths of PNR-1z (magenta) and PNR2 (black). (A) Plan view and (B) 3D plot. The x, y and z-axis show the Easting, Northing and depth in metres, respectively.

3.2 Preston New Road-1z (PNR-1z) Event Catalogue, Velocity Model and Injection Data

Schlumberger processed the PNR-1z continuous downhole dataset in real-time (Clarke, Verdon, Kettlety, Baird & Kendall 2019) using the Coalescence Microseismic Mapping (CMM) method (Drew et al. 2013) to create an event catalogue. Real-time monitoring using the CMM method was only possible in the case of the PNR-1z site because the continuous data were processed on a supercomputer during injection operations (Verdon 2020). This is impractical and computationally costly when compared to applying a deep learning model on the data for rapid event detection on a CPU or a single GPU.

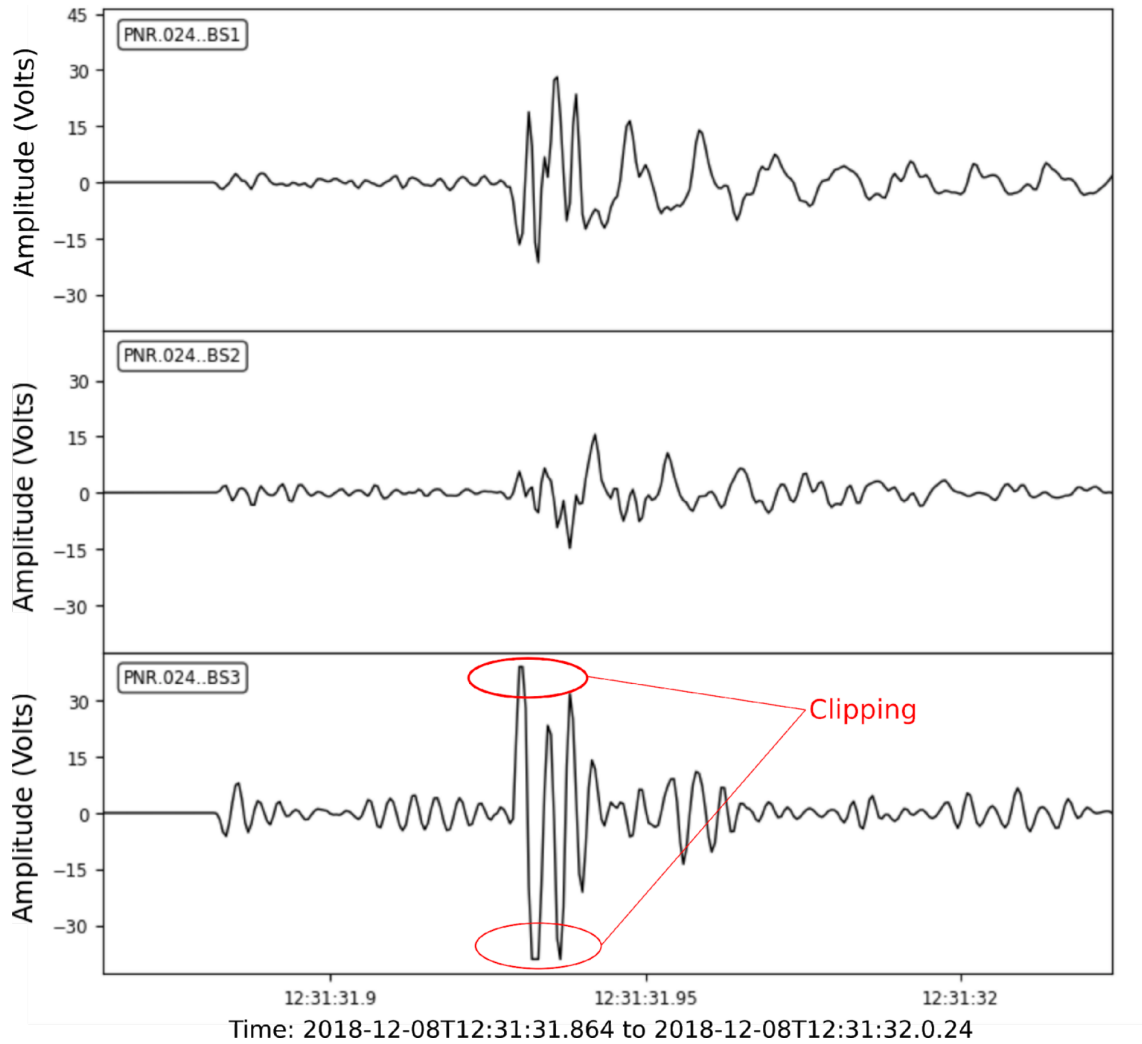


Figure 3.2: Amplitude clipping on the 3-component seismogram for the $M_w = -0.237$ event on station 24.

The PNR-1z event catalogue lists a total of 38,452 events. This event catalogue will be used as ground truth in the comparison of the GPD model’s event detection performance with the CMM method (see subsection 4.3: multi-station comparison in the Methods section). The event catalogue contains events within a magnitude range of $-2.839 \leq M_w \leq 1.155$. The upper-end of the estimated event magnitudes are likely incorrect or have been underestimated as they have been affected by amplitude clipping (the amplitudes were saturated). Event magnitudes of small earthquakes (microseismicity) are difficult to accurately estimate (Kendall et al. 2019). Event magnitudes from the surface stations at Preston New Road have an accuracy of ± 0.5 , downhole measurements would have slightly better accuracy but it would still be closer to ~ 0.5 than 0.1 (Butcher 2021).

3.2. PRESTON NEW ROAD-1Z (PNR-1Z) EVENT CATALOGUE, VELOCITY MODEL AND INJECTION DATA

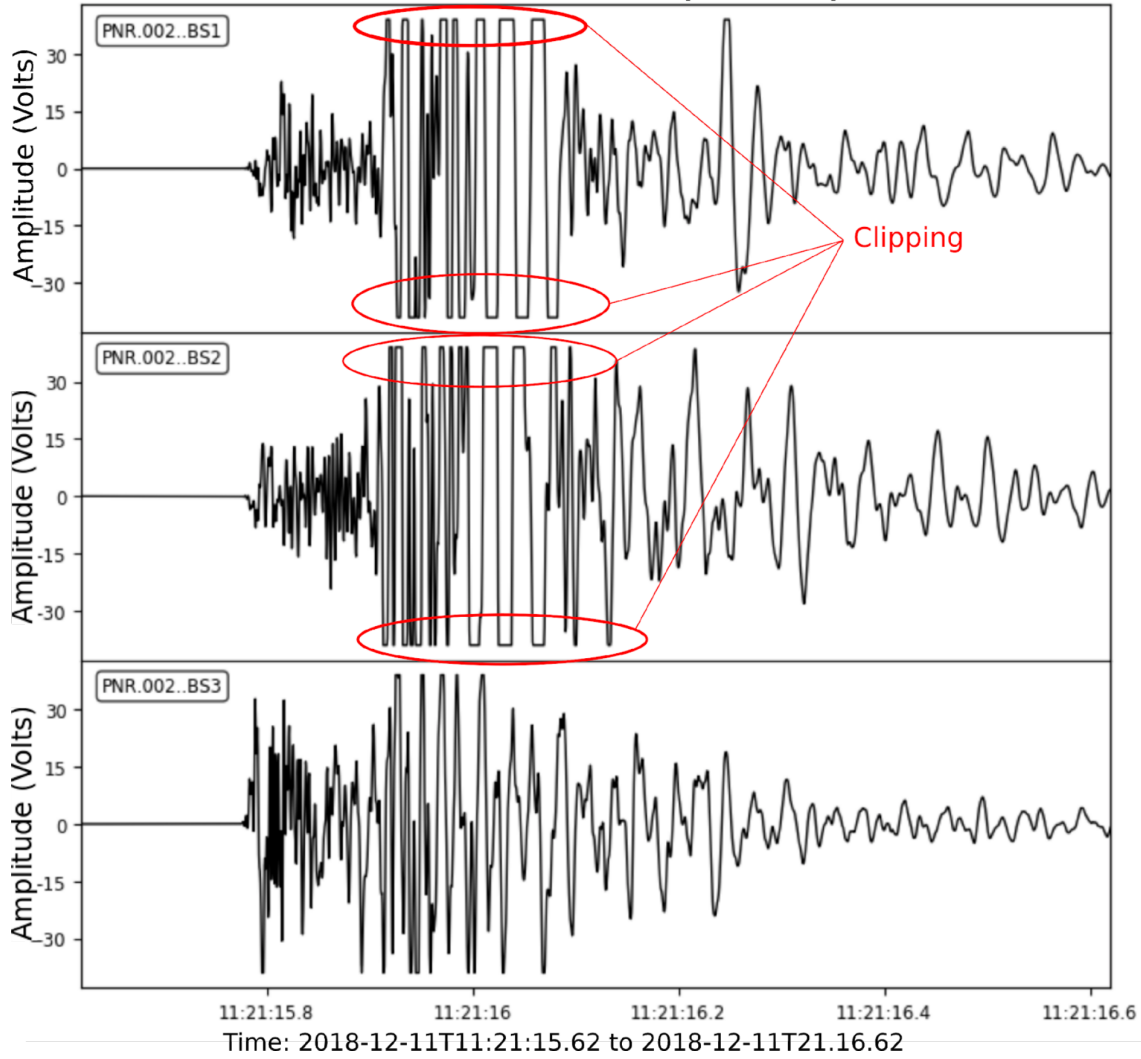


Figure 3.3: Amplitude clipping on the 3-component waveforms of the largest magnitude event, $M_w = 1.155$ on station 2.

Operational pumping data such as the injection rates and pressure measurements are publicly available (Oil and Gas Authority 2018). A velocity model of the PNR site was also provided by Cuadrilla Bowland Ltd. (2018) (Figure 3.4). Figure 3.5 shows all catalogued event magnitudes overlaid with the injection rate data plotted against time. By overlaying the injection rate data with the event magnitudes against time, we can observe some interesting features of the seismicity from the CMM event catalogue. The rate of seismicity is shown to escalate during high injection rates (see Figure 4.4(B) in the Methods subsection 4.4.1 for a zoomed in image). From Figure 3.5, we can see that a lot of the spikes in seismicity line up along the increase in injection rates. This period

of event onset during high injection rates will be explored in the subsection 4.4.1 about periods of interest in Methods.

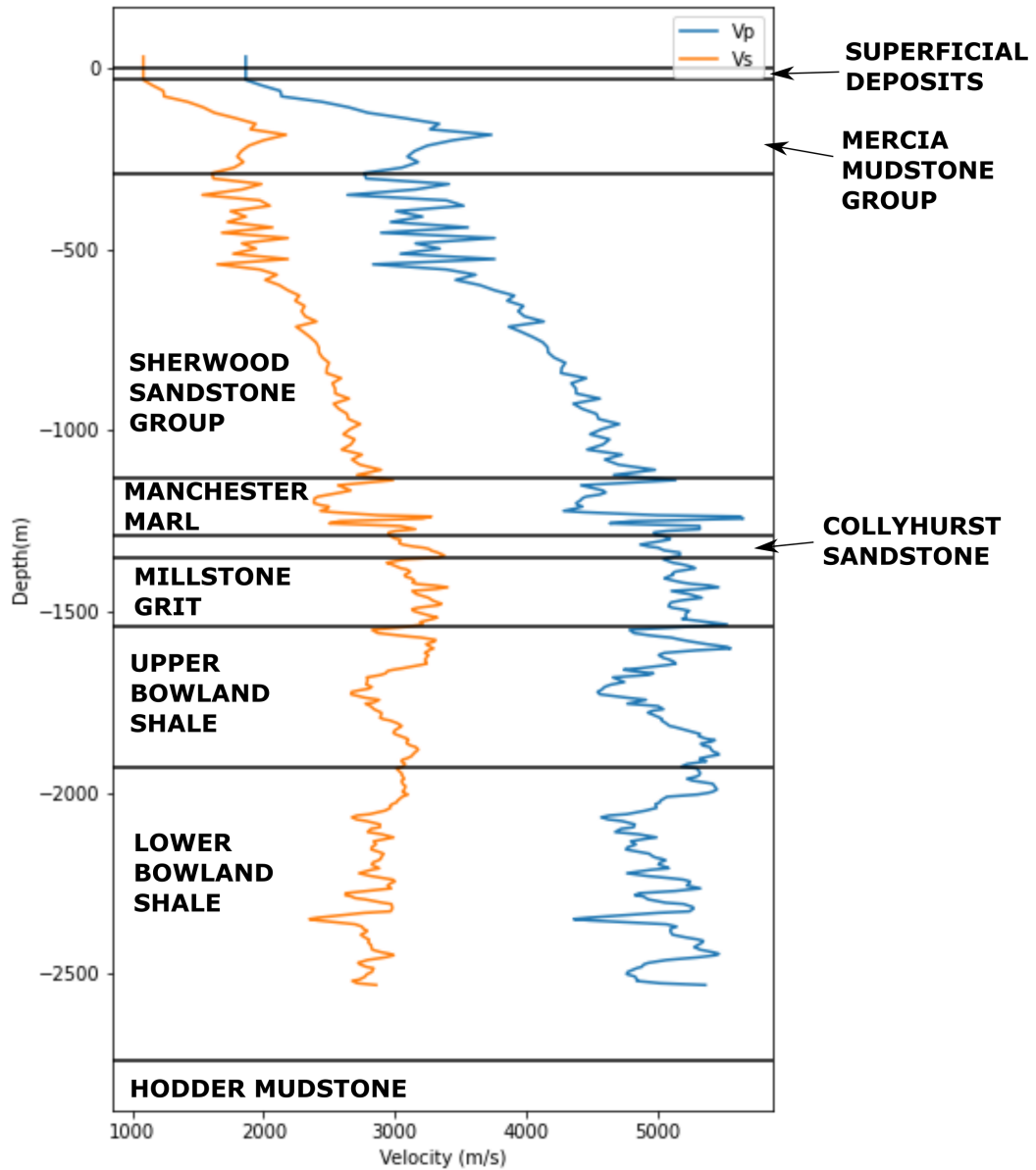


Figure 3.4: Velocity model of the Preston New Road site. Velocity of the P wave and the S wave are in blue and orange respectively. The lithology of the site has also been plotted on the figure.

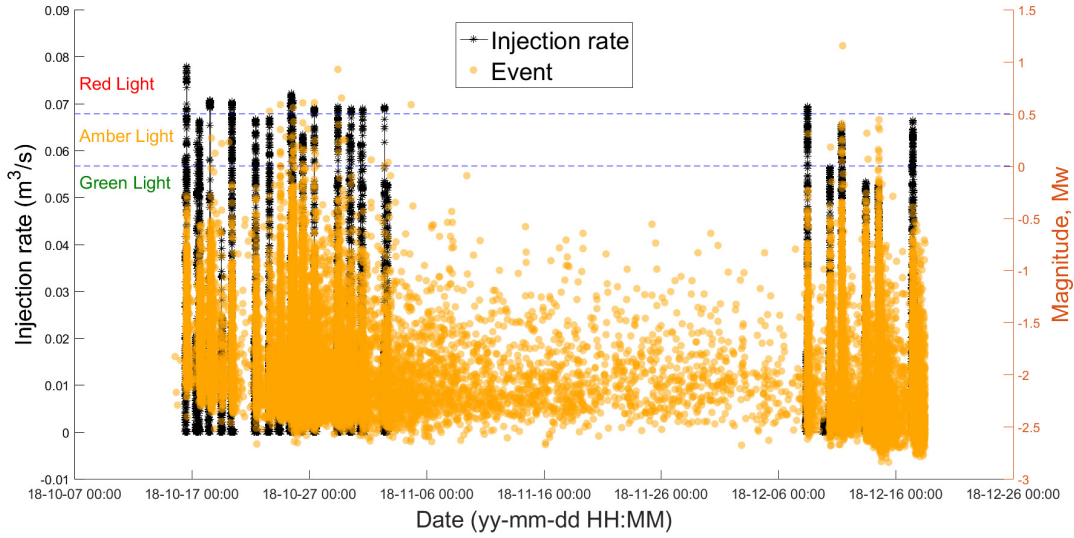


Figure 3.5: CMM catalogued events from 8 October to 17 December 2018. Injection rate in black and the events with their respective magnitude in orange. The injection hiatus took place from 3 November to 7 December 2018. The dashed blue lines separate the red ($M \geq 0.5$), amber ($0 \leq M < 0.5$) and green ($M < 0$) light stages of the UK traffic light system.

3.3 Preston New Road Event Waveforms

The PNR-1z event waveforms were also processed by Schlumberger. They have provided event associated waveforms for 30,726 earthquakes. Figure 3.6 shows the smallest catalogued event ($M_w = -2.839$) across a select number of stations (stations 2, 10, 23 and 24) along the downhole array. This figure shows that even on the event waveforms, some of the smaller magnitude events cannot be visibly seen on a number of stations closer to the heel of the well (i.e. farther from the location of the events) even if they can be seen on the traces recorded by stations closer to the toe of the well (Figure 3.6). This may be because microseismic events have low SNR and the stations closer to the heel of the well are farther from the locations of the events, and thus it is harder to visually detect the amplitudes of low SNR events in those stations. The CMM method used to create this event catalogue is a multi-station earthquake event method that uses the hypocentre of a coalescence grid where the signals coalesce at the event origin time (see Coalescence Microseismic Mapping subsection in Background and Objectives) and so, the CMM method can better detect the microseismic events across all stations. When comparing the CMM method with the GPD model, it is important to consider that the

GPD model only detects phase arrivals on single, independent stations- making it a single station method. Therefore, for smaller events (e.g. Figure 3.6), we note that it would be difficult for the GPD model or even a human analyst to detect these small events on the shallow stations.

3.3. PRESTON NEW ROAD EVENT WAVEFORMS

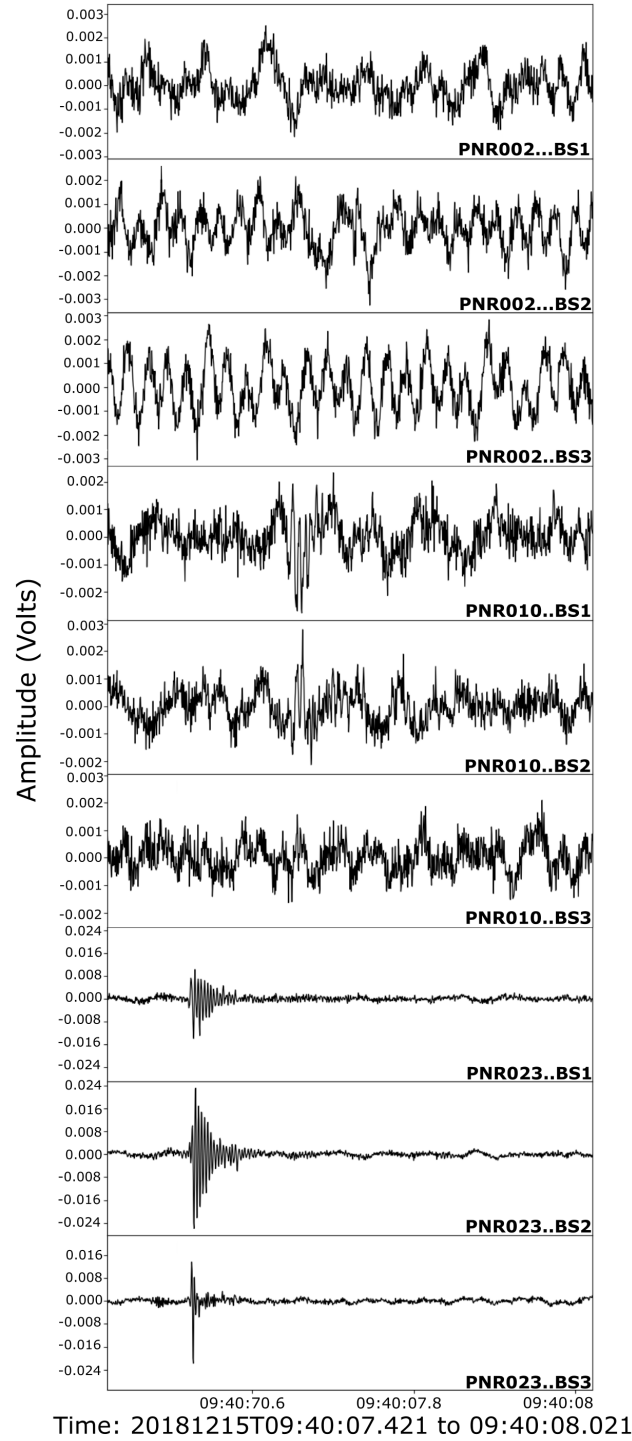


Figure 3.6: The smallest event, $M_w = -2.839$, recorded on 3-component seismograms of multiple stations (stations 2, 10, and 23). The station near the toe of the well (stations 23) displays the event more clearly because it is closer to the events. Stations near the heel (stations 2 and 10) do not seem to display clear phase arrivals.

Chapter 4

Methods

4.1 GPD model

The architecture of the GPD model consists of 6 layers: four convolutional layers for feature extraction and two fully connected layers for image classification. In feature extraction, the input data is sent through a series of layers that firstly convolves the input data with a set of weights (or filters), decimates the data so the image can be evaluated at different length scales (pooling)- and finally, an activation function is applied on the data to ensure positive output (Ross, Meier, Hauksson & Heaton 2018). The output of the feature extraction process is then concatenated into a long 1D-vector and is sent to two fully connected layers for classification (Figure 4.1). The fully connected neural network then produces a list of class probabilities of the waveforms through time. Figure 4.1(A) illustrates the workflow of the CNN for generalised phase detection from input waveforms to the output class probabilities.

For each station, the GPD model defines a sliding 400-sample window that moves through the data shifting at a pre-defined but adjustable number of samples (the n_{shift} parameter) on each component. Within each window, the model calculates three probabilities at the centre of the 400 sample window: the probability of the trace being a P phase, an S phase and a noise window. Figure 4.1(B) shows an example of the P and S probability traces derived from the continuous data. Ross, Meier, Hauksson & Heaton (2018) selected a 0.98 probability threshold for detecting regional events- this parameter was changed in this study for the use of detecting smaller microseismic phases. The model will have a trigger output when the probability of a P or S phase exceeds the user-defined probability threshold. In the case of multiple consecutive triggers, the pick

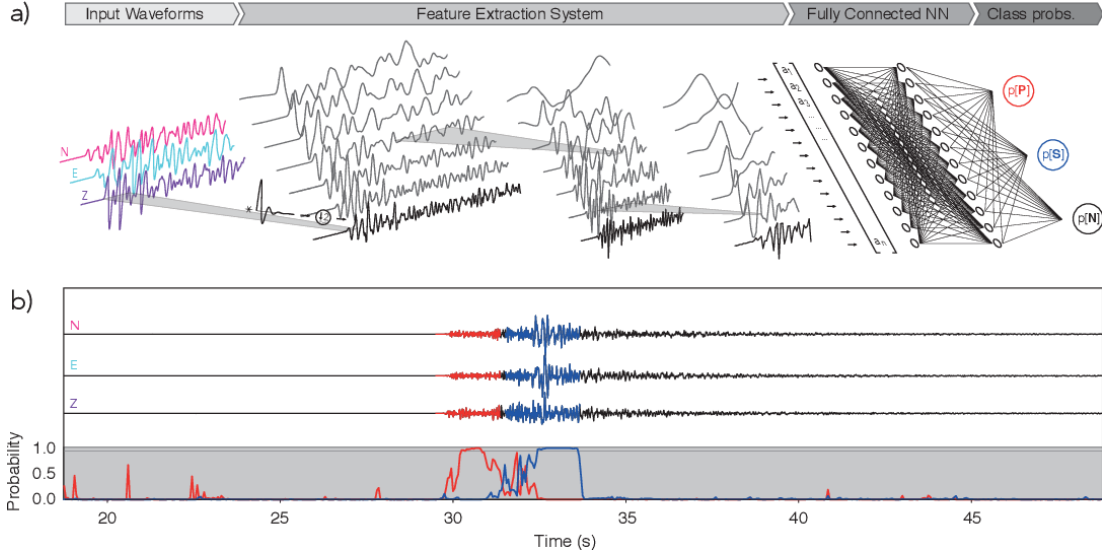


Figure 4.1: (A) The workflow of the GPD model from input waveforms to output class probabilities. (B) The top panel shows continuous seismic traces of 3 components (E, N and Z) and the bottom panel shows the probability traces of the P (red) and S (blue) phases through time (Ross, Meier, Hauksson & Heaton 2018).

time would be the time at which probability is at its maximum value.

Ross, Meier, Hauksson & Heaton (2018) trained their Generalized Seismic Phase Detection (GPD) model on detrended 100 Hz data filtered with a 2.0 Hz high-pass filter. The data is then normalised by the absolute maximum amplitude on any of the E, N or Z components. We note that the GPD model was trained on regional earthquakes on surface stations (i.e. not higher frequency microseismicity recorded on a downhole array of 15 Hz geophones). This can be a potential issue because HFIS is not represented in the training data. Although the model was trained with a 2.0 Hz highpass filter, Ross, Meier, Hauksson & Heaton (2018) found that the GPD model was still able to detect phases for a broad frequency range above and below 2.0 Hz. The flexibility of the model to pick phases outside the frequency range of the training data implies that it is possible for the model to pick phases in high frequency downhole data. This possibility will be further explored in the single station classification tests and the multi-station comparison sections (section 4.2 and 4.3 in Methods).

A few differences should be noted about the PNR-1z continuous downhole dataset and the GPD training dataset, namely: the sampling frequencies and filters used. The

different sampling frequencies change the duration of seismic data represented in the 400-sample sliding window. The GPD model was trained on 100 Hz data but the continuous downhole data has a sampling frequency of 2000 Hz. This means that the 400-sample sliding window feature will only contain 0.2 seconds of seismic data instead of 4 seconds (for 100 Hz data). A 0.2 second sliding window might seem short. However, event detection should still work because Figure 4.2 shows that individual P and S phases of the largest catalogued event are well within a 0.2 second window. For our study, we also selected an n_{shift} of 50 samples (0.025 s) for our sliding window. We specified a low value for the n_{shift} parameter so that the sampling would be sufficient enough to accurately capture the microseismic phase arrivals. Therefore, the microseismic phases recorded in the PNR-1z dataset are short enough to fit within a the 400-sample sliding window and thus are suited for high frequency phase detection by the GPD model.

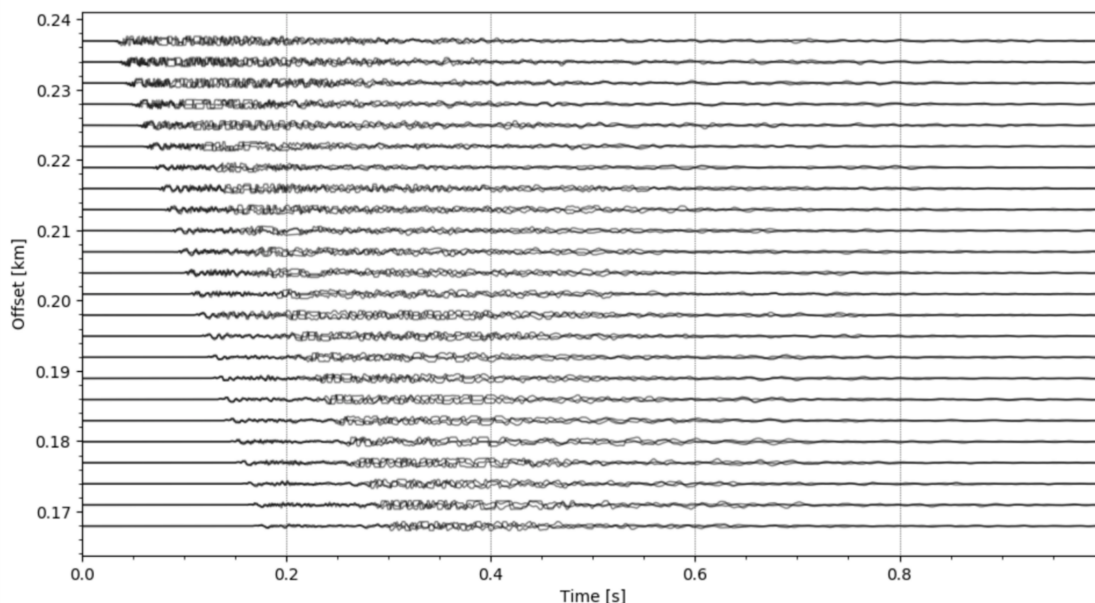


Figure 4.2: A horizontal section plot showing the moveout of the largest magnitude event ($M_w = 1.155$) across all 24 stations. The three component traces are plotted over each other at each station (total of 72 traces). The y-axis shows the offset, which is the distance from the source to each receiver/station. Each individual phase is within 0.2 seconds.

4.2 Data Pre-processing

Before running the pre-existing GPD model on the continuous downhole PNR-1z dataset, the raw data has to undergo preprocessing to ensure optimal conditions for event detection. Cuadrilla Ltd. recorded the raw continuous downhole data in their respective station orientations. Since Ross, Meier, Hauksson & Heaton (2018) trained the GPD model on millions of seismograms in the E, N and Z components, the continuous downhole PNR-1z traces (input) also have to be rotated to the E, N and Z components.

In addition, we created spectrograms of the largest and smallest magnitudes recorded to assess the dominant frequencies of the signal and noise as well as to choose a suitable filter for the raw data. We also conducted Fast Fourier Transform spectral analyses of the traces on each station to highlight a range of dominant frequencies of the noise and the signal. On the basis of these results, we chose a 50 Hz fourth-order Butterworth-highpass filter for GPD model testing on the continuous downhole PNR-1z dataset. Figure 4.3 shows that after filtering, the events are more easily recognisable which can make phase picking easier for the deep learning algorithm. Sufficient pre-processing can increase efficiency of this analysis (Kislov et al. 2020).

4.3 Single Station Classification test

We conducted a classification test to achieve Objective 1: to evaluate the ability of the GPD model to correctly classify seismic phases (P, S and noise) on single stations. This is done by using visually inspected 400-sample phase windows from single, randomly selected stations from the CMM event catalogue as the input for the model and observing how many the model can classify correctly in a test dataset. We chose to provide the model with 400-sample windows as this is primarily the input that the pre-trained GPD model classifies. We also chose windows on random stations to not introduce bias into this test, as the GPD model may perform better on certain stations (e.g. stations closer to the events) or when some stations are less noisy than others. We use the event waveforms from the CMM catalogue as ground truth.

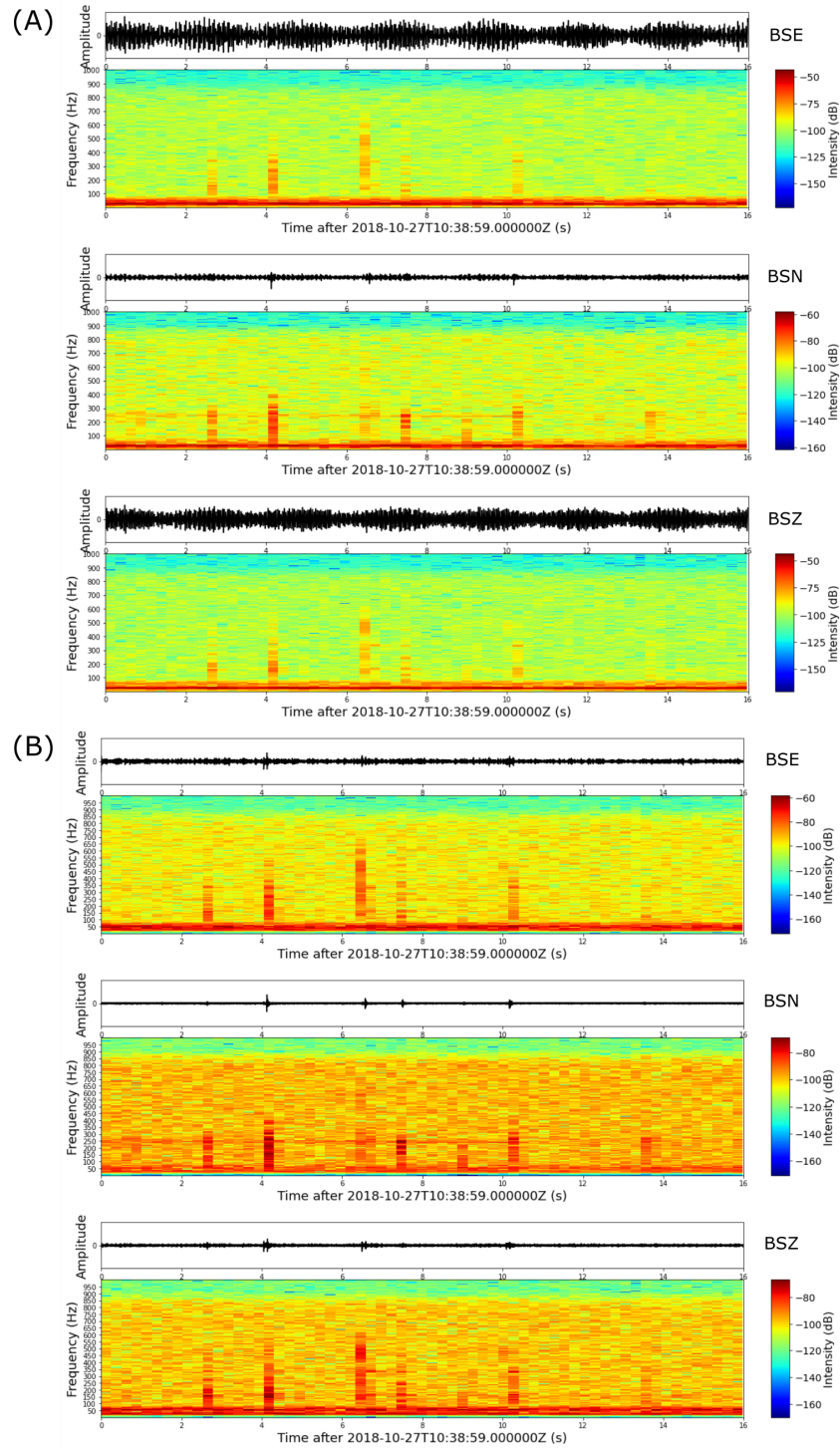


Figure 4.3: Continuous trace seismogram (top panel) with the respective spectrogram (bottom panel) for the E, N and Z components on station 17 (A) before and (B) after filtering with a 50 Hz Butterworth highpass filter.

4.3.1 Test Dataset for Single Station Classification

For this evaluation, we run the model against a test dataset of 250 events consisting of an equal number of P phases, S phases and noise 400-sample (0.2 second) windows. Thus, a total of 750 phase windows (250 P, 250 S, 250 noise) were used as the input for this classification test. The test dataset was compiled from the CMM event catalogue. Each of the event windows have been filtered and visually quality checked (QC-ed) to ensure that the selected phase is in the middle of the 400-sample window (where the model computes phase probabilities) and that the phase is visible on the traces of the randomly selected station. We need to make sure that a human analyst would be able to detect the phase visibly on the traces so that the test for the GPD model is appropriate. We chose 750 phase windows for this test so that it is sufficient for robust statistics and yet manageable to visually QC. The test dataset of 250 events includes the 50 largest magnitude events, the 50 smallest magnitude events, 50 mid-range magnitude events (25 from $-1 \leq M_w < 0$ and 25 from $-2 \leq M_w < -1$) and 100 random events that occur during injection. This is done to see if the model can detect phases of large, small and mid-range magnitude events, and how well the model performs generally over all 250 events.

4.3.2 Confusion matrix, classification metrics and probability thresholds

We will use the confusion matrix to represent the results of this classification test. The confusion matrix is commonly used to measure the performance of multi-class classification algorithms (Silva-Palacios et al. 2017, Tharwat 2020). The confusion matrix is a method that tabulates the results of a model and is a convenient way to calculate and gauge model performance. Metrics from the confusion matrix such as precision, recall and the F1-score of the GPD model will be crucial in qualitatively representing the performance of the model. We can also use these metrics to aid the selection of the probability detection threshold of the model.

The GPD model classifies each window based on the probability of the window being a P, S or noise window. The model labels the input (400-sample window) as the class with the highest probability that exceeds the user-defined probability threshold. We use a 3x3 confusion matrix that states the number of true positives and misclassifications for

Catalogue (actual)	Model (predicted)			
		P	S	Noise
	P	TP_{pp}	ε_{sp}	ε_{np}
	S	ε_{ps}	TP_{ss}	ε_{ns}
	Noise	ε_{pn}	ε_{sn}	TP_{nn}

Table 4.1: Layout of the confusion matrix table where TP_{xx} is a true positive of a phase and ε_{xy} is a misclassification for the phases P, S and noise. The subscripts on the labels show the phase the model predicts followed by the actual true phase (e.g., for ε_{ps} , the model predicts a P phase arrival when the actual phase arrival is an S phase- this makes it a misclassification).

the 3 seismic classes: P, S and noise windows (Table 4.1). During classification, when a window containing a P phase is fed into the model and the model classifies this window as a P, this counts as 1 TP_{pp} in the confusion matrix. When the algorithm misclassifies a P window as an S or noise window, the count for ε_{sp} or ε_{np} increase respectively. The count applies the same for both the S phase and noise windows.

Once we construct the confusion matrix, the classification performance of the model is further evaluated by calculating the recall (or sensitivity), precision and F1-score of the GPD model (Sokolova et al. 2006). Recall R is defined as

$$R = \frac{(TP)}{(TP + FN)} \quad (4.1)$$

where R is the recall, TP is the number of true positives and FN are false negatives.

R gives an indication of how well the model performs with regards to the number of false negatives (or missed events). The closer the R value approaches 1, the fewer events the model has missed. If $R = 1$, the model has missed no events. A low value of R indicates that the model has missed many known events.

Precision P is defined as

$$P = \frac{(TP)}{(TP + FP)} \quad (4.2)$$

where P is precision, TP is the number of true positives and FP is the number of falsely predicted positives.

Whereas R places more weight on false negatives, P sheds more light on how the model is performing with regards to false positives (or false picks). A high P value is desired as this means that the model produces a low number of false picks. Low P values imply that the model produces a large number of false picks.

We can calculate the recall and precision for individual classes (P, S or noise). To calculate individual recall of a phase (e.g. P), the FN for P would be the sum of the misclassifications of the P phase (i.e. $FN_p = \varepsilon_{sp} + \varepsilon_{np}$). For individual phase precision (e.g. for P), the FP for P would be the sum of misclassifications for other phases as a P phase (i.e. $FP_p = \varepsilon_{ps} + \varepsilon_{pn}$).

The F1-score can be computed as

$$F1 = \frac{2TP}{2TP + FP + FN} = 2 \times \frac{P \times R}{P + R} \quad (4.3)$$

where TP is the number of true positives, FP is the number of false positives, FN is the false negatives, P is the precision and R is the recall.

The F1-score represents the harmonic mean of both precision P and recall R (Sokolova et al. 2006) and weights P and R equally. The overall F1-score of a confusion matrix is calculated using the average recall and precision of all three phases. F1-score values can lie between a range of 0 and 1. Ideally, the F1-score should be close to 1 as this means the model possesses high values for precision and recall. To determine the optimal probability threshold for the model for microseismic phase detections, we construct several confusion matrices over a range of detection probability thresholds. We compute the metrics (i.e. precision, recall and F1-scores) for probabilities between 0 and 1 at 0.01 intervals. The results from the classification tests help determine the ideal probability threshold that yields the best F1-score (i.e. the best balance of precision and recall).

4.3.3 Classification Performance Review on Different Magnitude Bands

As the test dataset contains four different magnitude bands: the largest magnitude events (the largest 50 $M_w > 0$ events), mid-range magnitude events (25 events from $-1 \leq M_w < 0$ and 25 events from $-2 \leq M_w < -1$, so a total of 50 events from $-2 \leq M_w < 0$) and

the smallest magnitude events (the smallest 50 $M_w < -2$ events). We can compute four different confusion matrices respectively. The confusion matrix metrics (P , R , F1-score) are also be evaluated for each matrix. For this review, we did not set a probability threshold for the model and just used the classification with the highest probability for each window label. This is to evaluate how the GPD model performs with detecting events of different magnitude ranges.

4.4 Multi-station comparison

For multi-station comparison, we adapt a workflow of the GPD model to run the model on the continuous downhole PNR-1z dataset, associate the phase picks, locate the events and produce an event origin time catalogue to compare with the catalogue produced by the CMM method.

4.4.1 PNR-1z Continuous Downhole Data Model Tests during Periods of Interest

To assess the efficiency and multi-station performance of the GPD model, we run the model on continuous downhole data obtained from activities on PNR-1z. The continuous data files are firstly concatenated into a matrix. This data matrix is rotated to the E, N and Z components, filtered with a 50.0 Hz Butterworth highpass filter and then de-trended. A probability threshold of 0.5 was chosen for the multi-station event detection tests because findings in subsection 5.1.1 (Results chapter) show that it is difficult to gauge the optimum probability threshold as there is a trade-off between precision and recall values when classifying seismic phases from HFIS. Zou et al. (2016) also suggest that a probability threshold of 0.5 is the default value in interpreting classification using probabilities. The resulting continuous data matrix is then used as input for the GPD model.

These model tests were conducted on the University of Bristol’s ACRC Blue Pebble cluster with one GeForce GTX 1080 GPU on a node with one CPU. The computational runtimes of the model for earthquake detection on all stations will be reviewed in subsection 5.3.4 in the Results section. We focus on creating a catalogue of origin times of

earthquake events during specific periods of interests. To conduct a multi-station method comparison between the GPD and the CMM (Objective 3), we have chosen a 3-hour continuous data period that includes 2 different periods of interest to run the model on:

- High Magnitude Period:

An hour of data that captures time around a large magnitude event ($M_w > 0$). This is to assess if the GPD model is able to detect more aftershock events or if the model can detect any more events that could be potential precursory microseismic activity. Many early aftershocks often occur in the coda of larger events and therefore these events might be harder for the CMM method to detect. These missed aftershocks lead to short term aftershock incompleteness in a catalogue (Kagan 2004, Peng et al. 2007). For this period, we have chosen to look for aftershocks after the largest event in the CMM catalogue ($M_w=1.155$) on 11 December 2018 11am-12pm (Figure 4.4(A)).

- Event Onset Period during High Injection Rates:

These are periods that display increased seismicity during high injection rates. During injection, a large number of events arrive in close temporal proximity (i.e. high seismic rates). This period was selected to see if the GPD model can keep up with detecting high seismicity or even detect additional events (if any) that the CMM catalogue might have missed. We selected the period on 11 December 2018 from 9-10am to represent the event onset period (Figure 4.4(B)).

For the GPD model input, we selected a 3 hour continuous period from 9am to 12pm on 11 December 2018 to include our periods of interest (event onset during high injection rates period and the high magnitude period). The resulting output of the model will be a list of phase picktimes (P and S) from each of the 24 stations for the 3 hour period.

4.4.2 Event Detection Workflow: Phase Grouping, Phase Association and Event Location

To conduct a fair multi-station comparison of the CMM and GPD event catalogues (Objective 3), we need a workflow that converts individual lists of picktimes to an origin time catalogue of located events (Objective 2). The workflow needs to first concatenate

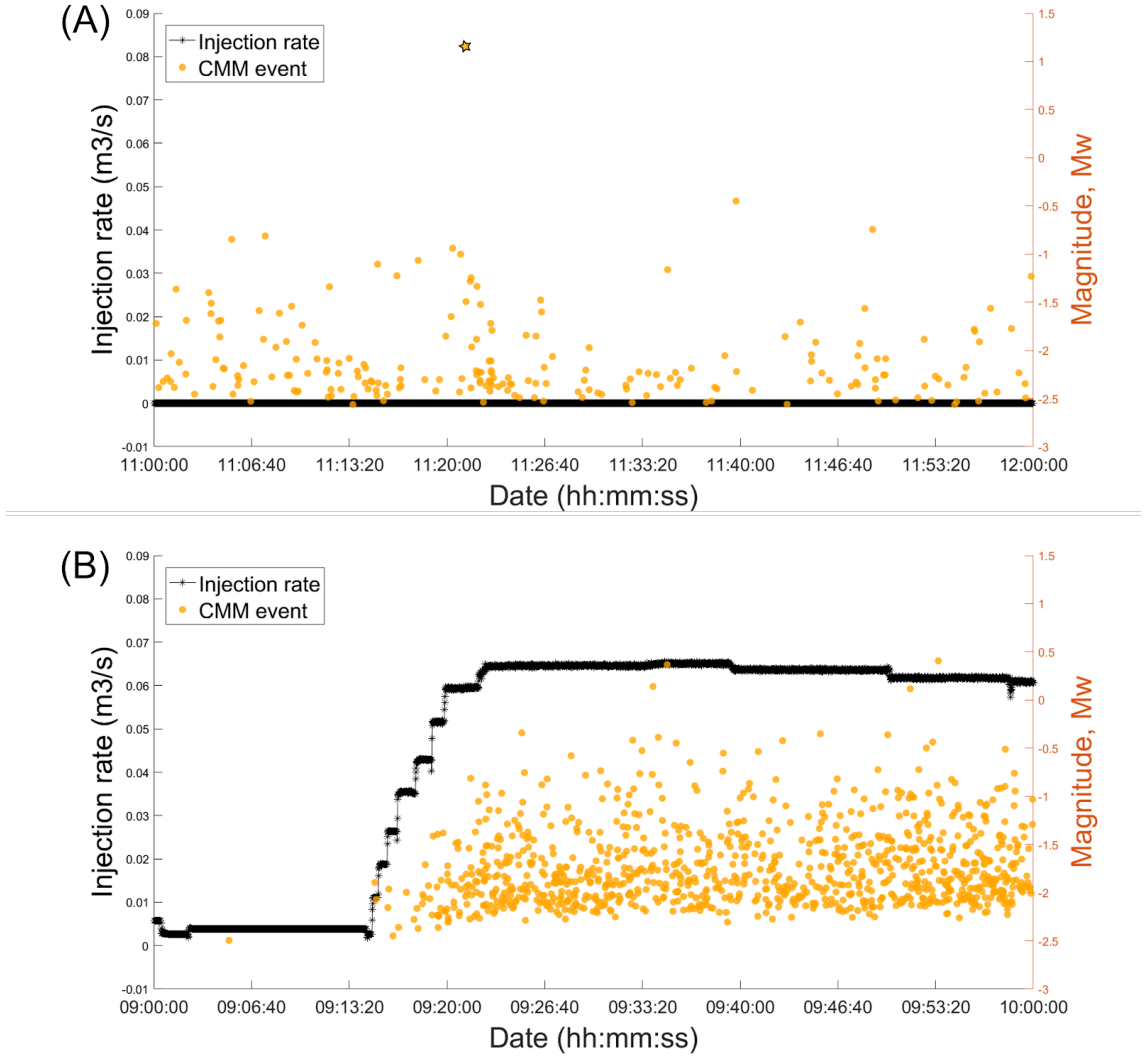


Figure 4.4: Temporal plots of injection data and events from the CMM catalogue on the periods of interest: (A) high magnitude period and (B) the event onset period during high injection rates. The yellow star in (A) represents the largest event in the CMM catalogue.

the lists of phase picktimes (GPD model output) for all stations, group the phase picks into events, associate the phases and locate the events to create the origin time event catalogue.

Before phase association, we group phase picks of the same phase and event across the stations. For this phase grouping, we define the longest travel time for a phase of the same event to arrive at all stations. To calculate this travel time, we used the equation

$$tt_{pha} = \frac{x_{max}}{V_{pha(min)}} \quad (4.4)$$

where tt_{pha} is the time the phase (P or S) takes to travel between the 2 furthest stations in the array, x_{max} is the maximum distance between the 2 furthest stations in the array and $V_{pha(min)}$ is the velocity of a phase (P or S) at the shallowest depth of the array. We group the P and S phase picks separately for phase association later.

The maximum distance x_{max} between the two furthest stations (station 1 and station 24) is estimated to be 681.81 meters. From the velocity model of PNR-1z, a lower bound V_p of 4823.817 m/s at 1559.305 meters depth was chosen (as station 1, the most shallow station, is located at a depth of 1561m). The maximum travel P wave travel time tt_p for this site is estimated to be 0.141 seconds. We round this tt_p value to 0.2 s to take error into account. For the S phase, we use the lower bound V_s of 2870.051 m/s around the same depth as station 1 on the same distance x_{max} for a tt_s of 0.238 seconds. We also round tt_s to 0.25 s.

Using these phase travel-time windows, the phase picks are grouped to make initial ‘event’ group lists. This is done by sorting the separate phase pick list by time and taking the earliest picktime as the ‘head’ unique event arrival time. Picks that are within the time window between the ‘head’ arrival time and the maximum phase travel time are assumed to belong to the same event. It is possible for multiple picks at the same station to be grouped into the same event. When multiple picks are made, we take the highest probability pick from the picks of the same stations because a higher probability pick is more likely to be an actual phase pick.

After phase grouping, we proceed to associate the two phases and obtain event locations using the NonLinLoc algorithm (see subsection 4.4.3). There are several methods of phase pick association to consider. In this thesis we considered 3 methods; the "PS" association, the "fixed-difference SP" association and the "fixed-difference PS" association.

The PS association is done by using only the P phase groups to create an observation file for initial event location. We filter the P phase groups by defining an ‘event’ to have a minimum of 4 phase picks for event location and drop groups that have less than 4 phase picks. Once we get the list of P phase arrivals, we use this to create an initial catalogue of origin times and locations using NonLinLoc with only the P picks. We add the initial

list of origin times to a dataframe. We then associate an S phase for each event origin time by calculating the longest (or maximum) travel time window for the corresponding S phase to arrive at each station. This maximum travel time window can be calculated by using the Wadati equation, defined as

$$(t_s - t_0) = \left(\frac{V_p}{V_s} \right) (t_p - t_0) \quad (4.5)$$

where V_p and V_s are the P and S wave velocity respectively, t_p and t_s are the P and S arrival time respectively and t_0 is the event origin time.

We assume that the S phase arrives within a time window between the origin time and the maximum S travel time calculated by using the maximum value of V_p/V_s of that event group ($t_0 \leq t_s \leq (t_0 + (V_p/V_s)(t_p - t_0))$). The maximum value of V_p/V_s is estimated from a list of V_p/V_s ratios between the event depth and each station depth. This is done for each event group so each event will have a unique maximum S arrival time. For the same station, we associate each S pick with its corresponding P phase. If there are more than one S pick associated with the P pick, we use the S pick with higher probability as higher probability picks are more likely to be an actual phase pick. After the association, we obtain a new list of origin times and location estimates by relocating the events with the associated P and S phase arrival information. As a result, we obtain a final event catalogue of origin times and locations during the selected periods of interest that we use to compare with the CMM catalogue.

Another way to associate the P and S phases is by using the “fixed difference” PS association method. In this method, we use the $t_s - t_p$ of the largest event ($M_w = 1.115$) from the furthest station as the fixed-difference time window for phase association. We use the $t_s - t_p$ of the largest event because it gives the clearest time difference between the P and S phase. We determine the $t_s - t_p$ of the $M_w = 1.115$ event to be 0.125 seconds. We round the $t_s - t_p$ to 0.13 seconds and define the S travel time window to fall between the origin time and 0.13 seconds after the event origin ($t_0 \leq t_s \leq (t_0 + 0.13)$). We use the same window for phase association at each station. If an S phase on a station falls within the time travel window, it will be associated with the P picktime of that same station. Here, we only compute the locations once using NonLinLoc with both the P and S associated phases.

The last phase associating method is the “fixed-difference” SP association. This is

done by using the S phase groups to associate to the P phase groups. We first use the S picks to create initial estimates of locations and origin times. We then define a possible P phase travel time window for each S pick using the fixed-difference window (maximum $t_s - t_p$) from the method before. We assume that the P phase can arrive between the S arrival time and 0.13 seconds before the S arrival time at each station ($(t_s - 0.13) \leq t_p \leq t_s$). After associating the P picks to the S picks, we use NonLinLoc to obtain event locations and origin times.

For each association method, we compared the residual root-mean-square (RMS) errors, horizontal and vertical error estimates in the locations to choose the most robust phase association method. The output of this finalised workflow will be an origin time event catalogue with their respective locations.

4.4.3 NonLinLoc Event Locations

To obtain event locations, we use the NonLinLoc (Non-Linear Location) package by Lomax et al. (2000). This software package uses probabilistic non-linear and global search methods to obtain absolute locations within a 3D subsurface grid. Within the NonLinLoc package, we used the Vel2Grid, Grid2Time and NLLoc programs to obtain event locations and origin times. We defined a $3 \times 3 \times 3$ km model grid with node spacing of 0.25 km and used the velocity model of the PNR-1z site provided by Cuadrilla. We decided to use the NonLinLoc Phase file format because the input for phase pick times can be provided to the microsecond. The resulting event origin time will also be precise to the microsecond. This is important as several microseismic events can occur within the same second.

The GPD model outputs a probability for each phase pick. In the phase file, we weighted each pick by assigning each pick probability with an error window and a corresponding time error. Table 4.2 shows how the phase picks were weighted. We chose relatively short error windows to capture microseismic phases because microseismic events last for short periods of time. A further study is needed to gauge how pick probabilities correspond to the error windows on the picks but for this study, we have chosen these errors empirically. Increasing error windows were chosen with decreasing pick probability for our 2000 Hz continuous data.

Pick Probability	Error Window (number of samples)	Time Error (seconds)
$\text{Prob} \geq 0.85$	5	0.0025
$0.85 > \text{Prob} \geq 0.70$	10	0.005
$0.70 > \text{Prob} \geq 0.60$	20	0.01
$0.60 > \text{Prob} \geq 0.50$	50	0.025

Table 4.2: Pick probability bands assigned to an error window around the phase pick in samples (chosen for 2000 Hz data) and the corresponding time error in seconds.

4.4.4 Event Catalogue Comparison

After producing an event origin catalogue from the event detection workflow with the GPD phase picks, we compare the resulting catalogue with the CMM catalogue. This was carried out by event-associating the GPD picked events with the events in the CMM catalogue. We impose a 0.2 second time window around each event origin time (t_0) in the CMM catalogue so that any GPD event that falls within the time window ($t_0 \pm 0.2$ seconds) will be associated with that CMM event origin time. This increases the possibility of mis-associating events (having more than one GPD event associated with a CMM event). Despite the long (for microseismic events) time window, the method only mis-associates 19 times (for the fixed-difference PS association method) out of 1972 possible CMM-event associations (0.1% of the time), 24 times (1.2%) for the PS association and 42 times (2.1%) for the fixed-difference SP method.

Chapter 5

Results

5.1 Model Classification Performance

5.1.1 Classification test results

We first present the model classification results without a user-defined probability threshold. When a probability threshold is selected, the probability must exceed the threshold to be classified under a phase. Here, the model does not have a set probability threshold and thus uses the highest phase probability to classify windows into the P, S or noise classes. Table 5.1 illustrates the confusion matrix of the GPD model classifying 750 phase windows. From the confusion matrix, we can derive the recall, precision and overall F1-score of the model.

TOTAL 250EV	MODEL (PREDICTED)				
CATALOGUE (ACTUAL)	Phase	P	S	NOISE	TOTAL
	P	64	4	182	250
	S	43	41	166	250
	NOISE	1	3	246	250
	TOTAL	108	48	594	750

Table 5.1: The GPD model confusion matrix for 750 400-sample phase windows filtered with a 50 Hz highpass Butterworth filter.

Recall R is the fraction of correct phase picks detected by the model out of all known correct picks from the catalogue. As explained in the Methods section (4.2.2), the R value approaches 1 when the number of false negatives (missed events) decreases to zero. The R values of the P, S and noise classes are 0.26, 0.16 and 0.98, respectively, giving a low average R of 0.47. These low recall values of the P and S phases indicate that the model fails to detect (misses) a large number of P and S phases.

Precision P is the fraction of all the picks made by the model that are true positives. The P value is 0.59 for the P phase, 0.85 for the S phase and 0.41 for noise windows, averaging 0.62. These results show that the GPD model produces fewer false positives for the S phase than the P phase. The low precision for noise windows shows that the model produces a large number of false positives- frequently misclassifying the P or S phase windows as noise.

The F1-score is a value that equally weighs the precision and recall. For individual P, S and noise classes, the model produces low F1-scores of 0.36, 0.28 and 0.58, respectively. We estimate the overall F1-score of the model to be a low value of 0.54 from the average R and P values of all three classes. A low F1-score indicates that the classifying ability of the model produces low precision and recall.

Next, we present results from repeating the classification test across a range of different probability thresholds (Figure 5.1). The optimal probability threshold is unknown a priori and could vary from the threshold Ross, Meier, Hauksson & Heaton (2018) used in their study (0.98 for regional earthquakes) and so we want to obtain the probability threshold that will yield the highest model F1-score for HFIS. The GPD model produces the best balance of precision and recall when the probability threshold equals 0.02, attaining an F1-score of 0.51. The optimal probability threshold of 0.02 is too low for application in our multi-station continuous model tests. Instead, we select a probability threshold of 0.5 empirically as we expect the picks on multiple stations to strengthen event detection. Figure 5.1(A) shows a decreasing trend of the overall F1-score with increasing probability threshold. The F1-scores of individual classes also decrease with the increasing probability threshold (Figure 5.1(B)).

Figure 5.2 shows a trade-off between recall and precision with increasing probability threshold for the P and S classes. This trade-off explains the decreasing trend of F1-

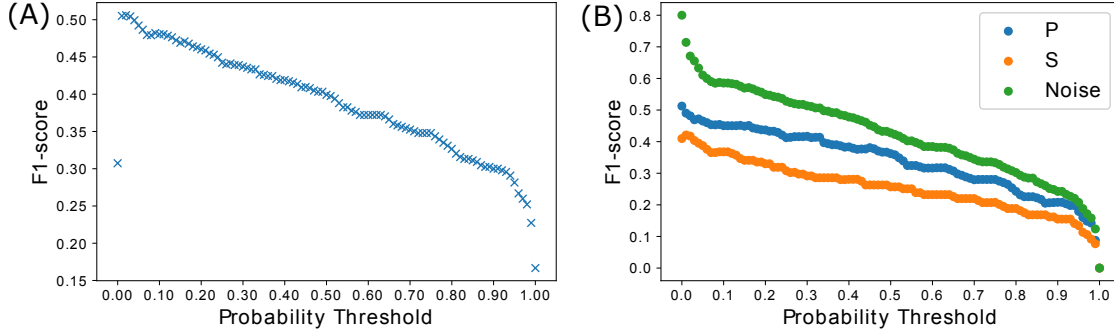


Figure 5.1: (A) Overall F1-scores and (B) F1-scores of individual classes (P, S, Noise) as a function of the probability detection threshold.

scores with increasing probability threshold for the P and S classes. For both phases, the precision first increases rapidly but then plateaus. Meanwhile, recall decreases steadily. Figure 5.2 indicates that the precision for the P phase plateaus at around 0.6 whereas it approaches 1 for the S phase. This suggests that the classified S phases tend to be true positives, while there are a lot of classified P phases that are false positives.

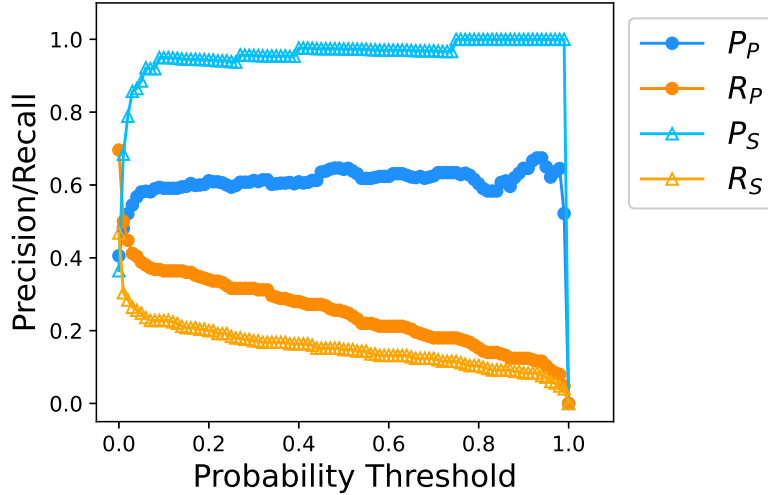


Figure 5.2: Precision (in shades of blue) and Recall (in shades of orange) as a function of probability detection threshold for the P (filled circles) and S phase classes (empty triangles).

5.1.2 Classification Test on Different Event Magnitude Bands

The confusion matrix displays low recall of the P and S phases and low precision of noise windows. A possible explanation for this is that the current model does not recognise phase arrivals of small events ($M_w < -2$). To show this, we constructed confusion matrices from subsets of the test dataset for events of random magnitudes during injection (as the control confusion matrix) and 4 different magnitude bands: the 50 largest $M_w > 0$ events, 25 events from $-1 \leq M_w < 0$, 25 events from $-2 \leq M_w < -1$, and 50 of the smallest $M_w < -2$ events. The individual confusion matrices are shown in Appendix 1 but the results will be analysed in this section with the precision and recall values.

Subsequently, we calculate the precision and recall values for each confusion matrix and, thus, each magnitude band (Figure 5.3). Figure 5.3 shows that the precision of each class decreases significantly with descending magnitude bands. The figure also displays decreasing recall for the P and S phases while the noise recall varies little.

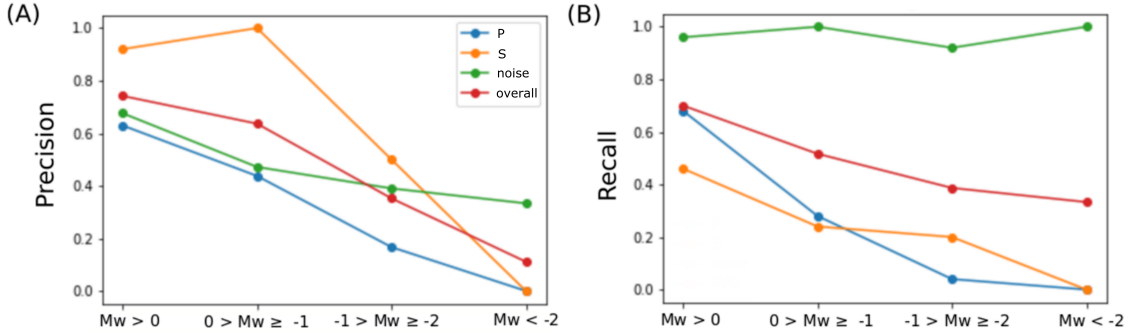


Figure 5.3: (A) Precision and (B) Recall values of all classes (P, S, noise, and the overall value) as a function of moment magnitude bands ($M_w > 0$, $0 > M_w \geq -1$, $-1 > M_w \geq -2$ and $M_w < -2$).

The significant proportion of low magnitude events in the test dataset may explain why the GPD model seems not to perform well. Results for the smallest events ($M_w < -2$) show that the model did not detect any P or S phases and misclassified all 100 of the non-noise windows as noise. In the confusion matrix for all 250 events, a total of 90 events have $M_w < -2$ (36% of the test dataset). These results may explain the low recall values- the model frequently misses these small magnitude events.

The model produces higher average P and R values as the moment magnitude increases. The higher recall values on the larger magnitude events indicates that the model misses fewer events when the events are large. Although we have higher P and R values for larger magnitude events, the model's classifying performance is still moderate as the model only produces overall P and R values in the 0.7 range. These values mean that the model still produces a number of false positives and misses a proportion of M_w events.

For mid-magnitude events, the model produces an overall precision value of 0.49 and an overall recall value of 0.45. These mid-magnitude results indicate that, to some extent, the GPD model is still able to detect events with magnitudes lower than the magnitude range represented ($-0.81 < M < 5.7$) in training dataset- albeit missing a majority of the $M_w < -2$ events.

5.2 Multi-station Event Detection Workflow

5.2.1 Phase Picks and Phase Grouping

The GPD model picked a total of 66,484 phases (28,922 P phases and 37,562 S phases) over 24 stations of continuous data between 9am and noon on 11 December 2018. If we separate the observations on 11 December 2018 into 1-hour periods, there are 34,057 picks (15,530 Ps, 18,527 Ss) for the event onset period during high injection rates (9am-10am), 28,042 picks (11,918 Ps, 16,124 Ss) from 10am-11am, and 4,385 picks (1474 Ps, 2911 Ss) during the large magnitude period (11am-12pm). After phase grouping (where we group phase picks on at least 4 stations into events), the phase numbers reduced to a total of 2036 P (hourly: 1108, 826 and 102 groups) and 2,641 S groups (hourly: 1330, 1120 and 191 groups) for phase association. These results again indicate that the GPD model picks more S phases than P phases for microseismic events.

From visual inspection, the model is able to produce accurate phase picks (Figure 5.4). However, there is a large proportion of high probability ($\text{Prob} \geq 0.85$) inaccurate phase picks that are slightly off by a few milliseconds for an event within the 24 geophone array (Figure 5.5). Figure 5.5 displays an inaccurate P phase pick with a low pick probability ($\text{Prob} = 0.5$) and a high probability ($\text{Prob} = 0.95$) but inaccurate S phase pick. These inaccurate phase picks can severely affect event locations in the later stages of the workflow.

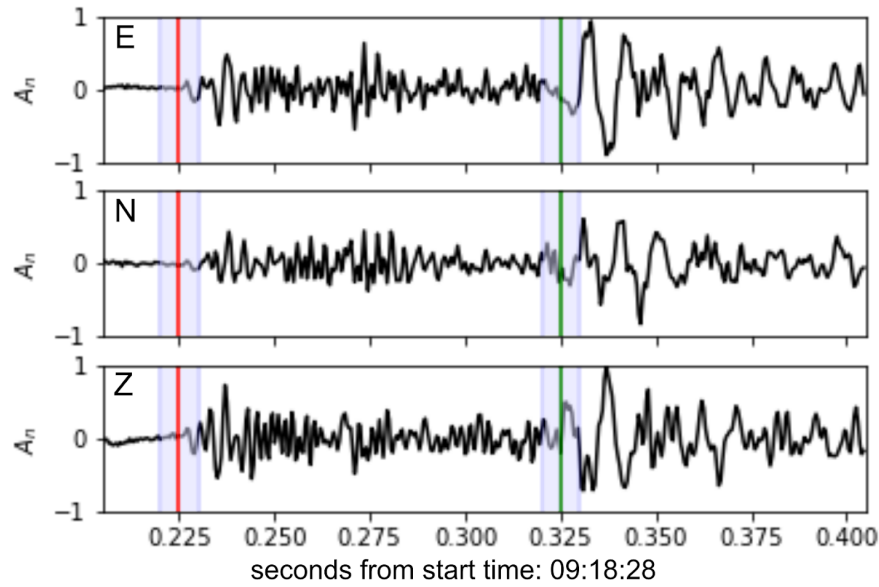


Figure 5.4: Accurate model picks for the P (red) and S (green) phases. Subplots from top to bottom show the waveforms with normalised amplitude for the E, N and Z components on station 11. The shaded areas around the pick show the assigned time error ($P = 0.0025$ s, $S = 0.0025$ s) around the pick according to pick probability ($P = 0.9995$, $S = 0.9734$).

5.2.2 Results of the Phase Association Methods

For each phase association method (PS, fixed-difference SP and fixed-difference PS), we compare location errors and select the most robust phase association method for the event detection workflow. We obtain the RMS of the travel time residuals, horizontal and vertical location errors of each method from the NonLinLoc program.

Figure 5.6 illustrates the cumulative distribution functions (CDFs) of the RMS errors between the three methods. The RMS errors indicate that the fixed-difference PS method performs slightly better, while the fixed-difference SP method shows the worst RMS errors. The 95th percentiles of the RMS errors for the PS, fixed-difference SP and the fixed-difference PS method are 0.0126, 0.015 and 0.0116 seconds, respectively. However, the differences between these errors are not very significant. Figure 5.6 shows that for all three methods, 80% of the RMS errors are within 0.008 seconds.

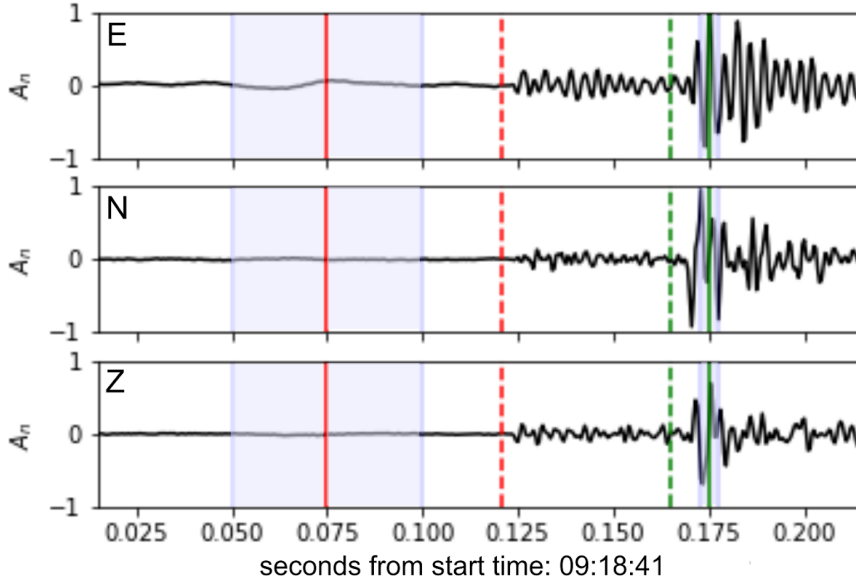


Figure 5.5: An early P phase pick (in red) and a late S phase pick (in green) on normalised waveforms for the E, N and Z component on station 8. The shaded areas represent the time error around each phase pick according to pick probability. The dotted lines are our manual phase picks. Time errors for the P and S phases are 0.025 and 0.0025 seconds, respectively (for probabilities $P = 0.5$ and $S = 0.95$).

In addition, we construct CDF plots to assess the horizontal and vertical location errors of the association methods (Figure 5.7). Figure 5.7(A) shows that 80% of the fixed-difference PS and fixed-difference SP locations possess horizontal errors of less than 0.3 km whereas 80% of the PS method are within 0.4 km. Results for the vertical errors are similar- 80% of the fixed-difference methods are within 0.2 km while the PS method has a larger vertical error of 0.3 km (Figure 5.7(B)). All methods at the 95th percentile share a horizontal error of 0.5 km and a vertical error of 0.3 km. Although the fixed-difference association methods produce lower errors than the PS association, these results are still not substantially different to select one particular method over another.

As we did not see significant differences in the phase association methods in the RMS, horizontal or vertical location errors, we also produced CDF graphs for the absolute x, y and z differences (Easting, Northing and depth) between the CMM and GPD event locations (Figure 5.8). We observe that the fixed-difference methods possess lower x and z absolute location differences than the PS method.

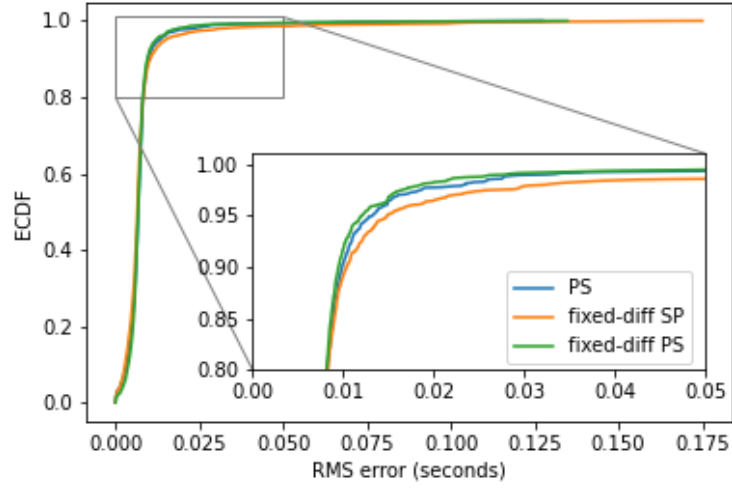


Figure 5.6: Empirical Cumulative Distribution Function (ECDF) of the RMS errors for the PS (blue), fixed-difference SP (orange) and fixed-difference PS (green) association methods with a zoomed-in window to show the slight difference in the methods.

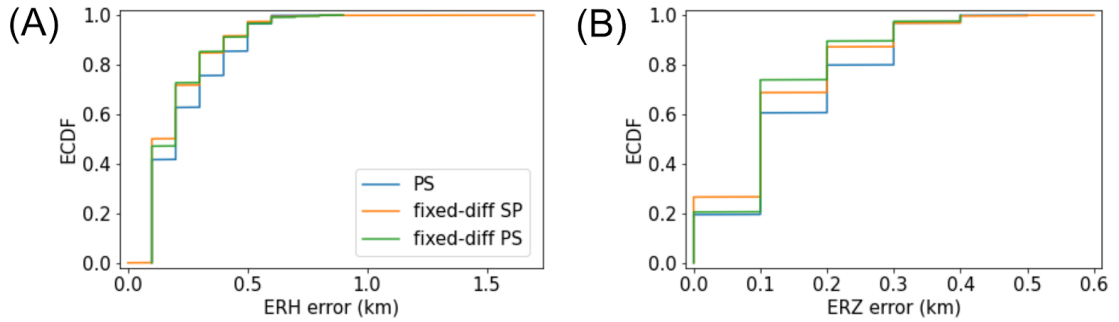


Figure 5.7: (A) ECDF of the horizontal location errors (ERH) and (B) ECDF of the vertical/depth location errors (ERZ) for the PS (blue), fixed-difference SP (orange) and fixed-difference PS (green) association methods.

Out of the phase association methods, the PS method performs the worst in the vertical and horizontal errors as well as the absolute location differences. Although the fixed-difference PS method produces the worst absolute location differences in the y axis, the differences are not very considerable. Therefore, we infer that the PS association method performs consistently the worst amongst the association methods we observed.

To decide between the remaining two association methods, we compare the event cata-

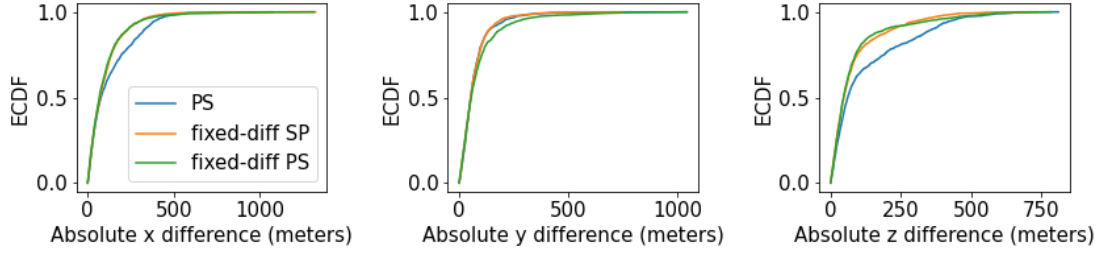


Figure 5.8: CDF as a function of (A) absolute x, (B) absolute y and (C) absolute z location differences the three phase association methods.

logues from the fixed-difference PS and fixed-difference SP methods. We use the resulting catalogues from the multi-station continuous downhole model tests for comparison (3 hours of continuous data on 11 December 2018, 9am to 12pm). The CMM catalogue lists 1972 seismic events during this period. The workflow with the fixed-difference PS method detected 1589 events (81%) while missing 383 events (19%) of the CMM catalogued events. This method also detected 429 new events that are not listed in the CMM catalogue. The workflow with the fixed-difference SP method detected 1760 events (89.2%), missed 212 events (10.8%) and found an additional 839 events.

We visually reviewed the new events detected with both phase association methods. We selected a sample of 172 (40%) out of the 429 new events from the fixed-difference PS association and 168 (20%) of the 839 SP associated new events. The review indicates that 3.5% of the fixed-difference PS associated new events are false positives whereas 16% of the fixed-difference SP associated new events are false positives. Assuming these percentages are representative for the whole sample, we can scale the results of each method with the respective total number of new events. The fixed-difference PS method produces approximately 414 new true positives while 15 false positives and the fixed-difference SP method produces 704 new true positives and 135 false positives during the three hour period.

On the basis of these results, we decided to use the fixed-difference PS association method in the event detection workflow to minimise the number of false positive events. Selecting the fixed-difference SP method could be equally valid if the desired objective of the application is to achieve more events. However, this method would require performing a quality check on all events to eliminate false positives. This would make event detection more time-consuming and labour intensive- especially when the rate

of induced seismicity escalates during injection operations. The finalised GPD event detection workflow is illustrated in Figure 5.9.

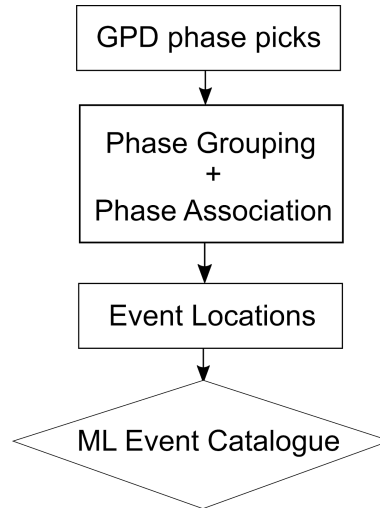


Figure 5.9: The finalised GPD event detection workflow from multi-station phase picks to an event origin time catalogue with locations.

5.3 Multi-station Catalogue Comparison

5.3.1 High Magnitude Period

We executed the model over one hour of continuous data of 24 stations on 11 December 2018 from 11am to 12pm (Figure 5.10). During this selected high magnitude period, the CMM catalogue contains a total of 218 seismic events, including the largest magnitude event ($M_w = 1.155$). Our workflow detected and located 96 events (44%) of the CMM catalogued events but missed 122 events (56%). For this period, 74% of the events (161 events) have moment magnitude $M_w < -2$. Figure 5.10 shows that all of the missed events are below $M_w -2$.

During this period, the GPD model only detected 6 new events. After the largest event (blue star on Figure 5.10), the model detected 3 new immediate aftershocks and 2 more new aftershocks in the next 10 minutes. We visually inspected all the new events during this period and can confirm that 5 out of 6 are real events. Figure 5.11 is an example of a new aftershock event detected after the M_w 1.155 event. These waveforms

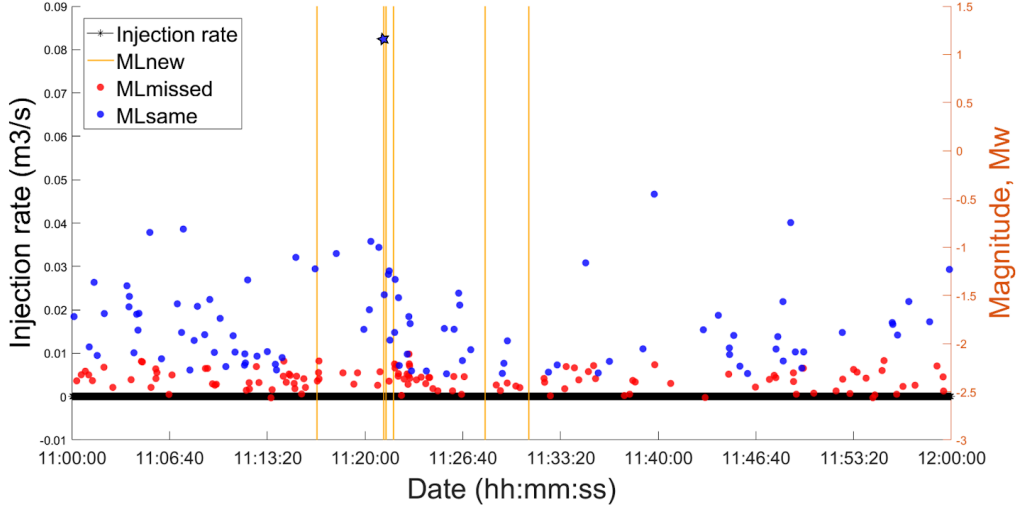


Figure 5.10: A temporal plot of the GPD model events (new events as vertical orange lines, missed events in red and identified events in blue) for the high magnitude period on 11 December 2018 (11am to 12pm). The new events are represented by orange lines because the respective magnitudes are not known. The star at 11:21:15.62 represents the largest magnitude event in the CMM catalogue.

also exhibit smaller events that both the CMM method and the GPD model did not detect. The model only detected one new aftershock event within the coda of the M_w 1.155 event (Figure 5.12). Although it looks like there is a visible feature within the waveforms, we classified this as a false positive event because the phases are not very clear and it is not certain whether it is an event or just noise.

5.3.2 Event Onset Period During High Injection Rates

For this model run, we applied the model on the event onset period (11 December 2018, 9am to 10am). As the injection rate increases, the rate of seismicity escalates (Figure 5.13). A total of 986 seismic events occurred during this period according to the CMM catalogue. The GPD model detected 865 events (87.7%) of the CMM catalogued events but missed 121 events (12.3%). There are 190 $M_w < -2$ events but only 85% (103) of the 121 missed events were $M_w < -2$. This suggests even though the model misses the majority of the low magnitude events, it is still possible for the model to detect $M_w < -2$ events.

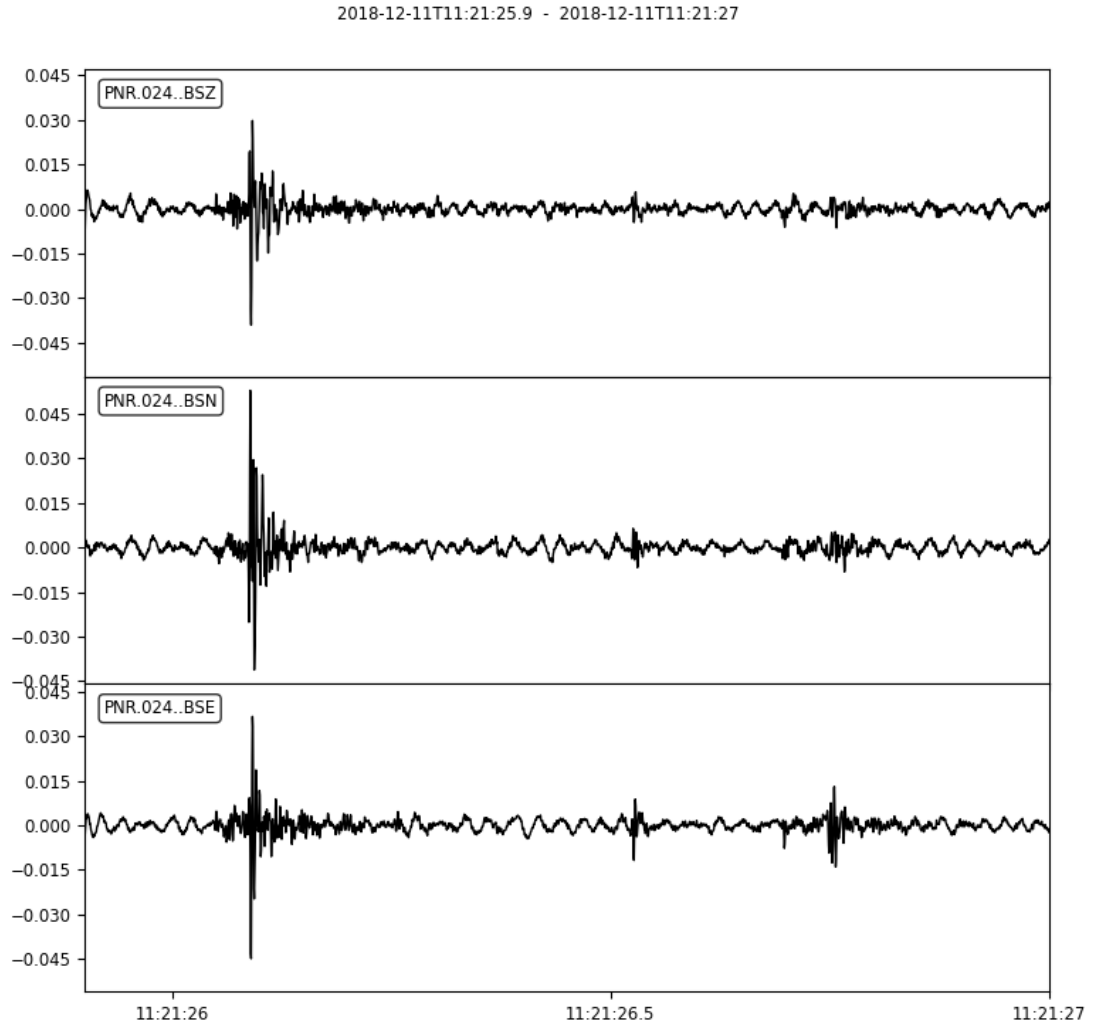


Figure 5.11: E, N and Z waveforms of the new aftershock (at 11:21:26.02) after the high magnitude event on station 24. The CMM and GPD model did not detect smaller events after this first event.

The GPD model detected 230 additional events within the hour. We reviewed a sample of 100 new events and determined that 97% of the events are true positives and 3% are false positives. From this, we can interpolate that only 7 out of the 230 new events are false. A large proportion of these new events during high injection rates are events that occur in close temporal proximity. Figure 5.14 shows an example of the newly detected events during high seismicity rate- i.e. when there are several events within a second. The CMM method missed the first event in Figure 5.14, but both the model and CMM detected the subsequent event within one second.

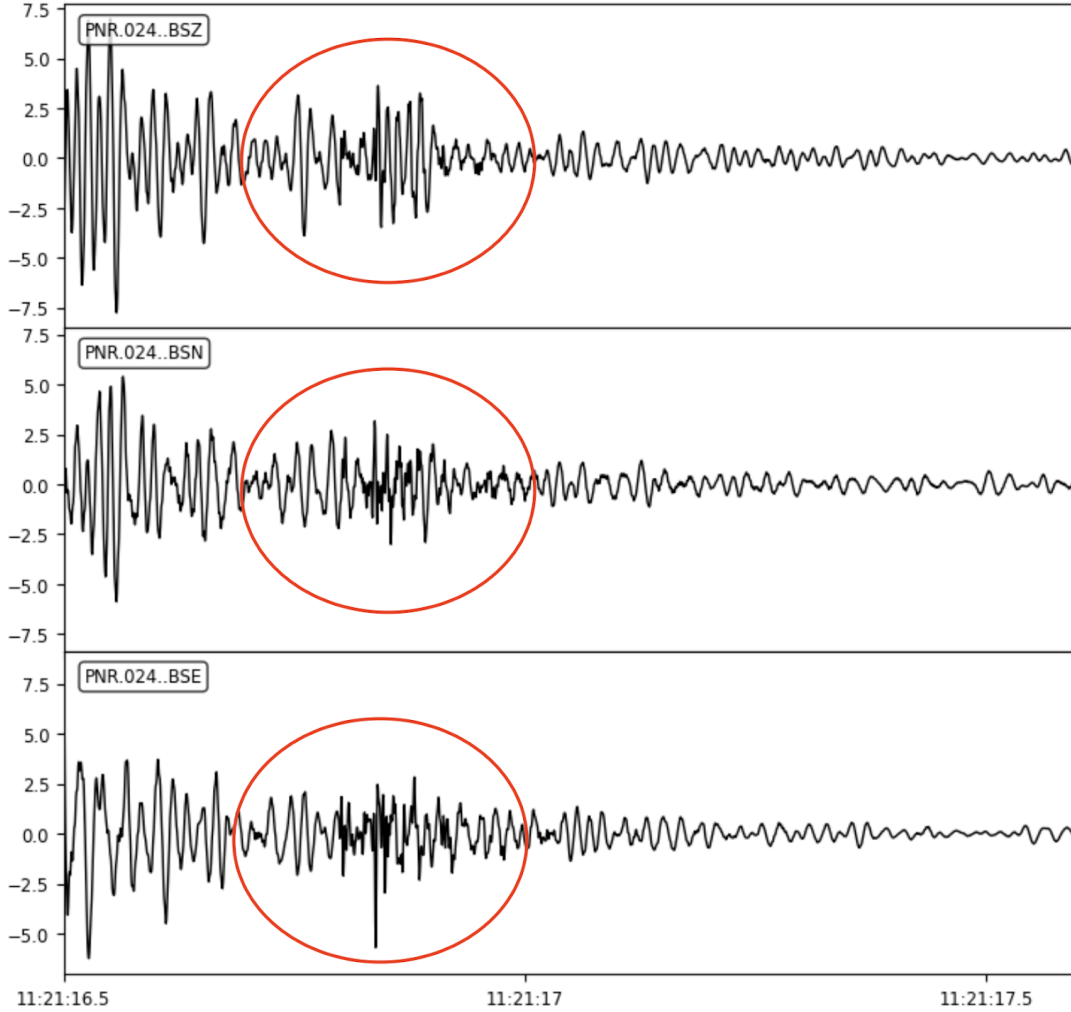


Figure 5.12: E, N and Z waveforms of a newly detected aftershock (at 11:21:16.67) within the coda of the high magnitude M_w 1.155 event on station 24. This new model event was classified as a false positive event because the phases of the event are not clear.

5.3.3 Event Location Comparison

We determined event locations for the entire 3 hour period (Figure 5.15). The locations derived by NonLinLoc with picks from the GPD model are less tightly clustered around the injection well than the CMM locations. Figure 5.16(A) illustrates the features from the additional and shared event locations. Model event locations that were also catalogued (blue) by the CMM method displayed the main NE-SW trending seismicity cloud (Feature 1) dipping towards the east and a small group of events that dip along the west (Feature 2).

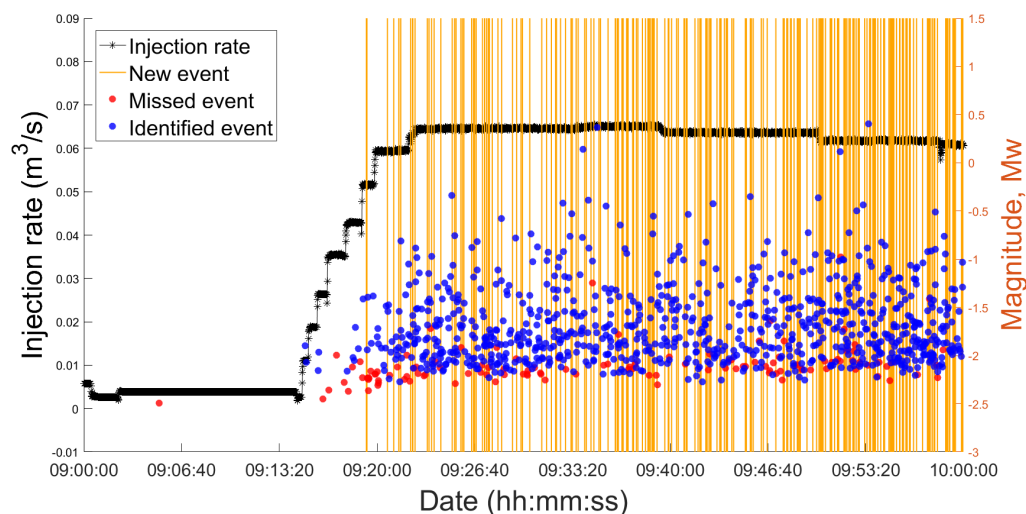


Figure 5.13: A temporal plot of the GPD model events (new events as orange lines, missed events in red and identified events in blue) for the Event Onset Period on 11 December 2018 (9am to 10am).

The new model event locations emphasise features found in both catalogues. Additional events (orange) in Figure 5.16(B) display another seismicity cloud that forms more of Feature 2 that dips west-wards. With the addition of the new model events, the seismicity better highlights a deep linear feature (Feature 3) observed in both 5.16(A) and (B). Feature 3 lies SW of Stage 38, dipping slightly towards the east (9 degrees). In the y-z plane (Figure 5.15(F)), we observed a streak of events that extend along the northing in both directions. This feature can be explained by the lack of azimuthal coverage along the North-South direction.

Figure 5.17 illustrates the RMS residuals in the GPD model picks. RMS residuals are the root mean square of the travel time residuals (or misfits) calculated for the final earthquake location. RMS residuals are commonly used as an indicator of how precise the locations are. Assuming we are aiming for good microseismic event locations to be within an error of 30 meters (average geophone spacing in the array), an event location that has RMS error > 6 ms is categorised as a bad location (using the average V_p across the geophone array depths = 5054 m/s). The model produces a median RMS pick residual of 6.75 milliseconds, corresponding to 33 meters (Figure 5.17). Figure 5.17 also shows that the 95th percentile gives an RMS error of 12.02 milliseconds, corresponding to a 60

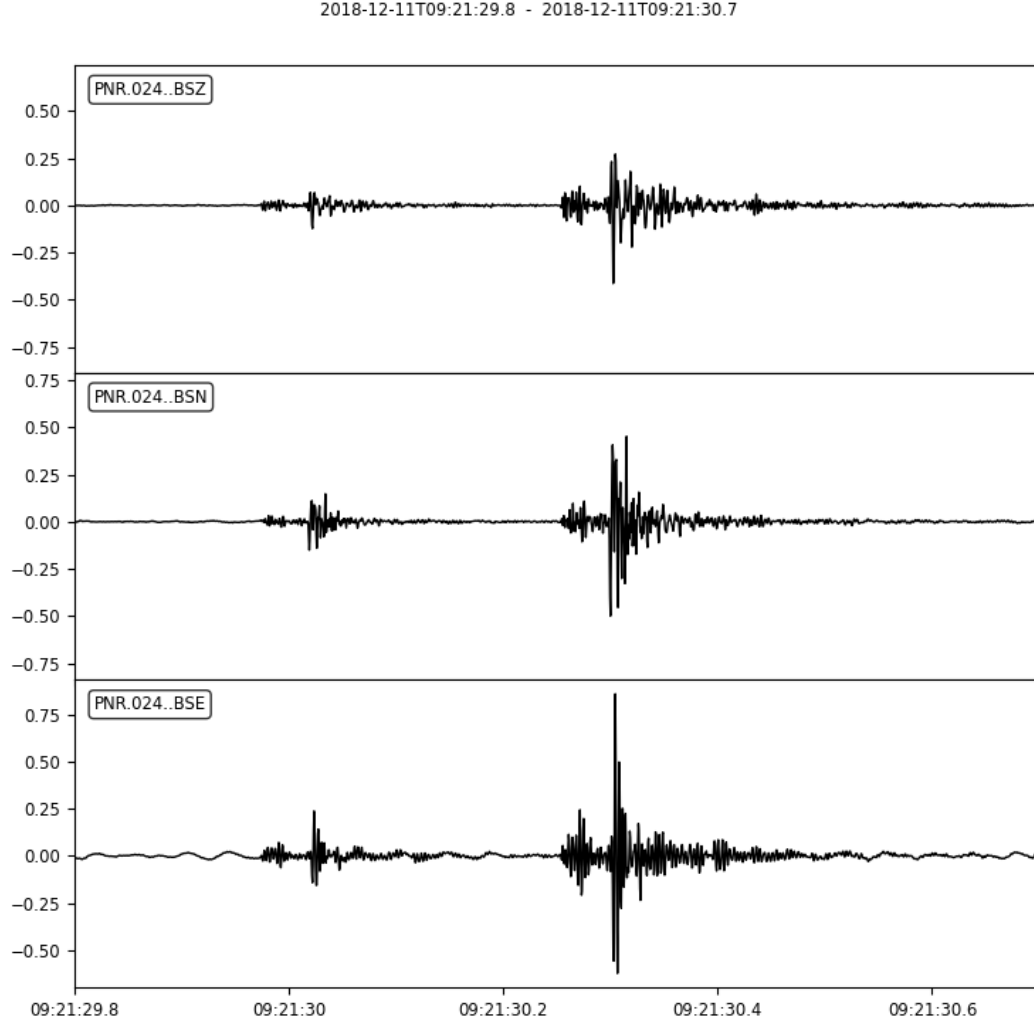


Figure 5.14: E, N and Z waveforms of the newly detected event followed by a catalogued event (detected by both the CMM and GPD model) during high injection rates on station 24.

meter uncertainty. Approximately 68% of the model locations (1401 events) have RMS error > 6 ms. Therefore, a large proportion of the events located using GPD model picks are bad locations.

We obtained horizontal and vertical errors from the NonLinLoc program for the located events (Figure 5.18). Figure 5.18(A) shows that the median and mean horizontal errors on the events are 0.2 and 0.21 km, respectively. It also shows that the 95th percentile is within 0.5 km. The event locations have median and mean vertical errors of 0.1 km and 0.12 km, respectively. 95% of the events have vertical errors below 0.3 km.

We can also observe comparable location errors looking at the spread of event locations using error probability density functions (PDFs). Figure 5.19 shows the error PDFs from 2 NonLinLoc locations; subplots A,C and E display the PDFs for the high magnitude event during the high magnitude period (also the largest event in the catalogue). The error PDF for an event during the high injection period are in subplots B, D, F on Figure 5.19.

We produced CDF graphs as a function of distance from the well for the two sets of event locations to compare event clustering around the injection well (Figure 5.20). We selected Stage 38 (Easting: 336411 m, Northing 432669 m, Depth: 2323 m) of the PNR-1z well to represent the well location because injection occurred at this stage during this three hour period (Figure 5.15). Figure 5.20 reveals that the easting, northing and depth (x, y, z) differences between the events and the well point for the CMM method cover a smaller range of distances- the 95th percentile of the easting, northing and depth differences are within 93 m, 176 m and 89 m, whereas the GPD set of locations cover a larger range. The 95th percentiles for the model location (easting, northing and depth) differences correspond to 310 m, 383 m and 372 m, respectively. The CMM locations are therefore clearly more tightly clustered (in all three dimensions) and confined to a smaller volume of the subsurface.

In addition, Figure 5.20(A) shows that more events have less event-to-well distances along the easting than along the northing for both the model and CMM locations. The shorter distances along the easting might be caused by the geometry of the array on the well (PNR-2). The array and well are oriented sub-parallel along the east (seen in Figure 5.15(A, B)). The network has a larger azimuthal coverage along the easting which results in a larger aperture; thus, it better constrains the easting locations than in the northing (Havskov et al. 2012). Depth locations are also more well-constrained than locations along the northing because the array geometry is sub-vertical (Figure 5.15(C, D)). A sub-vertical network orientation increases the aperture along depth so the array is more sensitive to depth.

In our event locations, we assumed that all high probability phase arrival picks within an event group are accurate. However, it is possible that during phase picking, a proportion of high probability picks among the 24 stations are not accurate- this can strongly influence the event locations. The location uncertainties were obtained from an

absolute location method (NonLinLoc), further re-location could reduce relative location errors and improve event clustering. We will discuss the results of the event locations and future improvements further in the Discussion section (Chapter 6).

5.3.4 Computational Runtime Comparison

In this section, we compare the computational runtime of the GPD model with other currently available event detection methods (i.e. autocorrelation, FAST and ConvNetQuake). For one hour of continuous data on 24 stations at 2000 Hz, the GPD model had a runtime of 20 minutes and 33 seconds on a NVIDIA GeForce GTX 1080 GPU. Yoon et al. (2015) and Perol et al. (2018) provided the runtimes of different earthquake detection methods, namely autocorrelation, FAST and ConvNetQuake, for 1 week of continuous data (from 8 January 2011 to 15 January 2011 in Northern California) on one station at 100 Hz. We applied the GPD model to the same dataset and compared the runtime results for event detection in Table 5.2. All the detection methods in Table 5.2 used a CPU to generate their respective results. Specifically, Perol et al. (2018) executed ConvNetQuake on an Intel i5 2.9 GHz CPU, Yoon et al. (2015) ran autocorrelation and the FAST algorithm on an Intel Xeon Processor E5-2620 2.1 GHz CPU and I ran the GPD model on an Intel Core i5 3.1 GHz CPU.

Detection Method	Autocorrelation	FAST	ConvNetQuake	GPD
Reported Runtime	9 days, 13 hours	48 min	1 min, 1 s	11 min, 4 s

Table 5.2: Computational runtimes of the available event detection methods. Autocorrelation, FAST and ConvNetQuake runtime results obtained from Yoon et al. (2015) and Perol et al. (2018).

With the exception of ConvNetQuake, the GPD model has a much shorter runtime than autocorrelation and FAST- taking only 11 minutes and 4 seconds for 1 week of continuous data on one station. For seismic event detection, the GPD model uses the 3 components (E, N and Z) of continuous data whereas the runtime tests for autocorrelation and FAST only used only one component. The GPD model takes less time than the acquisition time to process data, so it is presumably much faster than the beam-forming method (CMM) used to obtain the PNR-1z catalogue (Verdon, Pers. Comm, 8 September 2020). The GPD model would be more practical as it produces fast results on a single CPU and even faster on a GPU, rather than a supercomputer- which is required for the

CMM method to produce near real-time results.

5.3. MULTI-STATION CATALOGUE COMPARISON

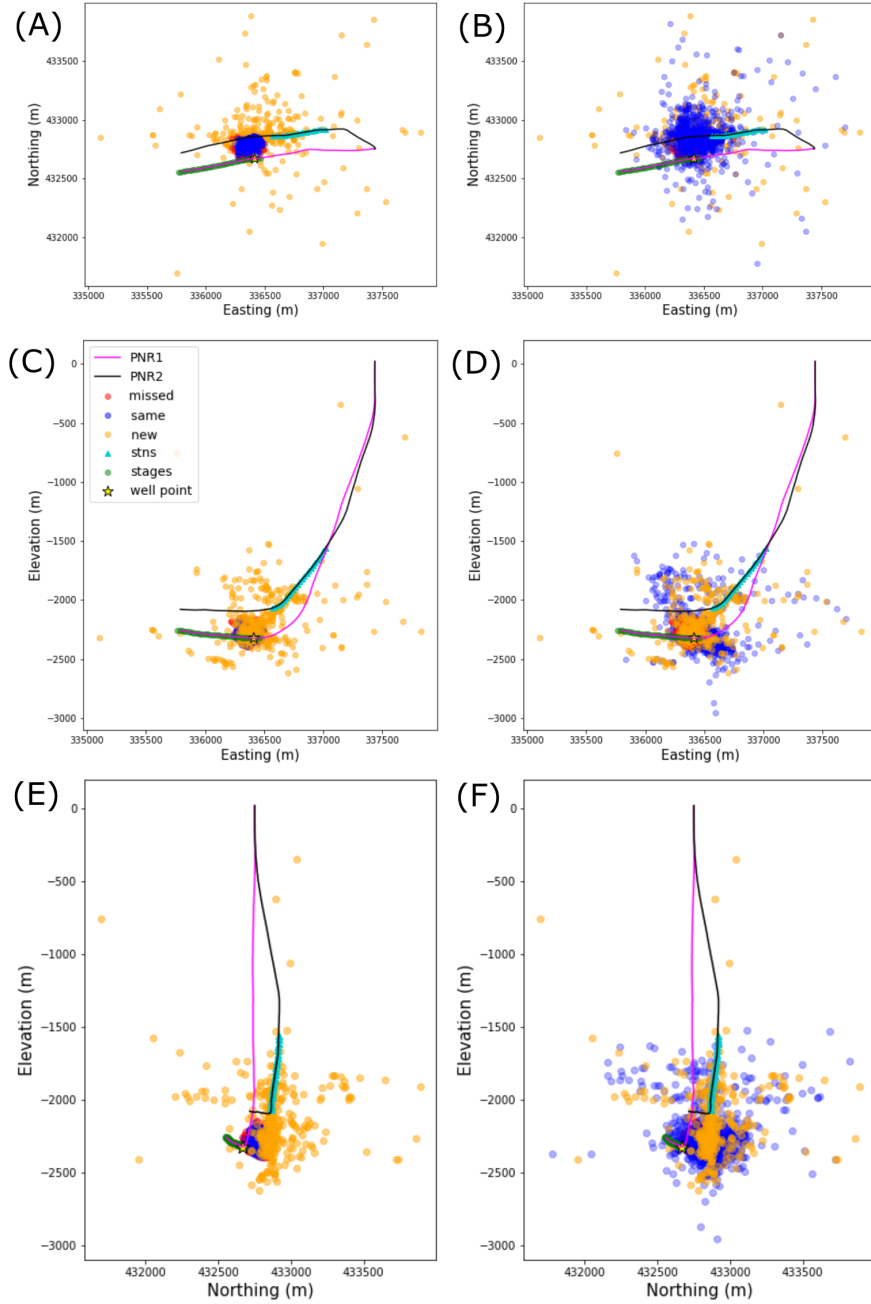


Figure 5.15: Location of the PNR-1z (magenta line) and PNR2 (black line) wells with geophones (cyan triangles) and stages (green circles) overlaid with events during the 3 hour period on 11 December 2018 (9am to 12pm). Left panels (A, C, E) display the missed (red) and same/identified (blue) event locations using the CMM catalogued locations whereas the right panels (B, D, F) use model locations. GPD model locations for new events (orange) were used in all panels. The first row shows the map view of the event locations and the subsequent two rows are cross-sections along the easting and northing, respectively. Stage 38 is where injection took place during the 3 hour period (yellow star).

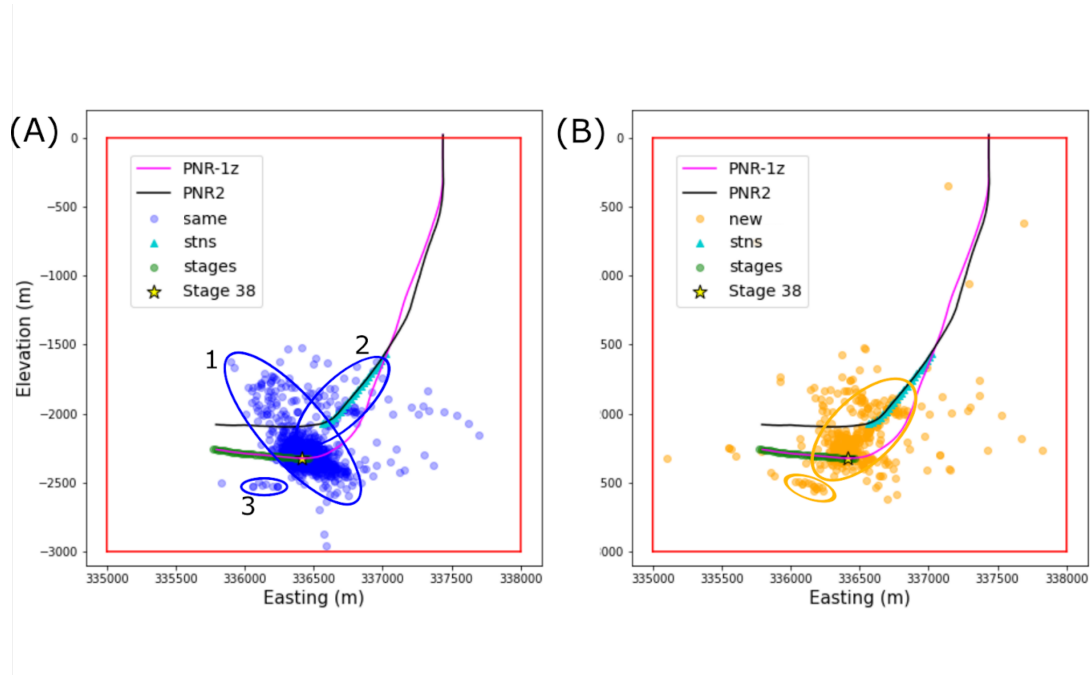


Figure 5.16: Seismic features in the cross-section along the easting on the locations of (A) the model events that were also catalogued by the CMM (blue) and (B) the new events only detected by the GPD model (orange). Magenta and black lines are the PNR-1z and PNR2 well, respectively. Red lines represent the subsurface grid set in the NonLinLoc location algorithm. Stage 38 is the representative well location (yellow star).

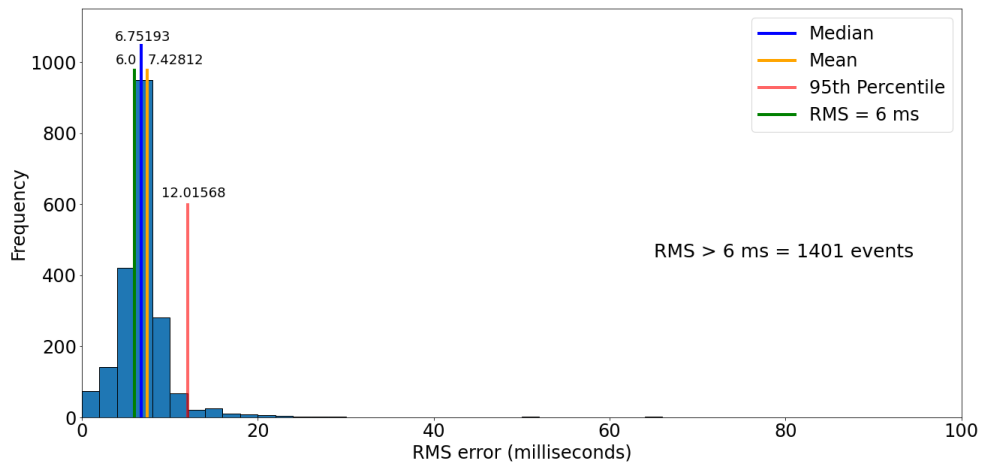


Figure 5.17: Histogram of the RMS pick residuals of the located events in milliseconds. Four events with $\text{RMS} \geq 100$ ms are off the graph. The blue, orange and red lines are the median (6.75 ms), mean (7.43 ms) and 95th percentile (12 ms) of the event picks. The green line represents the RMS error at 6 ms.

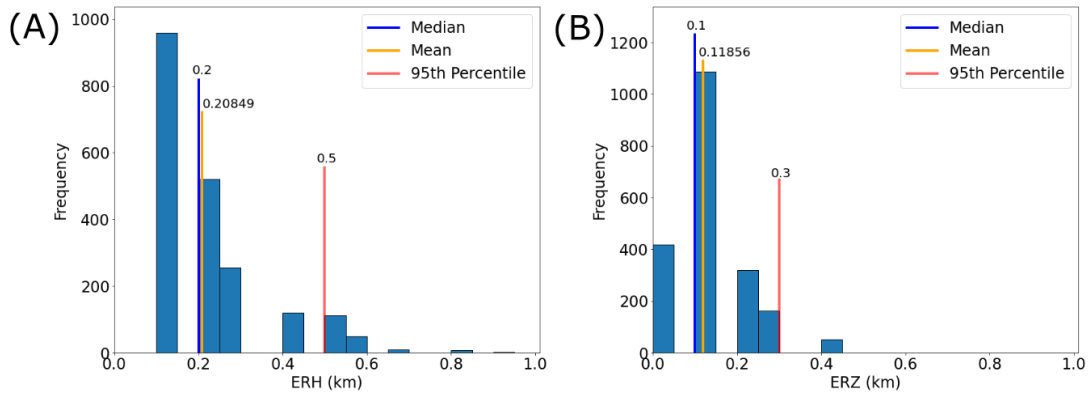


Figure 5.18: Histograms of the (A) horizontal (ERH) and (B) vertical error estimates (ERZ) on the model event locations. In both graphs, the blue, orange and red lines are the median, mean and 95th percentile.

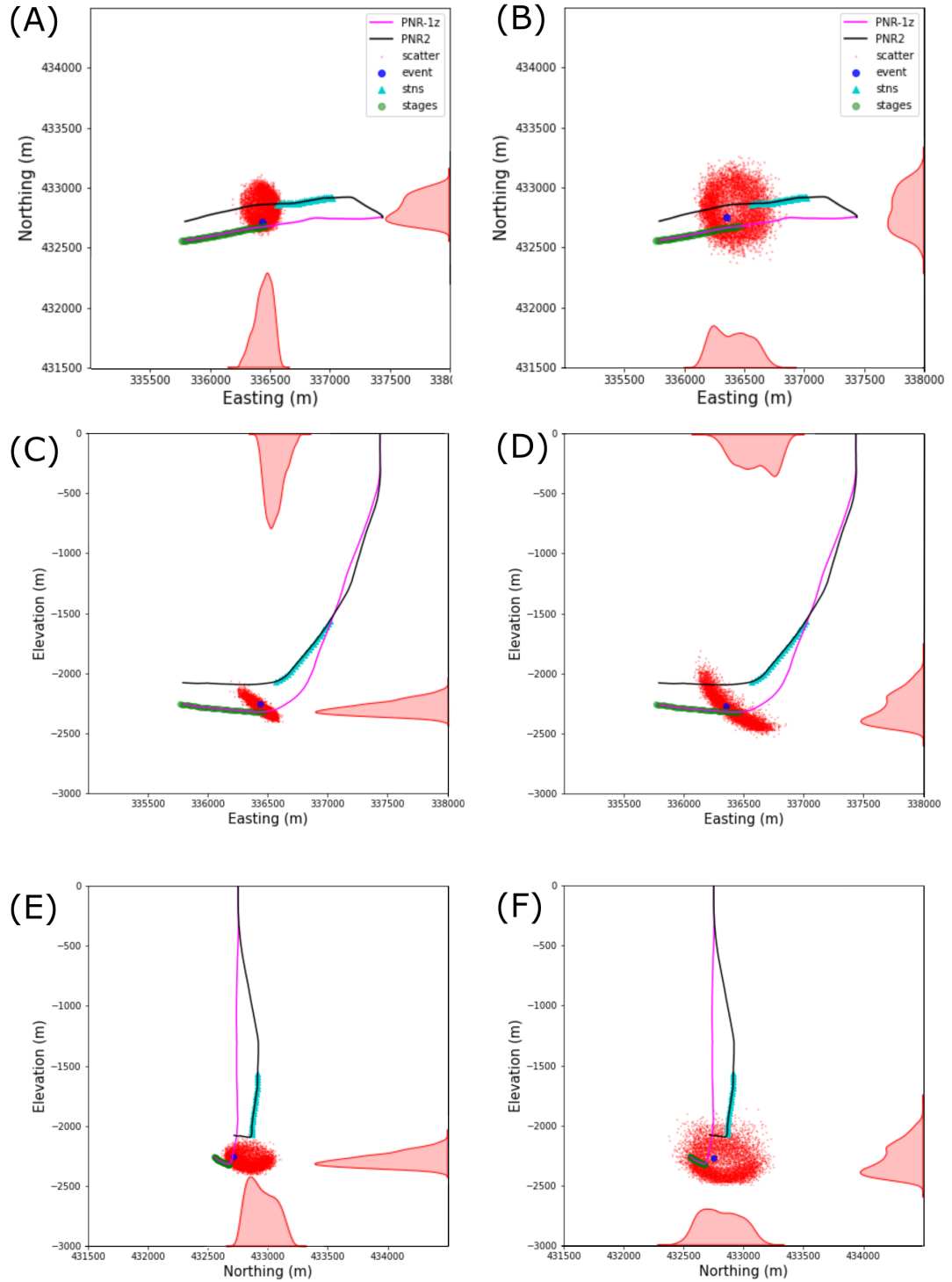


Figure 5.19: Error probability density functions (red points) for the locations (blue) of (A, C, E) the largest event in the event catalogue and (B, D, F) a M_w -1.5 event during the high injection period. The red curves on the x and y axis on all subplots are the density curves for the PDFs.

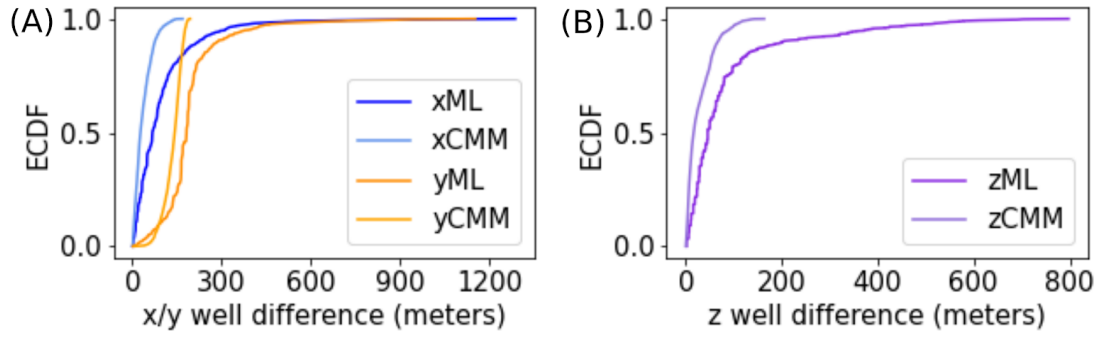


Figure 5.20: CDF of the event location difference to the well in the (A) easting (blue) and northing (orange), and along (B) depth (purple). In both panels, the darker lines represent the locations from the GPD model and the lighter shaded lines denote the CMM locations.

Chapter 6

Discussion

6.1 Single Station Classification and Multi-station Detection Performances

The GPD model’s performance deteriorated with decreasing event magnitudes in both the classification test and the multi-station detection assessment. We observed a decreasing trend in precision and recall values from the classification tests and did not detect a majority of the small ($M_w < -2$) events in the multi-station workflow. Results from the single-station classification test showed that even in the high magnitude bands, the model still missed a proportion of catalogued events (recall of 0.70) and had a number of false positives (precision of 0.74). Our values are significantly lower when compared with those achieved by Ross, Meier, Hauksson & Heaton (2018). They showed that even with varying probability thresholds, the GPD model can reach precision and recall values near or above 0.90 for classifying phases. Low precision and recall values for the lower magnitude events are expected as Ross, Meier, Hauksson & Heaton (2018) trained the GPD model with larger magnitude Southern Californian events ($-0.81 < M < 5.7$). The F1-scores derived from classification tests (a peak F1-score of 0.506 using a probability threshold of 0.02) are also low compared to other deep learning detection algorithms, including PhaseNet (Zhu & Beroza 2019). They showed that even with a 0.5 probability threshold, their deep learning classifier obtained F1-scores above 0.90 for both the P and S phases.

The multi-station workflow performed better in detecting the $M_w > -2$ events than the single station classification test. This is a direct outcome of using observations (or phase

6.1. SINGLE STATION CLASSIFICATION AND MULTI-STATION DETECTION PERFORMANCES

picks) from multiple stations. The model is more likely to detect an event on stations closer to the event as the phases will arrive more clearly. The multi-station workflow shows that the GPD model is able to detect most (81%) of the CMM catalogued events. Even with multiple stations, the GPD model, as in the single station classification test, cannot detect the smaller events well. The model can detect a few $M_w < -2$ events, but a large proportion of the missed events are in the $M_w < -2$ magnitude band. Out of the 383 missed events, 6% (21 events) of them are $M_w \geq -2$ and 94% (362 events) are $M_w < -2$. We used the CMM catalogue as the ‘ground truth’ for these comparisons and have assumed that it is a fairly complete catalogue. We visually inspected the phase arrivals of $M_w < -2$ events and confirm that they can be seen clearly in the waveforms. In other words, the low magnitude ($M_w < -2$) events are real and the model is definitely missing real events.

The failure to detect $M_w < -2$ events by both the single-station and multi-station methods may be because the signal-to-noise ratio is too low on a single station for detection. In contrast to the GPD model, the CMM method can detect these weak phases because it combines the observations of multiple stations by stacking the phase arrival signals. To explain the larger missed events ($M_w \geq -2$), the event waveforms may appear different from the generalised phase arrival trained on the GPD model. Our data are sampled at a higher frequency, were obtained from downhole geophones instead of surface stations and contain very low magnitude (down to $M_w = -2.839$) events. This means that the ability of the GPD model to classify and detect microseismic events needs to be improved for application.

A potential solution to the low detection rates of low magnitude events is to retrain or fine-tune the model to detect better on downhole data and smaller events in general. Lapins et al. (2019) used fine-tuning on the GPD model to improve earthquake detection for a seismic array at a volcanic setting. They show that the fine-tuned model can surpass off-the-shelf model performance. Lapins et al. (2019) kept the extensively pre-trained feature extraction layers in the convolutional neural network and fine-tuned by continual training with a small dataset (i.e. hundreds of events instead of thousands or millions usually needed for training a model from scratch). In the medical field, Huang et al. (2017) and Tajbakhsh et al. (2016) also show better performance on fine-tuned models using small datasets and large datasets respectively. If the fine-tuned GPD model is able to detect events better in the continuous downhole data, it might be able to outperform

the CMM with better event detection and shorter runtimes on large datasets.

6.2 Event Detection Workflow: Phase Picks, Phase Associations and Event Locations

Results in subsection 5.2.1 show that the GPD model detects more S phases than P phases. Generally, P phase arrivals for microseismic events are observed to have smaller amplitudes than the S waves (Witten et al. 2012, Kuang et al. 2013). Weak P wave onsets may be harder for the model to detect. A higher number of S picks affects the number of phase associations depending on the selected phase association method. If we associate P to S phases, we have less phase-associated pairs than when we associate S to P phases. Therefore, more events will be detected with the fixed-difference SP association method. The model also did not detect sufficiently many P phases (four or more stations) to be counted as an event group. This is an issue as the model might be missing events even though a high number of detected S phases indicate an event. This issue may also be solved by applying transfer learning to the GPD model so it can recognise and detect more P phases of weak microseismic events.

The phase associations of the workflow seem to be robust, producing a low number of phase mis-associations. Phase mis-association occurs when there are more than one S phase detected within the association window to associate with the P phase. The workflow has a short fixed phase-associating window (0.13 seconds) so it does not mis-associate frequently. When events occur frequently within a second, the window might contain arrivals from multiple events.

However, we could associate more phases by taking the longest travel-time difference between the P and S instead of the time difference of the largest event. The longest travel-time difference ($t_s - t_p$) would be between the furthest event in the catalogue on the furthest station. A longer phase associating window would increase the number of model event detections. Other more sophisticated methods could improve phase association. For example, PhaseLink by Ross et al. (2019) is a deep learning algorithm that has been trained by millions of synthetic P and S sequences to associate seismic phases, and the Hyperbolic Event eXtractor by Woollam et al. (2020) is a seismic phase associator for regions with intense seismicity. These methods are more sophisticated as they do not

6.2. EVENT DETECTION WORKFLOW: PHASE PICKS, PHASE ASSOCIATIONS AND EVENT LOCATIONS

depend on a fixed time window (like in our workflow) to associate seismic phases. Both methods have also been stress tested on large synthetic datasets, demonstrated good performances. In addition to this, both Woollam et al. (2020) and Ross et al. (2019) have mentioned further applications to microseismic monitoring.

The GPD model phase picks resulted in more scattered event locations when compared with the CMM method. The CDF of the model event locations as a function of Easting, Northing and depth differences to the well showed that the model locations cover a larger volume of the subsurface than the volume covered by CMM locations. The CMM locations displayed tighter clustering with a smaller range of well distance differences.

The proportion of inaccurate phase picks might cause the loosely scattered event locations. The picks that were used in the event locations were weighted by probability but the assigned time error windows have been observed to be too short for the pick probabilities. If the time error windows on the pick are too short, they might not accurately capture the phase (as seen in 5.2.1 of the Results section). Even with a high probability phase pick, phases can arrive outside of the time error window assigned to the model picks. We might be able to improve event locations by defining more appropriate time errors for different pick probabilities. Thus, a proper review of the distribution of expected time errors for different pick probability bands needs to be conducted to select appropriate time errors. Alternatively, choosing a different property to weight individual phase picks might help improve event locations. Instead of using pick probabilities, we could use the time difference between manually picked phase times and GPD model phase times to define the time error. However, this method would require manually picked event phases.

The model also produced a large number of high probability picks that are slightly inaccurate (by a few milliseconds) because of the phase picks represented in the training dataset. The training dataset consists of larger magnitude earthquakes ($-0.81 < M < 5.7$) and in contrast to microseismic events, these regional events usually last longer than a second. Phase pick arrivals that are off by milliseconds would not be noticeable on large regional events but would be significant on microseismic phase picks. These regional phase picks would affect the model's ability to pick on microseismic events accurately and thus severely affect event locations. Ultimately, to produce better event locations, we

would need to ensure that the model's phase picking is accurate by fine-tuning the model with accurately picked microseismic phases.

In addition to enhancing event locations by improving the model's pick accuracy, we could also improve event locations by enhancing relative event locations. This can be done by using HypoDD, a double-difference earthquake location program (Waldhauser & Ellsworth 2000), which is good at obtaining relative locations. The catalogued PNR event locations were obtained from CMM locations and appear more clustered than the present model event locations. Smith et al. (2015) and Drew et al. (2013) show that the refinement of phase arrival time picks (by event SNR values), location by Hypoinverse (Klein 2002), followed by double-difference relocation using HypoDD (Waldhauser & Ellsworth 2000) improves event locations. They reveal geological structures more clearly with tighter event clustering using HypoDD on manually refined (by event SNR values) events when compared with just the CMM-derived locations (Figure 6.1). Both manually refining events and double-difference relocation have the potential to enhance our event locations. However, manually refining events will decrease the number of events found by the model- especially when microseismic events have low event SNR values. If we want to keep the same number of events, we will just have to apply double-difference relocation on the event locations to achieve clustering with good relative event locations.

6.3 Periods of Interest

During the high magnitude period, the GPD model did detect a small number of new aftershocks. During this period, no fluid is injected and it is relatively quiet in terms of seismic activity when compared with the period of event onset during high injection rate. The new events make up 6.25% of the total events found during this period. This is lower when compared with higher seismicity during high injection rates- where 23% of the catalogue are new events. As the GPD model did not detect a lot of new events that the CMM missed, we can assume that the CMM catalogue is fairly complete for events $M_w \geq -2$ during seismically quiet periods after a high magnitude event. This assumption is for $M_w \geq -2$ events because we know the model frequently misses small ($M_w < -2$) events. It is possible that both methods could be missing the same events but it is hard to gauge without a third event detection method.

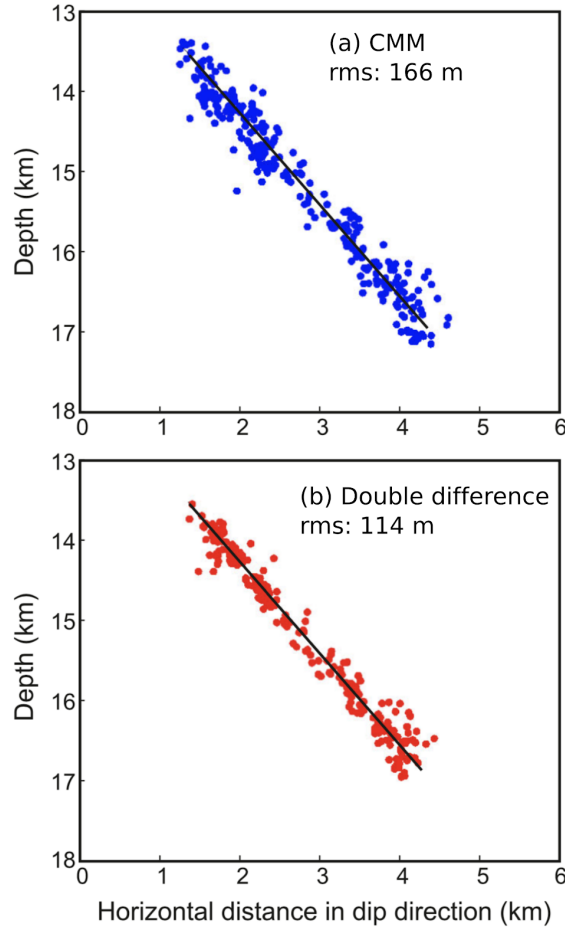


Figure 6.1: From Drew et al. (2013), the panels show the event clusters on a cross-section along the strike of a dyke intrusion in Iceland located by (a) the CMM method and (b) after manual refinement of traveltimes, Hypoinverse locations (Klein 2002) and relocations using double-difference (HypoDD).

Disregarding missed $M_w < -2$ events, the GPD model is more sensitive than the CMM method in detecting seismic events during periods of high injection. Results show that the model produces a 10% increase of detected events. The model detects events more frequently than the CMM method during injection- when the seismic rate is high. This may be because the GPD model's short 0.2 second sliding window can detect individual phases within the continuous data with every 50-sample (0.025 s) window shift. This highlights a shortcoming of the CMM method for monitoring injection activity. When seismic activity is high, the CMM method cannot detect events that occur in close temporal proximity to each other.

The CMM method missed these events because when multiple events occur within the same time window, the algorithm only uses the maximum coalescence value (the largest magnitude event in the window) to migrate into the 3-D subsurface traveltimes grid. When the injection rate increases, the rate of seismicity also increases so it is more likely for several events to be within the same time step. Therefore, the CMM could not detect these events because of the time step window defined by the user. Reducing the duration of this window will increase the number of event detections for the CMM on event onset periods during high injection but will make the process more costly and computationally intensive.

6.4 Comparison of the GPD model with Other Earthquake Detection Methods

6.4.1 Computational Runtime

The GPD model has shorter runtimes in comparison to methods like the CMM method, autocorrelation and the FAST algorithm. Shorter runtimes can provide more time for the operators to make more informed decisions for real-time microseismic monitoring during injection. This is useful for optimising project operations and increasing resource recovery rates. Faster runtimes also allow the events to be located sooner and thus will provide vital additional time to indicate fault reactivation. Short computational runtimes also provide advantages to post-hoc analysis. With improved fast processing on large datasets, it is easier to re-run analysis and achieve results quickly without the use of massive supercomputers.

6.4.2 Event Detection and New Events

We have determined that the GPD model can detect events in high frequency downhole data. From the multi-station workflow, the model managed to detect 81% of the CMM events. Most of the missed events were small ($M_w < -2$). Prior to this study, it was not known if the surface-station trained model would be able to detect events in higher sampling frequency data without first decimating the data to the same sampling frequency

during model training (100 Hz).

Other than being applicable to high resolution downhole data, the GPD model detected new events that the CMM method missed. The model has detected a lot more $M_w > -2$ events than the CMM method when the seismic rate increases during injection. The new model events can better delineate or highlight features in the subsurface quicker. Further analyses, such as magnitude estimates and better relative event locations, are needed for the new events. Good locations, along with magnitude estimates (which were beyond the scope of this thesis) could potentially help us learn more about the fluid injection and the physical mechanisms that induce these events- although this still remains to be seen. We can track the growth of seismic fracture networks by monitoring HFIS (Hainzl & Ogata 2005, Gale et al. 2014, Orlecka-Sikora et al. 2019). More located events could more quickly illuminate the migration pathways of the fracturing fluid if the induced seismicity is caused by fluid migration. This is to make sure the fracks are well-contained, not reactivating a pre-existing fault, and that the fluid is being injected into the target formation. Overall, it would be promising to utilise the model's ability to detect events during high seismic activity for the real-time monitoring of projects that involve fluid injection.

This study has found events that were not detected by both the CMM method and the GPD model (e.g. Figure 5.11 in subsection 5.3.2, Results). These events might not be detected by the CMM method because it only detects the signals of the largest event within the time step. The GPD model could not detect these events because these small events might have $M_w < -2$, and we know from previous results that the model can not detect these events well. Therefore, this gives another motivation for fine-tuning the model: after fine-tuning, the model could potentially detect additional small ($M_w < -2$) events that were not detected by the CMM method.

6.5 The GPD model, Other Deep Learning Models and HFIS

The GPD model is suitable for detecting seismic events in continuous data recorded on downhole geophones. It is also able to detect seismic events with smaller magnitudes ($M_w \geq -2$) than the training data the model was trained with ($-0.81 < M < 5.7$). This

may be because the model works reasonably well in detecting microseismicity, where the events are short enough to fit in the 400-sample (0.2 second) sliding window, on data with a high sampling frequency (2000 Hz).

There are a range of other deep learning earthquake detection models that can potentially perform better than the GPD model in detect HFIS (e.g. PhaseNet by Zhu & Beroza in 2019 and ConvNetQuake by Perol et al. in 2018). Results from comparing computation runtimes of several event detection methods suggests that the ConvNetQuake could process large datasets faster than the GPD model. Besides ConvNetQuake, PhaseNet is another deep learning model that could perform better than the GPD model. The PhaseNet model takes into account the accuracy of phase picks by producing Gaussian-shaped probability distribution peaks around each phase pick. In PhaseNet, phase arrival times are measured from the peak of the probability distribution so accuracy does not depend on when the probability threshold is exceeded (Zhu & Beroza 2019). Accurate arrival times can produce absolute event locations that will be more accurate than GPD model locations. However, we do not know how well the ConvNetQuake nor PhaseNet model can detect microseismic events (with low SNR) in high frequency downhole data. Therefore, the model ability to detect HFIS in high resolution downhole data should be explored further with comparisons between the different models.

We chose the GPD model because it is one of the most extensively trained models, and is thus expected to generalise more easily, including to high-frequency phase arrivals on geophones. Extensive training helps deep learning models to better generalise representations and recognise features that may indicate presence of a phase arrival and correctly classify waveforms to produce picktimes. Ross, Meier, Hauksson & Heaton (2018) showed the adaptability of the GPD model by detecting (larger) magnitudes out of the magnitude range of the training data. On this basis, our results partially support the general applicability of the model. The model could detect lower magnitude events out of the trained magnitude range but also struggles to detect events below $M_w < -2$. The model works reasonably well when taking into account that it has not been trained on detecting HFIS. However, the model has the potential to improve downhole event detection with fine-tuning.

Chapter 7

Conclusions

We demonstrated the flexible applicability of the GPD model- as it is able to detect seismic events within continuous data recorded on downhole geophones even though the original model was trained on larger events recorded on surface station data. Through classification tests and multi-station event detection results, we found that the performance of the off-the-shelf GPD model with high frequency downhole signals is mixed. Although the model does manage to detect microseismic events within the continuous data, it misses most $M_w < -2$ events, has a tendency to detect more S than P phases, and only has precision and recall values of 0.7 for the largest ($M_w > 0$) induced events. Tajbakhsh et al. (2016), Huang et al. (2017) and Lapins et al. (2019) have shown that fine-tuning either with large or small datasets have improved CNN model performances. On this basis, we conclude that the off-the-shelf GPD model needs to be fine-tuned to improve phase classification of hydraulic fracturing induced seismicity and event detection of microseismicity.

From the catalogue comparison between the GPD model and CMM method, we determined that the model is able to find new events that went undetected by the CMM method during the event onset period during high injection rates. The additional events found by the GPD model reveal a weakness in the CMM model to detect rapidly increasing seismicity during injection operations. Moreover, the lack of additional events (or new aftershocks) found during seismically quieter periods (when no injection occurred) after a high magnitude event ($M_w > 0$) suggests that the event catalogue of the CMM method for $M_w > -2$ events is relatively complete.

New events found during high injection rates can potentially offer insights into the physical mechanisms that trigger induced seismicity during fluid injection. These additional events could help us track fluid flow or changes in stress more easily. We can also more accurately constrain seismic rates and outline seismic features more clearly within the subsurface using the additional events.

This study also showed that event locations derived from off-the-shelf GPD phase picks produced more scattered locations compared to the more tightly clustered events located by the CMM method. The lack of clustering in model locations can be explained by the model's inaccurate picks on microseismic phases. This is a further motivation for retraining or fine-tuning the GPD model. More accurate phase picks will improve event locations.

We also demonstrated that the GPD model has a computational runtime advantage over other earthquake detection methods- such as the Fingerprint and Similarity Thresholding (FAST), autocorrelation, and the Coalescence Microseismic Mapping (CMM) method. The deep learning model yields a shorter computational runtime for the same duration of continuous data (1 week of 100 Hz data on one station). The efforts of this study show the potential of improving real-time monitoring in fluid injection operations (e.g. hydraulic fracturing, geothermal or carbon capture and storage projects) by exploiting the short runtimes offered by deep learning earthquake detection algorithms.

Appendix A

Confusion Matrices from the Classification Test on Different Event Magnitude Bands

DURING INJECTION	MODEL (PREDICTED)				
CATALOGUE (ACTUAL)	Phase	P	S	NOISE	TOTAL
	P	22	0	78	100
	S	10	7	83	100
	NOISE	0	0	100	100
	TOTAL	32	7	261	300

Table A1: Confusion matrix for 100 random-magnitude events during injection.

TOP 50 $M_w > 0$	MODEL (PREDICTED)				
CATALOGUE (ACTUAL)	Phase	P	S	NOISE	TOTAL
	P	34	0	16	50
	S	20	23	7	50
	NOISE	0	2	48	50
	TOTAL	54	25	71	150

Table A2: Confusion matrix for the 50 largest ‘high magnitude’ ($M_w > 0$) events.

BOT 50 $M_w < -2$	MODEL (PREDICTED)				
CATALOGUE (ACTUAL)	Phase	P	S	NOISE	TOTAL
	P	0	0	50	50
	S	0	0	50	50
	NOISE	0	0	50	50
	TOTAL	0	0	150	150

Table A3: Confusion matrix for the 50 smallest ‘low magnitude’ ($M_w < -2$) events.

MID 50 $-2 \leq M_w < 0$	MODEL (PREDICTED)				
CATALOGUE (ACTUAL)	Phase	P	S	NOISE	TOTAL
	P	8	4	38	50
	S	13	11	26	50
	NOISE	1	1	48	50
	TOTAL	22	16	112	150

Table A4: Confusion matrix for the 50 ‘mid magnitude’ ($-2 \leq M_w < 0$) events.

MID 25 $-1 \leq M_w < 0$	MODEL (PREDICTED)				
CATALOGUE (ACTUAL)	Phase	P	S	NOISE	TOTAL
	P	7	0	18	25
	S	9	6	10	25
	NOISE	0	0	25	25
	TOTAL	16	6	53	75

Table A5: Confusion matrix for the 25 ($-1 \leq M_w < 0$) events.

CONFUSION MATRICES FROM THE CLASSIFICATION TEST ON DIFFERENT
EVENT MAGNITUDE BANDS

MID 25 $-2 \leq M_w < -1$	MODEL (PREDICTED)				
	Phase	P	S	NOISE	TOTAL
	P	1	4	20	25
	S	4	5	16	25
	NOISE	1	1	23	25
	TOTAL	6	10	59	75

Table A6: Confusion matrix for the 25 ($-2 \leq M_w < -1$) events.

References

- Allen, R. (1982), ‘Automatic phase pickers: Their present use and future prospects’, *Bulletin of the Seismological Society of America* **72**(6B), S225–S242.
- Allen, R. V. (1978), ‘Automatic earthquake recognition and timing from single traces’, *Bulletin of the Seismological Society of America* **68**(5), 1521–1532.
- Anderson, I. & Underhill, J. R. (2020), ‘Structural constraints on lower carboniferous shale gas exploration in the craven basin, nw england’, *Petroleum Geoscience* **26**(2), 303–324.
- Apostolopoulos, I. D. & Mpesiana, T. A. (2020), ‘Covid-19: automatic detection from x-ray images utilizing transfer learning with convolutional neural networks’, *Physical and Engineering Sciences in Medicine* p. 1.
- Azad, M., Garagash, D. & Satish, M. (2017), ‘Nucleation of dynamic slip on a hydraulically fractured fault’, *Journal of Geophysical Research: Solid Earth* **122**(4), 2812–2830.
- Baisch, S., Weidler, R., Voros, R., Wyborn, D. & de Graaf, L. (2006), ‘Induced seismicity during the stimulation of a geothermal hfr reservoir in the cooper basin, australia’, *Bulletin of the Seismological Society of America* **96**(6), 2242–2256.
- Baluja, S. & Covell, M. (2008), ‘Waveprint: Efficient wavelet-based audio fingerprinting’, *Pattern recognition* **41**(11), 3467–3480.
- Bao, X. & Eaton, D. W. (2016), ‘Fault activation by hydraulic fracturing in western canada’, *Science* **354**(6318), 1406–1409.
- Barrett, S. A. & Beroza, G. C. (2014), ‘An empirical approach to subspace detection’, *Seismological Research Letters* **85**(3), 594–600.
- Bauer, A., Schwarz, B. & Gajewski, D. (2017), ‘Utilizing diffractions in wavefront tomography’, *Geophysics* **82**(2), R65–R73.

- Bergen, K. J. & Beroza, G. C. (2018), 'Detecting earthquakes over a seismic network using single-station similarity measures', *Geophysical Journal International* **213**(3), 1984–1998.
- Bergen, K. J., Johnson, P. A., Maarten, V. & Beroza, G. C. (2019), 'Machine learning for data-driven discovery in solid earth geoscience', *Science* **363**(6433), eaau0323.
- Bestmann, I. (2020), Detection and location of microseismic events using the subspace detector, Master of Science Thesis, University of Alberta.
- Bhattacharya, P. & Viesca, R. C. (2019), 'Fluid-induced aseismic fault slip outpaces pore-fluid migration', *Science* **364**(6439), 464–468.
- Bommer, J. J., Crowley, H. & Pinho, R. (2015), 'A risk-mitigation approach to the management of induced seismicity', *Journal of Seismology* **19**(2), 623–646.
- Bradford, I., Probert, T., Raymer, D., Özbek, A., Primiero, P., Kragh, E., Drew, J. & Woerpel, C. (2013), Application of coalescence microseismic mapping to hydraulic fracture monitoring conducted using a surface array, in '75th EAGE Conference & Exhibition-Workshops', European Association of Geoscientists & Engineers, pp. cp–349.
- Brown, J. R., Beroza, G. C. & Shelly, D. R. (2008), 'An autocorrelation method to detect low frequency earthquakes within tremor', *Geophysical Research Letters* **35**(16).
- Brown, S. P. & Yücel, M. K. (2013), 'Shale Gas and Tight Oil Boom: US Economic Gains and Vulnerabilities', *Council on Foreign Relations* .
- Butcher, A. (2021), personal communication.
- Cappa, F., Scuderi, M. M., Collettini, C., Guglielmi, Y. & Avouac, J.-P. (2019), 'Stabilization of fault slip by fluid injection in the laboratory and in situ', *Science advances* **5**(3).
- Chauvin, Y. & Rumelhart, D. E. (1995), *Backpropagation: theory, architectures, and applications*, Psychology Press.
- Chen, H., Meng, X., Niu, F., Tang, Y., Yin, C. & Wu, F. (2018), 'Microseismic monitoring of stimulating shale gas reservoir in sw china: 2. spatial clustering controlled by the preexisting faults and fractures', *Journal of Geophysical Research: Solid Earth* **123**(2), 1659–1672.

REFERENCES

- Clarke, H., Eisner, L., Styles, P. & Turner, P. (2014), 'Felt seismicity associated with shale gas hydraulic fracturing: The first documented example in europe', *Geophysical Research Letters* **41**(23), 8308–8314.
- Clarke, H., Soroush, H., Wood, T. et al. (2019), Preston New Road: The role of geomechanics in successful drilling of the UK's first horizontal shale gas well, in 'SPE Europepec featured at 81st EAGE Conference and Exhibition', Society of Petroleum Engineers.
- 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.
- Cuadrilla Bowland Ltd. (2018), 'Preston New Road 1z hydraulic fracture plan'. Accessed: 2020-02-19.
- URL:** https://consult.environment-agency.gov.uk/onshore-oil-and-gas/information-on-cuadrillas-preston-new-road-site/supporting_documents/Preston%20New%20Road%20HFP.pdf
- Deng, L., Hinton, G. & Kingsbury, B. (2013), New types of deep neural network learning for speech recognition and related applications: An overview, in '2013 IEEE international conference on acoustics, speech and signal processing', IEEE, pp. 8599–8603.
- Deng, L. & Liu, Y. (2018), *Deep learning in natural language processing*, Springer.
- Drew, J. E., Leslie, H. D., Armstrong, P. N., Michard, G. et al. (2005), Automated micro-seismic event detection and location by continuous spatial mapping, in 'SPE Annual Technical Conference and Exhibition', Society of Petroleum Engineers.
- Drew, J., White, R. S., Tilmann, F. & Tarasewicz, J. (2013), 'Coalescence microseismic mapping', *Geophysical Journal International* **195**(3), 1773–1785.
- Duveneck, E. (2004), 'Velocity model estimation with data-derived wavefront attributes', *Geophysics* **69**(1), 265–274.
- Dziewonski, A., Chou, T.-A. & Woodhouse, J. (1981), 'Determination of earthquake source parameters from waveform data for studies of global and regional seismicity', *Journal of Geophysical Research: Solid Earth* **86**(B4), 2825–2852.
- Ellsworth, W. L. (2013), 'Injection-induced earthquakes', *Science* **341**(6142).

- 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.
- Gajewski, D. & Tessmer, E. (2005), 'Reverse modelling for seismic event characterization', *Geophysical Journal International* **163**(1), 276–284.
- Gale, J. F., Laubach, S. E., Olson, J. E., Eichhubl, P. & Fall, A. (2014), 'Natural fractures in shale: A review and new observations', *AAPG bulletin* **98**(11), 2165–2216.
- Gibbons, S. J. & Ringdal, F. (2006), 'The detection of low magnitude seismic events using array-based waveform correlation', *Geophysical Journal International* **165**(1), 149–166.
- Goebel, T. H. & Brodsky, E. E. (2018), 'The spatial footprint of injection wells in a global compilation of induced earthquake sequences', *Science* **361**(6405), 899–904.
- Gopalakrishnan, K., Khaitan, S. K., Choudhary, A. & Agrawal, A. (2017), 'Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection', *Construction and Building Materials* **157**, 322–330.
- Green, C. A., Styles, P. & Baptie, B. J. (2012), 'Preese hall shale gas fracturing review and recommendations for induced seismic mitigation'.
- Gregory, K. B., Vidic, R. D. & Dzombak, D. A. (2011), 'Water management challenges associated with the production of shale gas by hydraulic fracturing', *Elements* **7**(3), 181–186.
- Grigoli, F., Cesca, S., Priolo, E., Rinaldi, A. P., Clinton, J. F., Stabile, T. A., Dost, B., Fernandez, M. G., Wiemer, S. & Dahm, T. (2017), 'Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A european perspective', *Reviews of Geophysics* **55**(2), 310–340.
- 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 mw 5.5 pohang earthquake: A possible case of induced seismicity in south korea', *Science* **360**(6392), 1003–1006.
- Grigoli, F., Cesca, S., Vassallo, M. & Dahm, T. (2013), 'Automated seismic event location by travel-time stacking: An application to mining induced seismicity', *Seismological Research Letters* **84**(4), 666–677.

REFERENCES

- Hainzl, S. & Ogata, Y. (2005), ‘Detecting fluid signals in seismicity data through statistical earthquake modeling’, *Journal of Geophysical Research: Solid Earth* **110**(B5).
- Harris, D. B. (2006), Subspace detectors: theory, Technical report, Lawrence Livermore National Lab.(LLNL), Livermore, CA (United States).
- Havskov, J., Bormann, P. & Schweitzer, J. (2012), Seismic source location, in ‘New Manual of Seismological Observatory Practice 2 (NMSOP-2)’, Deutsches GeoForschungsZentrum GFZ, pp. 1–36.
- Herwanger, J. V., Mohamed, F. R., Newman, R. & Vejbaek, O. (2013), Time-lapse seismic data-calibrated geomechanical model reveals hydraulic fracture re-orientation, in ‘SEG Technical Program Expanded Abstracts 2013’, Society of Exploration Geophysicists, pp. 4949–4953.
- Hitzman, M. W., Clarke, D. D., Detournay, E., Dietrich, J., Dillon, D., Green, S. & Gibbs, C. (2012), Induced seismicity potential in energy technologies, in ‘GSA meeting, 4-7 Nov 2012, Charlotte, NC, USA’.
- Huang, L., Li, J., Hao, H. & Li, X. (2018), ‘Micro-seismic event detection and location in underground mines by using convolutional neural networks (cnn) and deep learning’, *Tunnelling and Underground Space Technology* **81**, 265–276.
- Huang, Z., Pan, Z. & Lei, B. (2017), ‘Transfer learning with deep convolutional neural network for sar target classification with limited labeled data’, *Remote Sensing* **9**(9), 907.
- Hubbert, M. K. & Willis, D. G. (1972), ‘Mechanics of hydraulic fracturing’.
- Huynh, B. Q., Li, H. & Giger, M. L. (2016), ‘Digital mammographic tumor classification using transfer learning from deep convolutional neural networks’, *Journal of Medical Imaging* **3**(3), 034501.
- Igonin, N. & Innanen, K. A. (2018), Analysis of simultaneous velocity and source parameter updates in microseismic fwi, in ‘SEG Technical Program Expanded Abstracts 2018’, Society of Exploration Geophysicists.
- Kagan, Y. Y. (2004), ‘Short-term properties of earthquake catalogs and models of earthquake source’, *Bulletin of the Seismological Society of America* **94**(4), 1207–1228.

- Kamei, R., Nakata, N. & Lumley, D. (2015), 'Introduction to microseismic source mechanisms', *The Leading Edge* **34**(8), 876–880.
- Kao, H. & Shan, S.-J. (2004), 'The source-scanning algorithm: Mapping the distribution of seismic sources in time and space', *Geophysical Journal International* **157**(2), 589–594.
- Kendall, J.-M., Butcher, A., Stork, A. L., Verdon, J. P., Luckett, R. & Baptie, B. J. (2019), 'How big is a small earthquake? challenges in determining microseismic magnitudes', *First Break* **37**(2), 51–56.
- Keranen, K. M., Savage, H. M., Abers, G. A. & Cochran, E. S. (2013), 'Potentially induced earthquakes in oklahoma, usa: Links between wastewater injection and the 2011 mw 5.7 earthquake sequence', *Geology* **41**(6), 699–702.
- Kettlety, T., Verdon, J., Werner, M. & Kendall, J. (2020), 'Stress transfer from opening hydraulic fractures controls the distribution of induced seismicity', *Journal of Geophysical Research: Solid Earth* **125**(1), e2019JB018794.
- Kim, K.-H., Ree, J.-H., Kim, Y., Kim, S., Kang, S. Y. & Seo, W. (2018), 'Assessing whether the 2017 mw 5.4 pohang earthquake in south korea was an induced event', *Science* **360**(6392), 1007–1009.
- Kim, S. & Hosseini, S. A. (2017), 'Study on the ratio of pore-pressure/stress changes during fluid injection and its implications for co2 geologic storage', *Journal of Petroleum Science and Engineering* **149**, 138–150.
- Kislov, K., Gravurov, V. & Vinberg, F. (2020), 'Possibilities of seismic data preprocessing for deep neural network analysis', *Izvestiya, Physics of the Solid Earth* **56**(1), 133–144.
- Klein, F. W. (2002), User's guide to hypoinverse-2000, a fortran program to solve for earthquake locations and magnitudes, Technical report, US Geological Survey.
- Kohli, A. H. & Zoback, M. D. (2013), 'Frictional properties of shale reservoir rocks', *Journal of geophysical research: solid earth* **118**(9), 5109–5125.
- Kong, Q., Trugman, D. T., Ross, Z. E., Bianco, M. J., Meade, B. J. & Gerstoft, P. (2019), 'Machine learning in seismology: Turning data into insights', *Seismological Research Letters* **90**(1), 3–14.

REFERENCES

- Kozłowska, M., Brudzinski, M. R., Friberg, P., Skoumal, R. J., Baxter, N. D. & Currie, B. S. (2018), ‘Maturity of nearby faults influences seismic hazard from hydraulic fracturing’, *Proceedings of the National Academy of Sciences* **115**(8), E1720–E1729.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2017), ‘Imagenet classification with deep convolutional neural networks’, *Communications of the ACM* **60**(6), 84–90.
- Kuang, W., Zhang, W. & Zhang, J. (2013), Locating microseismic events with s-wave data only, in ‘SEG Technical Program Expanded Abstracts 2013’, Society of Exploration Geophysicists, pp. 2243–2247.
- Lapins, S., Goitom, B., Kendall, J. M., Cashman, K. V. & Hammond, J. (2019), ‘Transfer learning for automated seismic phase arrival detection on local or temporary volcano-seismic networks with limited seismic catalogue (S43D–0691)’, *AGU Fall Meeting 2019*.
- Larmat, C., Tromp, J., Liu, Q. & Montagner, J.-P. (2008), ‘Time reversal location of glacial earthquakes’, *Journal of Geophysical Research: Solid Earth* **113**(B9).
- Le Calvez, J. H., Bennett, L., Tanner, K. V., Grant, W. D., Nutt, L., Jochen, V., Underhill, W. & Drew, J. (2005), ‘Monitoring microseismic fracture development to optimize stimulation and production in aging fields’, *The Leading Edge* **24**(1), 72–75.
- Le Calvez, J. H., Craven, M. E., Klem, R. C., Baihly, J. D., Bennett, L. A., Brook, K. et al. (2007), Real-time microseismic monitoring of hydraulic fracture treatment: A tool to improve completion and reservoir management, in ‘SPE hydraulic fracturing technology conference’, Society of Petroleum Engineers.
- LeCun, Y., Bengio, Y. & Hinton, G. (2015), ‘Deep learning’, *nature* **521**(7553), 436–444.
- 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 mw4. 7 induced by shale-gas hydraulic fracturing in sichuan basin, china’, *Scientific reports* **7**(1), 1–12.
- Li, L., Tan, J., Schwarz, B., Staněk, F., Poiata, N., Shi, P., Diekmann, L., Eisner, L. & Gajewski, D. (2020), ‘Recent advances and challenges of waveform-based seismic location methods at multiple scales’, *Reviews of Geophysics* **58**(1), e2019RG000667.
- 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.

- Liu, H. & Zhang, J. (2014), 'Sta/Ita algorithm analysis and improvement of microseismic signal automatic detection', *Progress in Geophysics* **29**(4), 1708–1714.
- Liu, J. & Zahradník, J. (2020), 'The 2019 mw 5.7 changning earthquake, sichuan basin, china: A shallow doublet with different faulting styles', *Geophysical Research Letters* **47**(4), e2019GL085408.
- Lomax, A., Virieux, J., Volant, P. & Berge-Thierry, C. (2000), Probabilistic earthquake location in 3d and layered models, in 'Advances in seismic event location', Springer, pp. 101–134.
- Majer, E. L., Baria, R., Stark, M., Oates, S., Bommer, J., Smith, B. & Asanuma, H. (2007), 'Induced seismicity associated with enhanced geothermal systems', *Geothermics* **36**(3), 185–222.
- Maxwell, S. C., Rutledge, J., Jones, R. & Fehler, M. (2010), 'Petroleum reservoir characterization using downhole microseismic monitoring', *Geophysics* **75**(5), 75A129–75A137.
- McGarr, A., Simpson, D., Seeber, L. & Lee, W. (2002), 'Case histories of induced and triggered seismicity', *International Geophysics Series* **81**(A), 647–664.
- Michel, O. J. & Tsvankin, I. (2014), 'Gradient calculation for waveform inversion of microseismic data in vti media', *Journal of Seismic Exploration* **23**(3), 201–217.
- Mignan, A., Broccardo, M., Wiemer, S. & Giardini, D. (2017), 'Induced seismicity closed-form traffic light system for actuarial decision-making during deep fluid injections', *Scientific reports* **7**(1), 1–10.
- Mignan, A., Landtwing, D., Kästli, P., Mena, B. & Wiemer, S. (2015), 'Induced seismicity risk analysis of the 2006 basel, switzerland, enhanced geothermal system project: Influence of uncertainties on risk mitigation', *Geothermics* **53**, 133–146.
- Mousavi, S. M., Zhu, W., Sheng, Y. & Beroza, G. C. (2019), 'Cred: A deep residual network of convolutional and recurrent units for earthquake signal detection', *Scientific reports* **9**(1), 1–14.
- Mukuhira, Y., Ito, T., Asanuma, H. & Häring, M. (2020), 'Evaluation of flow paths during stimulation in an egs reservoir using microseismic information', *Geothermics* **87**, 101843.

REFERENCES

- Nakamura, M. (2002), 'Determination of focal mechanism solution using initial motion polarity of p and s waves', *Physics of the Earth and Planetary Interiors* **130**(1-2), 17–29.
- Netherlands Government (2014), 'Natural gas production reduced and funds earmarked for Groningen.'
Accessed: 2020-04-20.
URL: www.government.nl/latest/news/2014/01/17/natural-gas-production-reduced-and-funds-earmarked-for-groningen
- Oil and Gas Authority (2018), 'PNR-1z hydraulic fracturing operations data'.
Accessed: 2020-02-19.
URL: <https://www.ogauthority.co.uk/exploration-production/onshore/onshore-reports-and-data/preston-new-road-pnr-1z-hydraulic-fracturing-operations-data/>
- Orlecka-Sikora, B., Cielesta, S. & Lasocki, S. (2019), 'Tracking the development of seismic fracture network from the geysers geothermal field', *Acta Geophysica* **67**(1), 341–350.
- Patterson, J. & Gibson, A. (2017), *Deep learning: A practitioner's approach*, " O'Reilly Media, Inc."
- Peng, Z., Vidale, J. E., Ishii, M. & Helmstetter, A. (2007), 'Seismicity rate immediately before and after main shock rupture from high-frequency waveforms in japan', *Journal of Geophysical Research: Solid Earth* **112**(B3).
- Perol, T., Gharbi, M. & Denolle, M. (2018), 'Convolutional neural network for earthquake detection and location', *Science Advances* **4**(2), e1700578.
- Pugh, D., White, R. & Christie, P. (2016), 'A bayesian method for microseismic source inversion', *Geophysical Journal International* **206**(2), 1009–1038.
- Pytharouli, S. I., Lunn, R. J., Shipton, Z. K., Kirkpatrick, J. D. & do Nascimento, A. F. (2011), 'Microseismicity illuminates open fractures in the shallow crust', *Geophysical Research Letters* **38**(2).
- Raleigh, C., Healy, J. & Bredehoeft, J. (1976), 'An experiment in earthquake control at rangely, colorado', *Science* **191**(4233), 1230–1237.
- Romanowicz, B. (2008), 'Using seismic waves to image earth's internal structure', *Nature* **451**(7176), 266–268.

- Ross, Z. E., Meier, M.-A. & Hauksson, E. (2018), 'P wave arrival picking and first-motion polarity determination with deep learning', *Journal of Geophysical Research: Solid Earth* **123**(6), 5120–5129.
- 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.
- Ross, Z. E., Yue, Y., Meier, M.-A., Hauksson, E. & Heaton, T. H. (2019), 'Phaselink: A deep learning approach to seismic phase association', *Journal of Geophysical Research: Solid Earth* **124**(1), 856–869.
- Rutqvist, J., Rinaldi, A. P., Cappa, F. & Moridis, G. J. (2013), 'Modeling of fault reactivation and induced seismicity during hydraulic fracturing of shale-gas reservoirs', *Journal of Petroleum Science and Engineering* **107**, 31–44.
- Rutter, E. & Hackston, A. (2017), 'On the effective stress law for rock-on-rock frictional sliding, and fault slip triggered by means of fluid injection', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **375**(2103), 20160001.
- Sabbione, J. I. & Velis, D. R. (2013), 'A robust method for microseismic event detection based on automatic phase pickers', *Journal of Applied Geophysics* **99**, 42–50.
- Savvaidis, A., Lomax, A. & Breton, C. (2020), 'Induced seismicity in the delaware basin, west texas, is caused by hydraulic fracturing and wastewater disposal', *Bulletin of the Seismological Society of America* **110**(5), 2225–2241.
- Schaff, D. P. & Waldhauser, F. (2005), 'Waveform cross-correlation-based differential travel-time measurements at the northern california seismic network', *Bulletin of the Seismological Society of America* **95**(6), 2446–2461.
- 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.
- Scuderi, M. M. & Collettini, C. (2018), 'Fluid injection and the mechanics of frictional stability of shale-bearing faults', *Journal of Geophysical Research: Solid Earth* **123**(10), 8364–8384.

REFERENCES

- Segall, P. & Lu, S. (2015), 'Injection-induced seismicity: Poroelastic and earthquake nucleation effects', *Journal of Geophysical Research: Solid Earth* **120**(7), 5082–5103.
- Shen, D., Wu, G. & Suk, H.-I. (2017), 'Deep learning in medical image analysis', *Annual review of biomedical engineering* **19**, 221–248.
- Shin, H.-C., Roth, H. R., Gao, M., Lu, L., Xu, Z., Nogues, I., Yao, J., Mollura, D. & Summers, R. M. (2016), 'Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning', *IEEE transactions on medical imaging* **35**(5), 1285–1298.
- Shirzaei, M., Ellsworth, W. L., Tiampo, K. F., González, P. J. & Manga, M. (2016), 'Surface uplift and time-dependent seismic hazard due to fluid injection in eastern texas', *Science* **353**(6306), 1416–1419.
- Silva-Palacios, D., Ferri, C. & Ramírez-Quintana, M. J. (2017), 'Improving performance of multiclass classification by inducing class hierarchies', *Procedia Computer Science* **108**, 1692–1701.
- Skoumal, R. J., Brudzinski, M. R. & Currie, B. S. (2016), 'An efficient repeating signal detector to investigate earthquake swarms', *Journal of Geophysical Research: Solid Earth* **121**(8), 5880–5897.
- Skoumal, R. J., Brudzinski, M. R., Currie, B. S. & Levy, J. (2014), 'Optimizing multi-station earthquake template matching through re-examination of the youngstown, ohio, sequence', *Earth and Planetary Science Letters* **405**, 274–280.
- Skoumal, R. J., Ries, R., Brudzinski, M. R., Barbour, A. J. & Currie, B. S. (2018), 'Earthquakes induced by hydraulic fracturing are pervasive in oklahoma', *Journal of Geophysical Research: Solid Earth* **123**(12), 10–918.
- Slaney, M. & Casey, M. (2008), 'Locality-sensitive hashing for finding nearest neighbors [lecture notes]', *IEEE Signal processing magazine* **25**(2), 128–131.
- 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.
- Sokolova, M., Japkowicz, N. & Szpakowicz, S. (2006), Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation, in 'Australasian joint conference on artificial intelligence', Springer, pp. 1015–1021.

- Sone, H. & Zoback, M. D. (2013), ‘Mechanical properties of shale-gas reservoir rocks—part 1: Static and dynamic elastic properties and anisotropy’, *Geophysics* **78**(5), D381–D392.
- Song, F., Kuleli, H. S., Toksöz, M. N., Ay, E. & Zhang, H. (2010), ‘An improved method for hydrofracture-induced microseismic event detection and phase picking’, *Geophysics* **75**(6), A47–A52.
- Spencer, C. & Gubbins, D. (1980), ‘Travel-time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media’, *Geophysical Journal International* **63**(1), 95–116.
- Tajbakhsh, N., Shin, J. Y., Gurudu, S. R., Hurst, R. T., Kendall, C. B., Gotway, M. B. & Liang, J. (2016), ‘Convolutional neural networks for medical image analysis: Full training or fine tuning?’, *IEEE transactions on medical imaging* **35**(5), 1299–1312.
- Tarasewicz, J., Brandsdóttir, B., White, R. S., Hensch, M. & Thorbjarnardóttir, B. (2012), ‘Using microearthquakes to track repeated magma intrusions beneath the eyjafjallajökull stratovolcano, iceland’, *Journal of Geophysical Research: Solid Earth* **117**(B9).
- Tharwat, A. (2020), ‘Classification assessment methods’, *Applied Computing and Informatics* .
- Trnkoczy, A. (2009), Understanding and parameter setting of sta/lta trigger algorithm, in ‘New Manual of Seismological Observatory Practice (NMSOP)’, Deutsches Geo-ForschungsZentrum GFZ, pp. 1–20.
- UK Government, Department for Business, Energy & Industrial Strategy (2013), ‘Traffic light monitoring system (shale gas and fracking)’.
Accessed: 2020-10-20.
URL: www.gov.uk/government/publications/traffic-light-monitoring-system-shale-gas-and-fracking
- Van Der Baan, M., Eaton, D., Dusseault, M. et al. (2013), Microseismic monitoring developments in hydraulic fracture stimulation, in ‘Isrm international conference for effective and sustainable hydraulic fracturing’, International Society for Rock Mechanics and Rock Engineering.
- van Thienen-Visser, K. & Breunese, J. (2015), ‘Induced seismicity of the groningen gas field: History and recent developments’, *The Leading Edge* **34**(6), 664–671.

REFERENCES

- Verdon, J. (2020), personal communication.
- 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.
- Viesca, R. C. (2015), Elastic stress transfer as a diffusive process due to aseismic fault slip in response to fluid injection, in 'AGU Fall Meeting Abstracts', Vol. 2015, pp. MR41E–02.
- Vlek, C. (2019), 'Rise and reduction of induced earthquakes in the groningen gas field, 1991–2018: statistical trends, social impacts, and policy change', *Environmental earth sciences* **78**(3), 59.
- Waldhauser, F. & Ellsworth, W. L. (2000), 'A double-difference earthquake location algorithm: Method and application to the northern hayward fault, california', *Bulletin of the Seismological Society of America* **90**(6), 1353–1368.
- Wang, J., Xiao, Z., Liu, C., Zhao, D. & Yao, Z. (2019), 'Deep learning for picking seismic arrival times', *Journal of Geophysical Research: Solid Earth* **124**(7), 6612–6624.
- Wang, Z. & Krupnick, A. (2015), 'A retrospective review of shale gas development in the united states: What led to the boom?', *Economics of Energy & Environmental Policy* **4**(1), 5–18.
- Warpinski, N. R., Du, J. et al. (2010), Source-mechanism studies on microseismicity induced by hydraulic fracturing, in 'SPE Annual Technical Conference and Exhibition', Society of Petroleum Engineers.
- Warpinski, N. R., Mayerhofer, M., Agarwal, K., Du, J. et al. (2013), 'Hydraulic-fracture geomechanics and microseismic-source mechanisms', *SPE Journal* **18**(04), 766–780.
- Whaley, J. (2009), *The Groningen gas field*, GEO ExPro Magazine.
- Wilkins, A. H., Strange, A., Duan, Y. & Luo, X. (2020), 'Identifying microseismic events in a mining scenario using a convolutional neural network', *Computers & Geosciences* **137**, 104418.
- Witten, B., Montgomery, S. & Artman, B. (2012), Shear wave arrivals in surface microseismic data, in 'SEG Technical Program Expanded Abstracts 2012', Society of Exploration Geophysicists, pp. 1–5.

- Woollam, J., Rietbrock, A., Bueno, A. & De Angelis, S. (2019), 'Convolutional neural network for seismic phase classification, performance demonstration over a local seismic network', *Seismological Research Letters* **90**(2A), 491–502.
- Woollam, J., Rietbrock, A., Leitloff, J. & Hinz, S. (2020), 'HEX: Hyperbolic event extractor, a seismic phase associator for highly active seismic regions', *Seismological Society of America* **91**(5), 2769–2778.
- Wu, Y., Lin, Y., Zhou, Z. & Delorey, A. (2018), 'Seismic-net: A deep densely connected neural network to detect seismic events', *arXiv preprint arXiv:1802.02241*.
- Yang, S., Hu, J., Zhang, H. & Liu, G. (2021), 'Simultaneous earthquake detection on multiple stations via a convolutional neural network', *Seismological Society of America* **92**(1), 246–260.
- Yoon, C. E., Huang, Y., Ellsworth, W. L. & Beroza, G. C. (2017), 'Seismicity during the initial stages of the Guy-Greenbrier, Arkansas, earthquake sequence', *Journal of Geophysical Research: Solid Earth* **122**(11), 9253–9274.
- Yoon, C. E., O'Reilly, O., Bergen, K. J. & Beroza, G. C. (2015), 'Earthquake detection through computationally efficient similarity search', *Science advances* **1**(11), e1501057.
- Zelt, C. & Smith, R. (1992), 'Seismic traveltime inversion for 2-d crustal velocity structure', *Geophysical journal international* **108**(1), 16–34.
- Zhai, G., Shirzaei, M., Manga, M. & Chen, X. (2019), 'Pore-pressure diffusion, enhanced by poroelastic stresses, controls induced seismicity in oklahoma', *Proceedings of the National Academy of Sciences* **116**(33), 16228–16233.
- Zhang, X., Yuan, C., Zhang, J., Liu, S., Chen, Z. & Li, W. (2018), Automatic microseismic detection and location via the deep-convolutional neural network, in 'SEG Technical Program Expanded Abstracts 2018', Society of Exploration Geophysicists, pp. 3057–3061.
- Zhu, W. & Beroza, G. C. (2019), 'Phasenet: a deep-neural-network-based seismic arrival-time picking method', *Geophysical Journal International* **216**(1), 261–273.
- Zoback, M. D. & Gorelick, S. M. (2012), 'Earthquake triggering and large-scale geologic storage of carbon dioxide', *Proceedings of the National Academy of Sciences* **109**(26), 10164–10168.

REFERENCES

Zou, C. (2017), *Unconventional petroleum geology*, Elsevier.

Zou, Q., Xie, S., Lin, Z., Wu, M. & Ju, Y. (2016), 'Finding the best classification threshold in imbalanced classification', *Big Data Research* **5**, 2–8.