Using Deep Neural Networks to Pick Phase Arrivals and Locate Microseismic Events in Askja, Iceland

Paul Charles Shuker
Supervisor: Prof. Nicholas Rawlinson
University of Cambridge

May, 2023

Online day-to-day notebook.

Except where specific reference is made to the work of others, this work is original and has not been already submitted either wholly or in part to satisfy any degree requirement at this or any other university.

Abstract

Icelandic seismic records continue to grow as numerous seismographs record data in real time. A manual pick is a specific timestamp when an analyst thinks an earthquake arrived first at the station based on the station's data. Picking accurately when primary and secondary waves (P- and S-phases) arrive at seismic stations from earthquakes/events is used to estimate earthquake origins, which builds a geological, tectonic and magmatic map under the surface that can vary over time. In the past, phase picking has been a human task. However, with recent work using large, human-analysed datasets, machine learning techniques like PhaseNet, EQTransformer and DeepPhasePick have been created that boast precision and recall comparable to humans from data in North America, central Europe, Asia and South America. I applied PhaseNet, DeepPhasePick and EQ-Transformer to manually picked seismic records in and around Askja, Iceland by various analysts. I found DeepPhasePick to have the lowest scoring with a recall below 15%, which I attribute primarily to being trained and optimised to only 30,000 waveforms. PhaseNet performed better than EQTransformer, with a 24.7%/5.9% improvement in precision/recall for P-phase pickings and a 23.4%/19.1% improvement for S-phase picking. I also trained PhaseNet initially with shallow ($\leq$ 8km depths) Askja events. Training PhaseNet from random weights on 18,158 waveforms led to a 12%/36.9% increase in precision/recall for P-phases and a 2.7%/26% increase for S-phases. PhaseNet training also shows improvements on deep ($>$ 8km depths) earthquake events of more complex earthquake origin since they are part of Iceland's ductile crust. The best PhaseNet model has a precision/recall of 95.8%/94.5% for the P-phases and 82.4%/80.9% for S-phases in shallow earthquakes. QuakeMigrate uses information gathered from the seismograph's Short Term Average to Long Term Average ratio (STA/LTA) over time to pick phase arrivals and locate earthquake origins. The most notable difference in picking is that QuakeMigrate arrival times have a 42.6% less precision and a 43.9% less recall than PhaseNet for shallow event P-phases due to QuakeMigrate's earlier picking times. I tried using

PhaseNet's continuous phase probabilities to replace an STA/LTA method for locating events in QuakeMigrate. When comparing the two methods south of Askja, it was found that the location uncertainties are smaller when using STA/LTA. Therefore, I suggest that PhaseNet's picking times are instead applied into phase association software followed by earthquake locating software, like Non-Linear Location (NonLinLoc), for automatic processing of seismic data because PhaseNet is designed for picking time accuracy and does not contain explicit information on picking confidence. The Askja-trained PhaseNet model also has potential uses in calculating phase polarisations and building velocity models to better understand Iceland deformation.

1 Introduction

Iceland is a unique, ever-changing region of geology on top of the North Atlantic ridge (figure 1). With the powerful 2010 eruptions of Eyjafjallajökull, south Iceland, grounding 107,000 flights over eight days, it is clear why there is great interest in understanding and predicting Iceland's seismic activity from an economic and safety standpoint [1].

The Northern Volcanic Zone (NVZ) is the surface response to the spreading rift between the Northern American and Eurasian tectonic plates. Volcanic systems are formed by rising, low-viscosity, basaltic magma filling extra space. Askja volcano, situated in the NVZ, is composed of three calderas - bowl-shaped structures formed by the evacuation and collapse of magma storage regions during volcanic eruptions - the largest of which ($\sim$ 8km in diameter) formed roughly 11,000 years ago. Askja last erupted in 1961, forming an 800m long fissure (a vent where magma can rise to the surface). Since the 1961 eruption, the long-term deformation trend has been an exponentially decaying rate [2]. However, since August 2021, the trend has reversed and Askja has been rapidly re-inflating.

Askja is microseismically active even when rifting is not occurring. It is thought to be linked to deep crustal melt and hydrothermal activity in Askja's volcano [3]. The region has numerous surrounding seismic stations (figure 1) with data covering more than 20 years. Therefore, Askja is a prime benchmark for machine learning models to detect, pick and locate exotic seismic events in Iceland. Energy released by seismic events is carried away from the point of origin (the earthquake's 'hypocentre') in the form of seismic waves. There are two broad classes of seismic body wave: the primary, or P-phase, and the secondary, or S-phase, which exhibit longitudinal and transverse (shear) particle motions, respectively. The arrival of these phases typically appear as sharp changes in the amplitude in a seismograph recording. It is the timing of this first motion that analysts record as the phase arrival or 'pick'. More complex waves can also be observed, such as those from reflections at rock boundaries, taking other paths to seismic stations. But, these are attenuated because they travel further and are harder to discern from the other phase signals. The majority of Askja's earthquakes have a local magnitude scale ranging between -0.5 and 3, the majority of which are below the threshold to which humans are sensitive. Seismographs also record various levels of noise, depending on their location. A primary source of noise in the area is called "secondary microseisms", which are generated by the coupling of energy/waves produced by wind-driven ocean storms with the solid Earth. These are dominant at 0.1 to 0.3Hz. Noise can also be a 0.4 to 0.5Hz low frequency signal, most notably caused by wind, oceans and human activity.

As seismic waves travel through the solid Earth, information about the geology and material properties of the Icelandic crust are encoded into the waveforms. They act as probes inside the Earth. With the growth of seismic networks in Iceland, and the consequent growth of digital archives of raw waveform data, building a reliable and automatic way to process seismic data with analyst-level accuracy is desirable. QuakeMigrate, created in 2020, is automatic seismic data-processing software that uses a waveform migration and a stacking algorithm to find corresponding events across seismic stations [4, 5]. It produces earthquake catalogues from raw seismic data, including their locations, origin times, P- and S-phase arrival times and local magnitudes with estimated uncertainties. QuakeMigrate gathers information about the P- and S-phase arrivals by using the Short Term Average to Long Term Average ratio (STA/LTA) algorithm to produce a continuous function over time that can be aligned and stacked (for example, see figure 13). There are, however, limitations to the STA/LTA approach, including: 1) STA/LTA peaks earlier than the actual phase arrivals because of the finite window size used to average. 2) Typically, the S-phase STA/LTA (made up of equal weightings from the N and E seismic directions) still contains information on the P-phase arrival. 3) STA/LTA is dependent on the Signal-to-Noise Ratio (SNR) of the data, creating a bias towards high SNR signals. The SNR quantitatively describes how prominent an earthquake signal is in a seismograph. With machine learning, learnt complex pattern recognition may lead to greater phase pick precision and a reduction in SNR bias.

Recently, machine learning has been applied to more seismological data around the world as data becomes available and vast, in the hope of developing improved automatic seismic interpretations. PhaseNet, published in 2018, took similar techniques used in image recognition and applied them to one-dimensional seismic data to build a continuous P- and S-phase probability (for example, figure 5) [6]. The model was built to pick phase arrivals during earthquake events in Southern California and reached new picking accuracy and recall heights beyond other, non-machine learning methods. EarthQuake-Transformer (EQT), published in 2020, takes a new, global approach, by training with a larger data size from around the world. EQT has more trainable parameters and uses a different model architecture [7]. The model first detects an event in the data as a window, is then used as a guide for making phase picks. This is achieved by producing the same continuous phase probabilities as PhaseNet. DeepPhasePick (DPP), published in 2021, optimised the model architecture for P- and S-phase picking separately using data from northern Chile [8]. Like EQT, DPP will detect then phase pick. They also introduce a quantitative measure of the picking error by using the Monte Carlo dropout method [9], where the picking stage is repeated with random neurons/nodes disconnected each time to mimic Bayesian inference.

EQT and PhaseNet have been applied to other regions [10, 11]. There has also been work done to retrain a model starting from its original using a small sample size [12]. With past advances in mind, testing how machine learning applies to Iceland's seismicity could offload labour-intensive tasks with greater reliability and precision than pre-existing algorithms. In this project, I wanted to push the boundary of machine learning work applied to Iceland by investigating some key questions:

- Are machine learning models applicable to pick microseismic phase arrival times in Iceland?
- Does training on Iceland data beforehand improve a model's phase picking ability in the region?

- How does transfer learning (training from a pre-trained model) compare to training from initially random weights?
- Can machine learning models go further by using the continuous probabilities to locate earthquakes in QuakeMigrate as a replacement for STA/LTA?

The data used, the phase picking and earthquake locating methods are described in section 2. In section 3 I show the results found, in section 4 I discuss findings and in section 5 I draw conclusions.

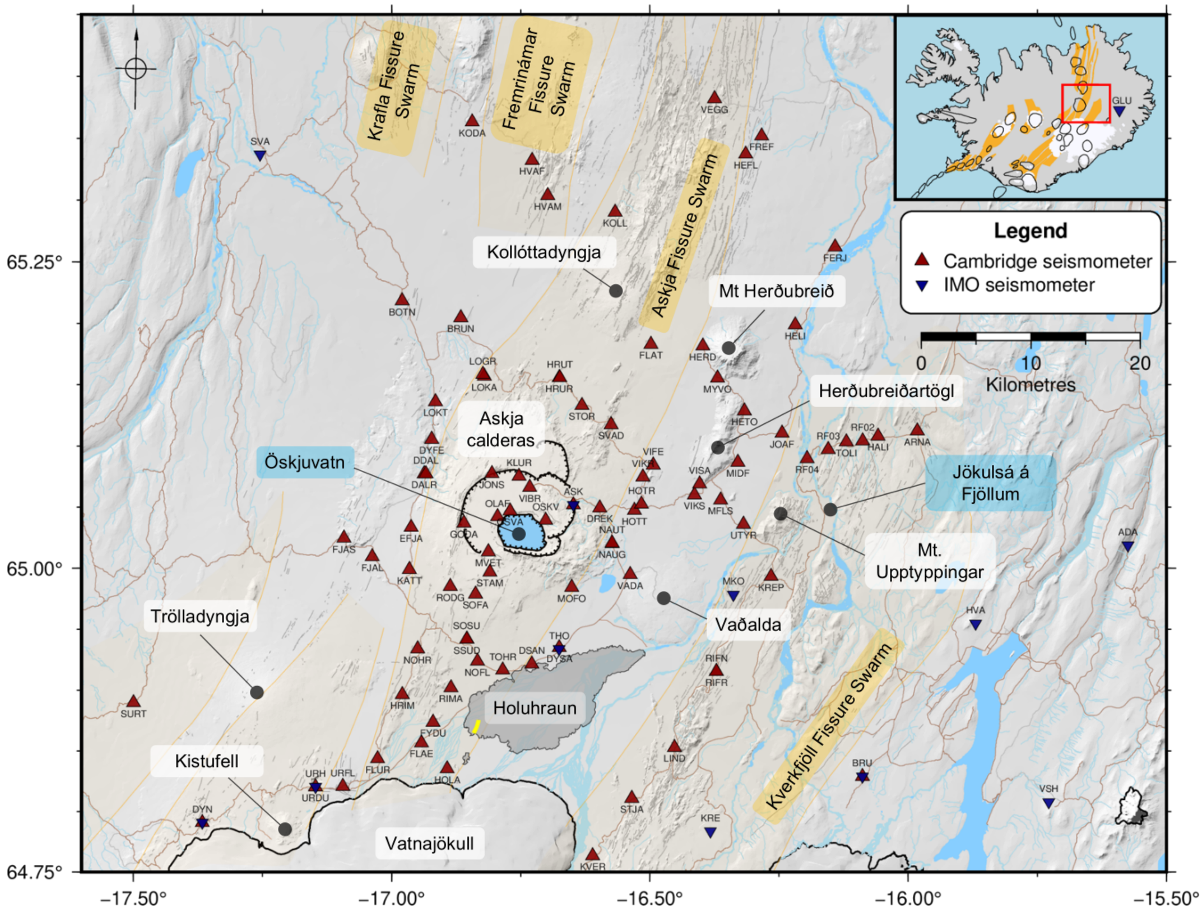

Figure 1: Simplified tectonic and network map around Askja, Iceland. IMO are Icelandic Meteorological Office seismic stations. Light brown lines are roads, blue represents rivers and lakes. Yellow shaded areas represent fissure swarms, volcanoes are outlined in black and calderas are shown by ticked, black lines [3].

2 Method

2.1 Data

The seismic data was collected by the Cambridge Volcano Seismology Group (CVSG) and consists of raw waveforms in and around Askja recorded across the seismic network. Parts of the waveforms, between 2007 and 2015, have been picked for P- and S-phase arrival times by analysts manually. The majority of seismic events originate between the most recently formed caldera, Öskjuvatn and Mt. Herðubreið in the north (figure 2). The seismic events used contained a P-phase followed by an S-phase no more than 5s apart. The waveforms are continuous, three-component (North (N), East (E) and vertical (Z) motion) velocity seismographs resampled from 50Hz to 100Hz, if necessary, using the Lanczos method, which were the inputs for machine learning models [13]. There are 94 unique seismic stations, with a range of Signal-to-Noise Ratios (SNRs) and earthquake local magnitudes typical to Iceland for both testing and training a model on (figure 3).

Grades were assigned to each phase arrival pick by the analysts, with each grade mapping to a picking error (measured in samples then translated to time). Manual picks with an error > 0.1s were removed (majority of the errors left are 0.01s and 0.02s) because a model's picking time is said to be true positive if its distance away from a manual pick (the residual time) is no further than ± 0.1s (section 2.2 details further). CVSG, using the software Non-Linear Location (NonLinLoc), estimated each event's location and hence its own estimated phase arrival times at each seismograph by calculating P- and S-phase travel times [14]. If the manual pick time is more than 0.4s away from the NonLinLoc's estimated arrival time, then the waveform was excluded since it is likely that the manual pick was inaccurate. The events are separated between deep ($>$ 8km depth) and shallow ($\leq$ 8km depth) earthquakes because seismic events happening below 8km are in a ductile part of Iceland's crust, thought to be caused by fluid migration or melt to generate the high strain rates necessary for fracture. In addition, deep events occur in short bursts of time at high frequency, a so-called swarm, making the individuals phase arrival times harder to untangle, so making a distinction is reasonable [15]. The catalogue of shallow events was taken from Winder (2021), which allows for QuakeMigrate phase picking results for an additional comparison (section 2.2.4) and local magnitude estimations for each event.

The full dataset is comprised of 22,695 waveform snippets, split into training (18,158, 80%), validation (2,264, 10%) and test (2,273, 10%) sets via stratified sampling, as well as 13,214 deep test samples.

2.2 Phase picking

PhaseNet, EQT, DPP and QuakeMigrate were all tested on the same 2,273 shallow events and every PhaseNet model was tested on 13,214 deep events. Each sample was input as a 120s window with the P-phase arrival centred to give more than enough padding. A model's phase pick is considered to be true positive when it is no further than 0.1s away from the manual pick time. A false positive is counted for a pick further than 0.1s but no further than 4s and false negative if no pick is made within the true positive window. Picks are only counted between ± 4s because there can be more than one seismic event in a 120s window (see figure 4). The metrics adopted to measure the performance of each model are precision, recall and F1 score as used by the creators of PhaseNet and EQTransformer. Precision, recall and F1 score are defined below:

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{1}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \tag{2}$$

$$\text{F1 score} = 2 \times \frac{\text{precision} \times \text{recall}}{(\text{precision} + \text{recall})}, \tag{3}$$

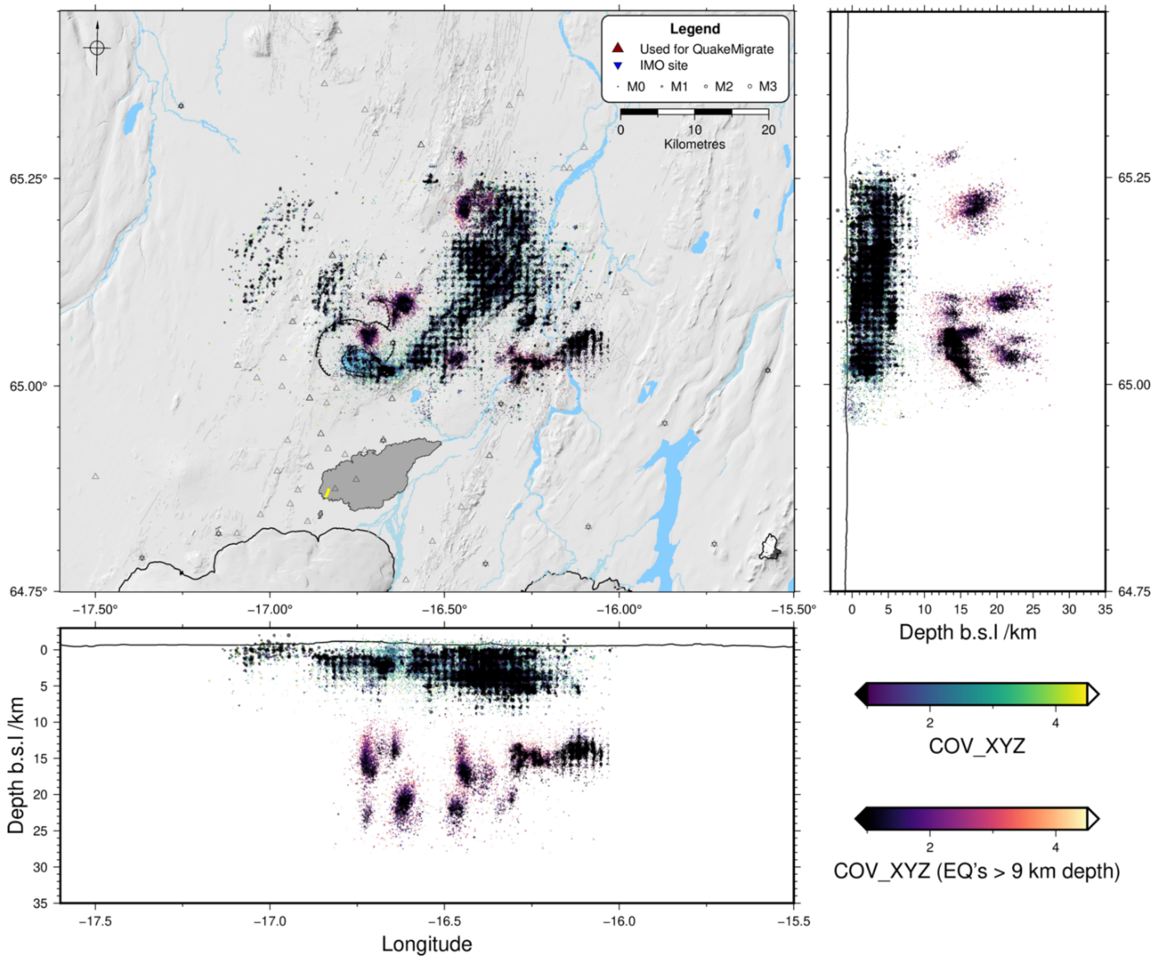

Figure 2: Askja event catalogue, 2007-2020. The plot shows 155,282 events retained after filtering QuakeMigrate's preliminary locations. The events are scaled by magnitude and coloured based on global covariance (a measure of location uncertainty) with deep and shallow events coloured differently [3].

such that each metric varies between 0% and 100%, where TP, FP and FN are the true positive, false positive and false negative counts respectively. F1 score combines the precision and recall scores for an overall model score. Maximising the F1 score guides towards precision and recalls of equal, highest values. To quantify the spread and precision of picking times, the mean, μ, and the standard deviation, σ, were calculated using all picks within ± 0.5s from the manual pick.

2.2.1 PhaseNet

PhaseNet is a deep, convolutional neural network that aims to accurately pick P- and S-phase arrival times from unfiltered, continuous seismic data. U-Net is a machine learning architecture used for image recognition [16]. PhaseNet is a modified U-Net style architecture to better suit seismic data. PhaseNet still has a "U" shaped network architecture using four convolution steps followed by four deconvolutions to return back to the same dimension as the input. PhaseNet has 268,000 parameters and was initially trained on roughly seven million Northern California Earthquake Data Center (NCEDC) waveforms. PhaseNet is input a 30s three-component window of seismic data normalised

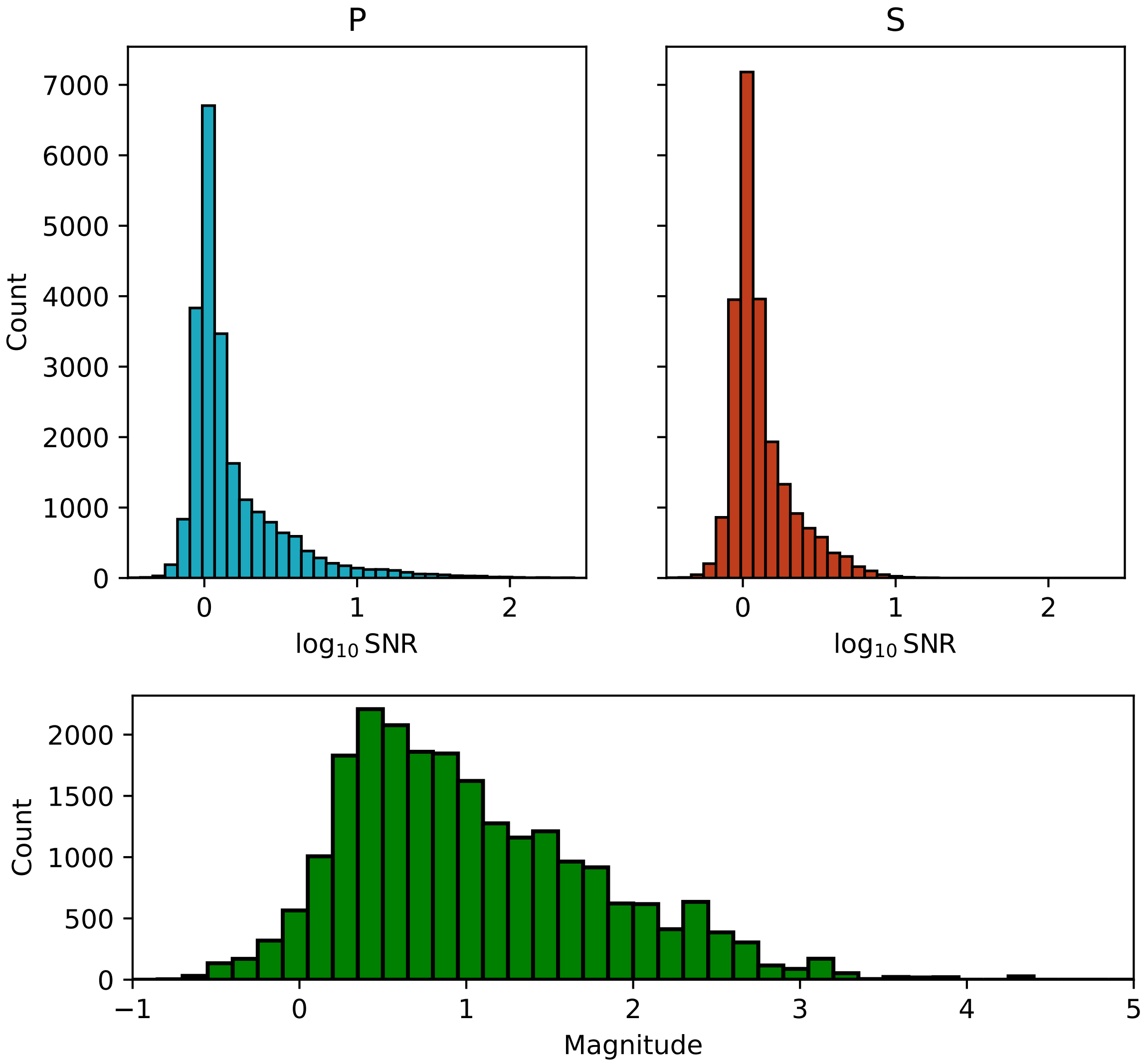

Figure 3: Askja's shallow dataset properties. The top row shows the Signal-to-Noise Ratio (SNR) distribution for P- and S-phases in the unfiltered shallow dataset, the bottom row shows the distribution of earthquake local magnitudes. The SNR is calculated as the ratio of the standard deviation of a 5s window proceeding to the standard deviation of a 5s window preceding the manual pick.

by its standard deviation and outputs a 30s continuous probability for P-, S- and noise over time, $\hat{P}(t)$, $\hat{S}(t)$ and $\hat{N}(t)$ respectively. $\hat{P}(t) + \hat{S}(t) + \hat{N}(t) = 1$ for all times, t, by using a cross-entropy loss function. A PhaseNet pick time is given by a probability local maximum when above a set probability threshold. If there are peaks within 0.5s of each other, then shorter peaks are ignored until peaks are further than 0.5s. The probability thresholds for P- and S-phases are chosen separately by testing values between $[0.1, 0.9]$ in 0.1 increments and choosing where the F1 score is maximum. If the best performing probability threshold is on an edge case (0.1 or 0.9), then increments of 0.01 were looked at.

PhaseNet was retrained starting from its model trained on Northern California data, a process known as transfer learning or retraining. This could potentially improve the robustness of the model and allow training with few samples. The waveforms were not

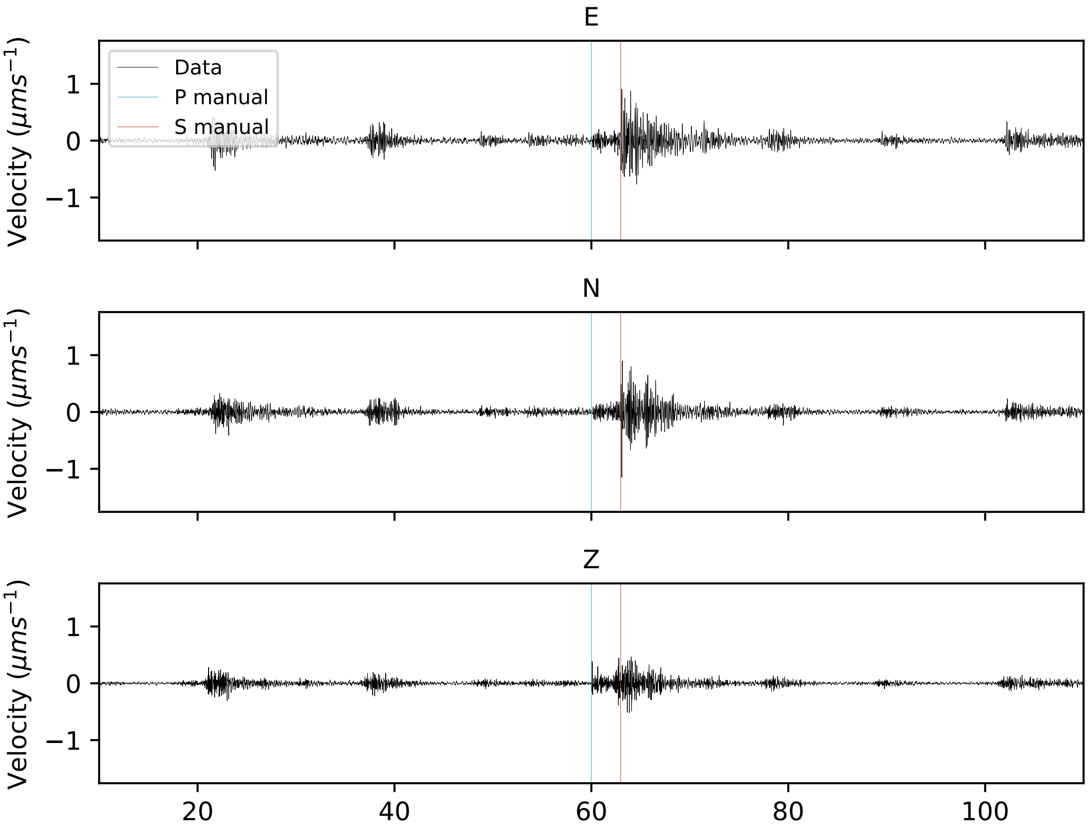

Figure 4: Numerous seismic events in one waveform. A sample taken from the deep earthquakes dataset, where the vertical lines represent manually picked phase arrivals, vertical solid lines represent the manual picking times. The black lines are the Earth's velocity recorded by the seismograph data bandpass filtered between 2Hz and 20Hz for clarity.

bandpass filtered to allow PhaseNet to distinguish noise. Each manual pick is represented by a truncated Gaussian of width 0.3s, so that the local maximum represents the arrival time (see figure 5). The 30s window of data is uniformly, randomly shifted such that the truncated Gaussian is still within the window's boundaries. This is to avoid PhaseNet learning a false positional dependence within the waveform window. The learning rate was set to a relatively small, constant value of 10^{-5}. The small learning rate stops the model from completely forgetting Californian training, known as "catastrophic forgetting" [17].

PhaseNet was also trained from uniform, random parameter weights between $[-L, L]$, where $L = \sqrt{1/3001}$ to test the effectiveness of two different learning methods. This is referred to as PhaseNet trained from scratch to distinguish the learning methods. The learning rate was set to 0.1 and decayed by a factor of 10 every 50 epochs, where an epoch is one pass through the entire training dataset in a random order. The initially high learning rate suppresses memorisation of noisy data, while the learning rate decay improves chances to learn complex patterns in the data [18].

In both training cases, the batch size is set to 20 and the trained model with the lowest validation loss was used, since the validation dataset was tested against but not trained on. This is to avoid the model memorising data, also called overfitting.

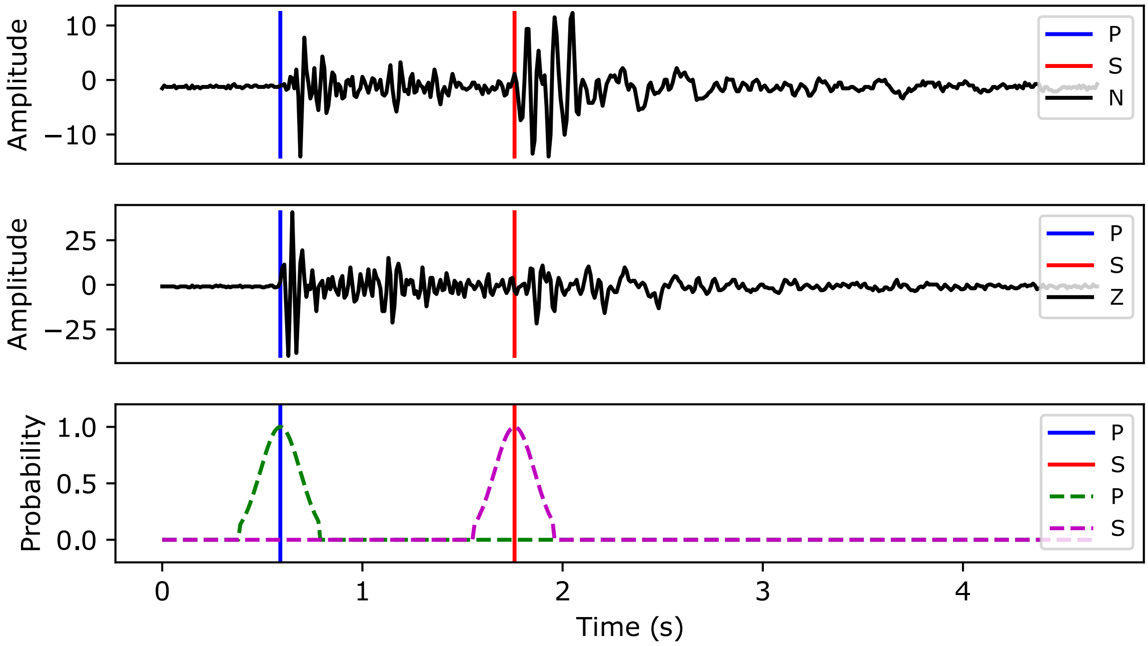

Figure 5: Example manual pick. How manual picking times are converted into a target probability over time (bottom row) for PhaseNet training [6].

2.2.2 EarthQuakeTransformer

EarthQuakeTransformer (EQT) is an artificial intelligence-based event detector and phase picker that runs on 60s seismic data slices. The model uses transformers and global/local-attention mechanisms to focus in on a smaller window where an event occurs, which the phase picks are made on. EQT is more complex than PhaseNet, with more layer types as well as a 41% greater parameter count of 372,000. The parameters were trained using one million event waveforms and 200,000 noise-only waveforms. The data is sourced globally, including a significant number of seismographs in North America, central Europe and South Asia. Like PhaseNet, no region of Iceland was trained on originally.

EQT was tested on Icelandic data the same way as PhaseNet, except for bandpass filtering the data between 1Hz and 45Hz as done in the original paper.

2.2.3 DeepPhasePick

DeepPhasePick (DPP) is a convolution neural network with two recurrent neural networks for detecting and phase picking respectively. Hyperparameters, parameters that are set before the model's training, are varied and chosen to optimise the model's performance in Northern Chile. DPP's original model was applied to shallow test waveforms, but recall was always below 15% even after attempting various filterings, including bandpass, highpass and lowpass. With less complete documentation than EQT and PhaseNet, it was decided to continue work only on EQT and PhaseNet.

2.2.4 QuakeMigrate

QuakeMigrate implements an exhaustive scan through an archive of raw data for earthquake origin times and hypocentres. This is achieved by migrating the STA/LTA onset functions back into the search volume using an assumed or predetermined velocity structure. At every node in the search volume, these migrated onset functions are stacked. Coherent peaks in the stacked onset functions localise the origin of a seismic event in both space and time. A Gaussian is fitted to the P- and S-phases in time to get a picking error uncertainty.

2.3 Locating

QuakeMigrate uses an onset function for P- and/or S-phases to represent continuous information on seismic events. The onset functions are migrated onto a three-dimensional grid of nodes, where each point is a potential hypocentre. The onset functions are aligned according to a pre-determined travel time table to each seismic station and repeated for a range of origin times, building a four-dimensional phase space (or "coalescence function"). The base onset function provided by QuakeMigrate is based on the STA/LTA algorithm, but the software has been designed as a framework for the broad class of partial waveform information migration and stacking methods. It is straightforward to incorporate alternative onset functions, provided they are stable and designed to produce peaks around the onset of a seismic phase arrival.

Locating was applied to August 2014 volcano-tectonic events in Skútustaðahreppur, $\sim$ 30km south of Askja. The events are included in QuakeMigrate's source code examples. They were located using the standard STA/LTA onset function and PhaseNet's continuous probabilities $\hat{P}(t)/\hat{N}(t) + 1$ and $\hat{S}(t)/\hat{N}(t) + 1$ for P- and S-phases respectively. The +1 is added because PhaseNet's onset functions will constructively combine for higher phase probabilities, as the coalescence is calculated based on the geometric mean (a product of n values followed by nth rooting). Adding values below +1 is a way of introducing destructive combinations of onset functions, potentially sharpening the earthquake locating process as a post-processing optimisation. The node spacing was changed from its default of 0.5 to 0.2km in all directions because PhaseNet's onset function can be narrower than STA/LTA, potentially placing the earthquake's location in between grid points.

3 Results

3.1 Training

Figure 6 shows the loss functions against epoch for both retraining and training PhaseNet from scratch respectively. The training was performed on the High Performance Computing (HPC) in Cambridge using one NVIDIA A100-SXM-80GB GPU, taking $\sim$ 60 hours for retraining and $\sim$ 20 hours for training from scratch.

3.2 Phase picking

The results for phase picking shallow events are tabulated in table 1, showing that the greatest F1 score is PhaseNet trained from scratch. Figure 7 shows shallow waveform examples with "bad" PhaseNet original predictions and how they change after training.

Varying the phase picking probability threshold will affect the precision and recall of a model as it acts similar to a confidence level. Figure 8 shows the precision and recall variance of PhaseNet before and after training to Iceland data. Figure 9 shows how

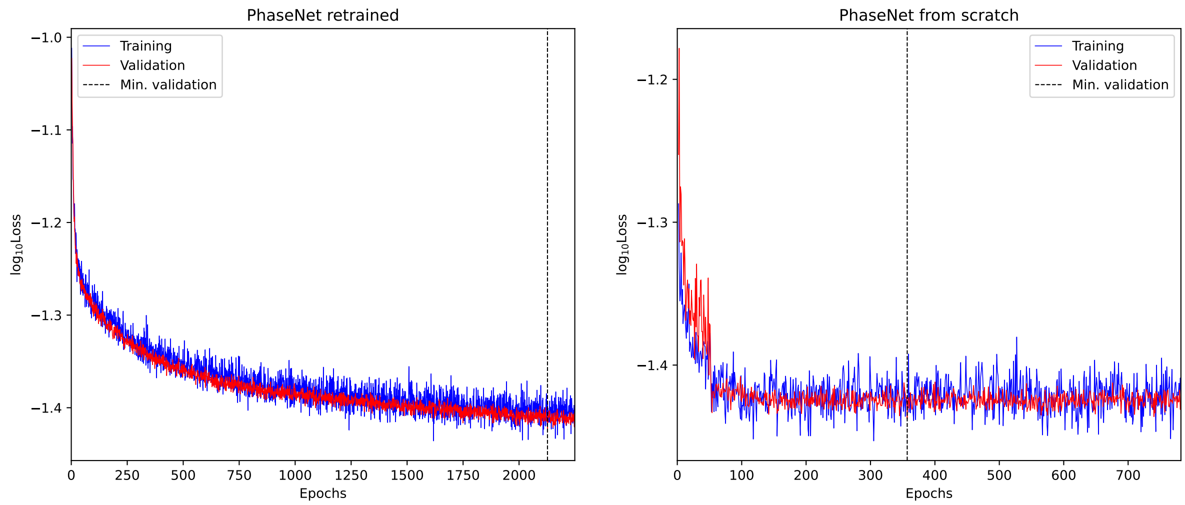

Figure 6: PhaseNet losses for retraining (left) and training from scratch (right). The dashed black line labels the model with the lowest validation loss found for each case, used in picking predictions.

SNR affects the S-phase significantly, with a decrease in pick scoring at high SNR values. QuakeMigrate often picks earlier than the analysts' picking times, shown in figure 10.

The results for applying machine learning models to the deep Askja events are shown in table 2, showing similar improvement for both PhaseNet retrained and trained from scratch. Figure 11 shows the picking time variation for PhaseNet before and after training from scratch. There is a minimal correlation between the local magnitude and the picking time difference and training causes a rise in picks made in the true positive window (± 0.1s). Figure 12 displays phase picking improvement after training PhaseNet from scratch on a deep event waveform as other P- and S-phases are being recognised by PhaseNet within the 120s window.

3.3 Locating

To illustrate how STA/LTA and PhaseNet's probability widths and timings compare, figure 13 shows the two onset functions on top of one another for an event in August 2014 at two different stations, showing how timings can vary and PhaseNet's onset can be narrower than STA/LTA. Figure 14 shows an event located using STA/LTA and figure 15 shows the same event located using PhaseNet retrained. PhaseNet locating has a greater uncertainty, but almost identical mean location.

4 Discussion

On shallow P-phase pick times, QuakeMigrate has the lowest overall F1 score because about half of the picking times are more than 0.1s earlier than the manual pick time, increasing both the false positive and false negative count. This is likely due to P-phases greater SNR compared to the S-phase, which is made more prominent when most noise is filtered out by bandpass filtering between 2 and 20Hz for STA/LTA. QuakeMigrate's S-phase picking is comparable to the PhaseNet retrained models.

Model	Prob. threshold	Precision	Recall	F1 score	μ(ms)	σ(ms)
PN original	0.50	0.838	0.576	0.683	-50	68
PN retrain	0.40	0.915	0.892	0.904	-46	43
PN scratch	0.40	0.958	0.945	0.952	-36	39
EQT original	0.06	0.592	0.517	0.552	-100	71
QM (1)	–	0.532	0.506	0.519	-86	63
QM (2)	–	0.980	0.931	0.955	-86	63

PN original	0.10	0.797	0.549	0.662	-61	85
PN retrain	0.30	0.830	0.771	0.799	-53	71
PN scratch	0.30	0.824	0.809	0.816	-52	68
EQT original	0.30	0.563	0.358	0.438	-6	86
QM (1)	–	0.822	0.782	0.801	-40	76
QM (2)	–	0.965	0.918	0.941	-40	76

Table 1: Shallow event test results for P- (top) and S-phase (bottom). QuakeMigrate (QM) has two scorings, one with the same true positive window of $\pm$0.1s (1) and the other with a true positive window of $\pm$0.2s (2).

Model	Prob. threshold	Precision	Recall	F1 score	μ(ms)	σ(ms)
PN original	0.05	0.706	0.405	0.515	-4	109
PN retrain	0.20	0.874	0.680	0.765	3	72
PN scratch	0.20	0.872	0.690	0.770	15	69

PN original	0.03	0.706	0.400	0.512	18	132
PN retrain	0.20	0.736	0.557	0.634	23	115
PN scratch	0.20	0.724	0.559	0.631	31	123

Table 2: Deep event test results for P- (top) and S-phase (bottom).

EQT original performed the worst on shallow events. This could be an indication that EQT's complex architecture is not as suitable to the Iceland region because the model may have overfitted to patterns in data from other regions, like North America, central Europe and Asia, while the simplicity of PhaseNet's design keeps the model robust. Therefore, PhaseNet can perform better in Iceland, even with training only from Northern California. However, EQT may still be a useful tool for Iceland phase picking after retraining or even training from scratch by future work. EQT tries to separate each seismic event into windows, this is especially problematic for close together swarm seismic events in deep earthquake waveforms.

PhaseNet's training shows significant improvements in phase picking for both shallow and deep events. On shallow events, PhaseNet trained from scratch performs better than retraining, which I attribute to an initially higher learning rate, allowing the model to fit to Askja data better. The 18,158 waveforms used to train PhaseNet from scratch is small in comparison to the seven million waveforms used in Northern California, but the good performance suggests that a small, high quality sample size can go far in training a model for a specific region. Moreover, if we assume each seismic event lasts for $\sim$ 5s in

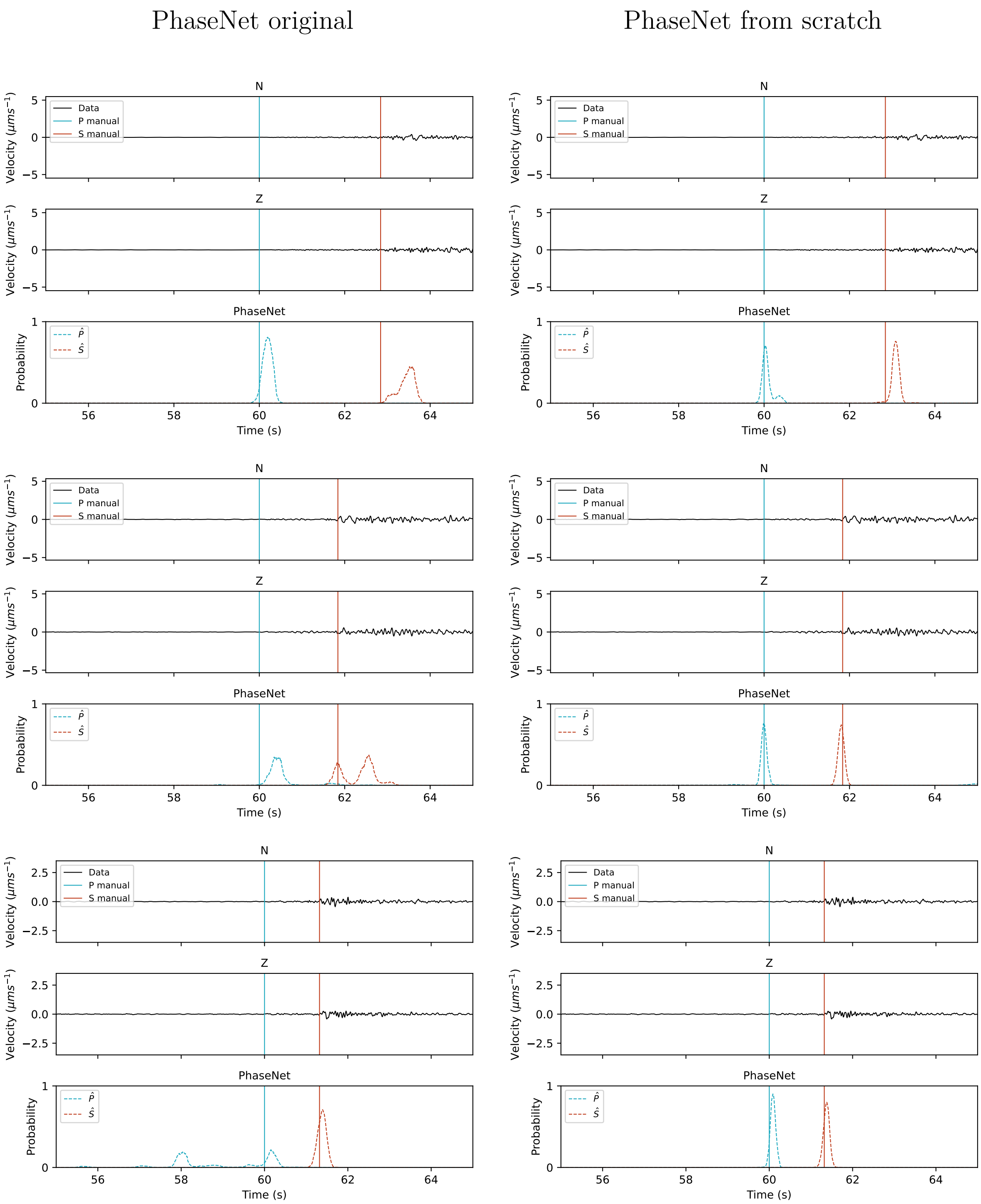

Figure 7: PhaseNet shallow waveform examples. A part of the test dataset. Each waveform is predicted before (left column) and after training from scratch (right column).

the 30s training waveforms, this gives 18158 $\times$ 5s$\times$100Hz = 1,089,480 data points which are not noise. This is a comparable number to PhaseNet's 268,000 trainable parameters, which explains why PhaseNet, with an effective choice of learning rate, can still train from random weights with fewer waveforms. Also, U-Net's architecture has been specifically designed for training with fewer samples by convolution steps, which pick out important patterns in data using few parameters. The remaining $\sim$ 25s from each waveform gives

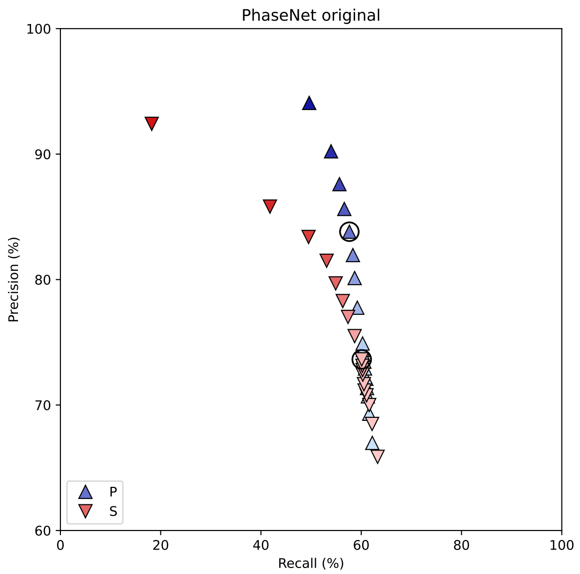

Figure 8: Precision-recall plot for PhaseNet original (left) and PhaseNet trained from scratch (right) tested on shallow events. A darker shade indicates a higher probability threshold. The circled points have the greatest F1 scores.

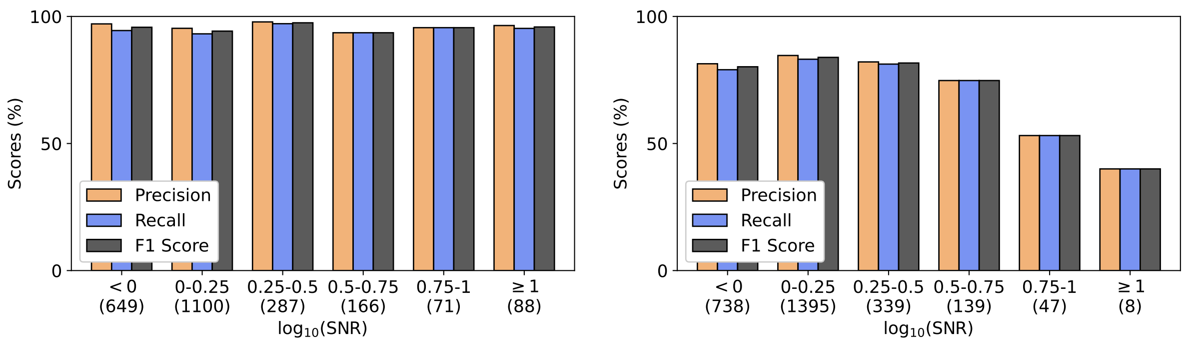

Figure 9: PhaseNet from scratch scoring against Signal-to-Noise Ratio (SNR) on shallow events. P-phase on left and S-phase on the right. The numbers in brackets shows the total count (TP + FP + FN) in the SNR range.

ample opportunity to distinguish noise from events.

It was found that P-phase picking had a low dependence on the SNR, whereas the S-phase had a stronger dependence. I believe that the SNR dependence is due to two effects: a smaller training sample size for large SNRs and large SNR values making the phase arrival easier to pick. S-phase is complicated further by the P-phase superimposed on top of it, making every S-phase arrival more unique in comparison to P-phase arrivals.

Models perform worse on the events deeper than 8km. Interestingly, both trained PhaseNet models perform the same within 1% on the deep events. The results indicate that deeper earthquake swarms are harder for algorithms to discern and precisely pick compared to the shallow events, but training on shallow events still improved their picking scores. A better deep event picking performance could be tackled with further work to try and train on deep events. However, the limitation with the the manual picks used in this paper is that they are sporadically picked arrivals throughout years of seismic data,

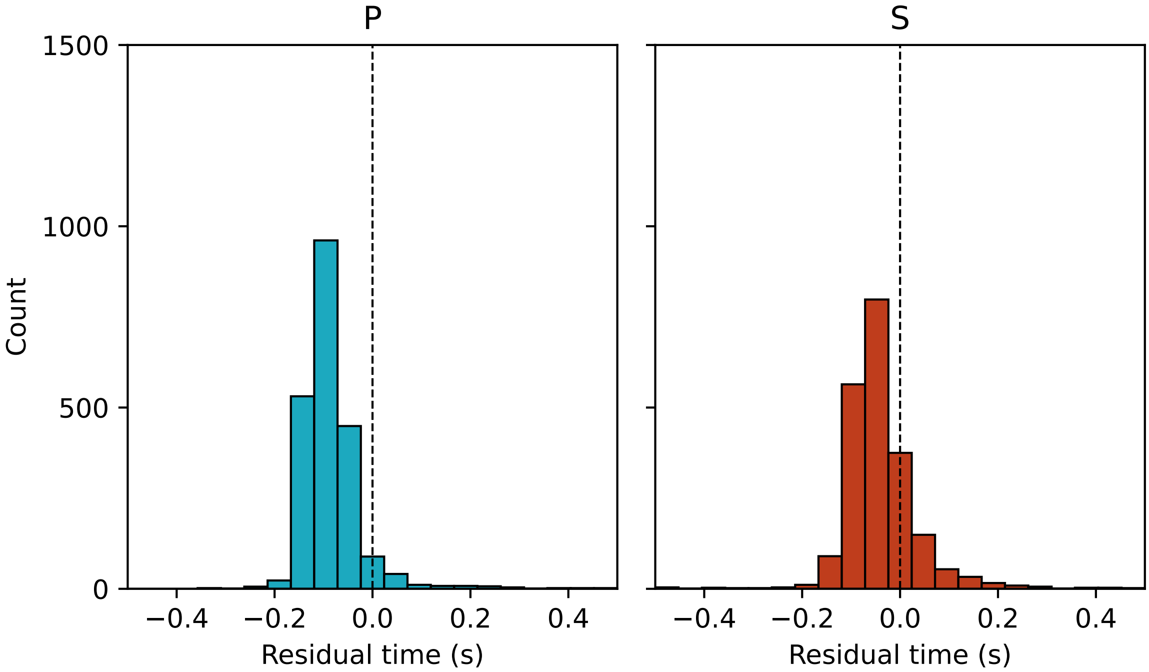

Figure 10: QuakeMigrate's (QM) residual times (QM pick time−manual pick time) between ±0.5s on shallow events. Negative residual times represent an earlier picking than the manual pick time.

and since the training window size is 30s, multiple phase arrivals can occur within each window which have not been manually labelled. The same argument applies to shallow seismic events too, but the events tend to be less frequent, i.e. more spaced out, so the training data was a higher quality. To tackle this problem, I would suggest manually picking every phase arrival within a short period of time, say one hour, at multiple seismic stations and do this for various years of data. This gives a higher quality dataset because no phase arrivals, to the best of the analysts' abilities, are mistakenly labelled as noise during training.

A limitation of PhaseNet is it does not quantify the picking error. One way to do this could be applying DeepPhasePick's Monte Carlo dropout technique, which may be tuned to produce similar errors to the analysts. Since a pick's probability over time is supposed to be a Gaussian, another way of quantifying error could be to calculate the difference between the pick probability over time versus the 0.3s truncated Gaussian that PhaseNet trains to try output. Exactly how the difference maps to a picking error would be something to explore further.

The volcano-tectonic earthquakes used in this study are typically modelled as point source double-couple failures. The mode of failure of the originating fault, and its orientation in space, can be determined from the spatial pattern of first arrival polarity data (whether the first motion was 'up' or 'down'). Making these observations, however, is an equally time-intensive task like phase picking. With high precision phase pickings from a machine learning model, data just beyond the picking time could be looked at to decide the direction that the seismograph moves.

PhaseNet demonstrated the ability to locate events using QuakeMigrate software with almost identical locations to STA/LTA. However, STA/LTA as an onset function has a

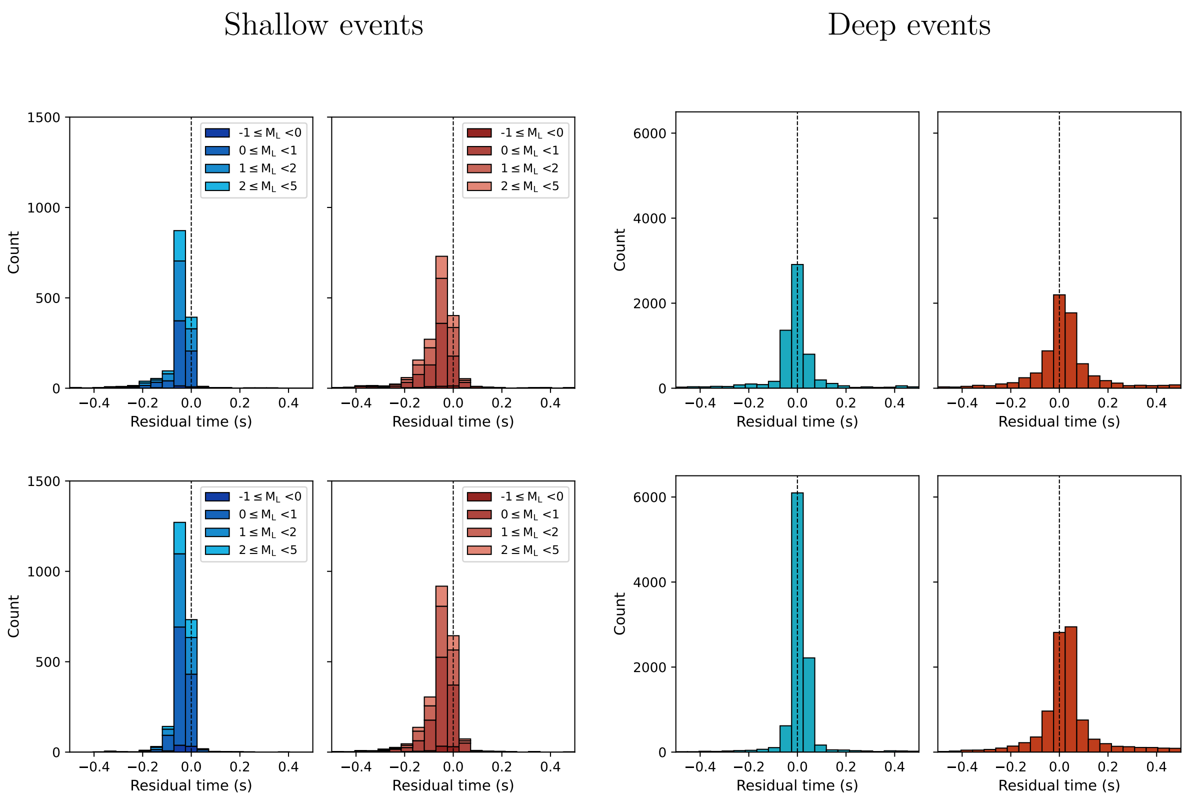

Figure 11: PhaseNet residual time histograms. Top row shows PhaseNet original, bottom row is PhaseNet trained from scratch. The left column shows shallow events, right column shows deep events. The residual time is (PhaseNet pick time − manual pick time) such that a positive residual time represents PhaseNet picking later than the analyst. Before training, $(73 \pm 6)\%$ of shallow test samples are shown and $(52 \pm 5)\%$ of deep test samples are shown, the remaining samples were not picked between ± 0.5s. After training, $(97.6 \pm 0.3)\%$ of shallow test samples are shown, and $(73.21 \pm 0.11)\%$ deep test samples are shown.

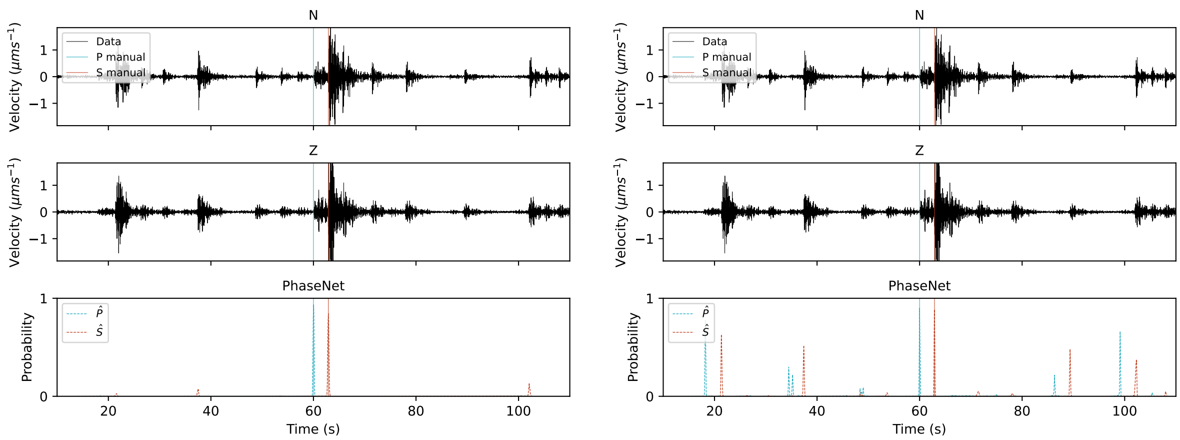

Figure 12: PhaseNet deep waveform example. A part of the test dataset, left is PhaseNet original's prediction, right is PhaseNet trained from scratch.

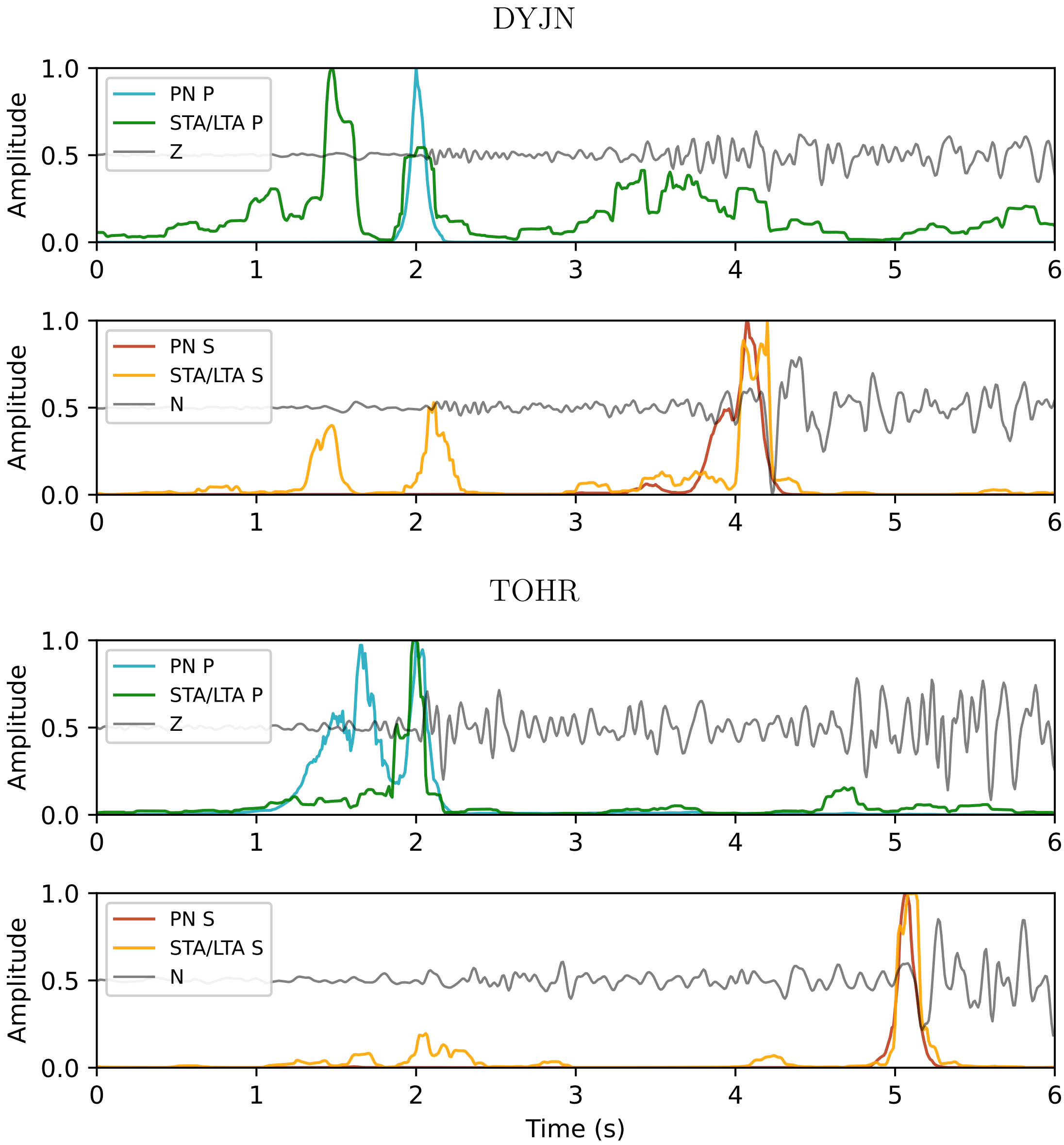

Figure 13: STA/LTA and PhaseNet trained from scratch comparison. One seismic event on 24 August 2014 at DYJN station (top) and TOHR station (bottom). Zero samples from DYJN station were included in the training dataset, 153 TOHR station samples were included. STA/LTA and PhaseNet (PN) probabilities are normalised between 0 and 1. The seismic data for the N and Z component are 2Hz to 20Hz bandpassed, normalised between ± 0.5 and centred about 0.5, with their relative sizes preserved. STA/LTA is input 2Hz to 20Hz bandpassed data with average window sizes of 0.2s and 1s.

lower location uncertainty in almost all cases because QuakeMigrate has been originally built around the STA/LTA. For example, STA/LTA has values varying between $[0, \inf)$ compared to PhaseNet's $[1, 2]$ range. PhaseNet has no explicit measure of picking confidence, whereas STA/LTA does have a built-in confidence based on the SNR. PhaseNet's onset function is also narrower than STA/LTA in some cases. Therefore, QuakeMigrate's grid spacing had to be reduced for PhaseNet to reach sharper coalescence maps such that the event origin does not land in between grid points. This costs greater computational time and a requirement for more RAM, which may not be available depending on the hardware being used. To avoid the computational time increase, a future implementation for QuakeMigrate's software would be a boxcar filter convolved with the onset functions to correct for the finite grid spacings by incorporating information about the coalescence values in between nodes. But, since PhaseNet has a high picking precision, in the future I would suggest that PhaseNet's estimated phase arrival times be input into a phase association software (for example, PhaseLink [19]) to combine phase picks thought to belong to the same seismic event, which can then be input into NonLinLoc to locate earthquake hypocentres and their uncertainties to compare against QuakeMigrate's locating for the same period of time.

5 Conclusion

Although Icelandic shallow events can originate from various earthquake mechanisms, machine learning has demonstrated effective phase picking by training on the "gold standard" of picking microseismic events using analysts. With significantly fewer waveforms than PhaseNet's original California training, PhaseNet was trained successfully on Askja events, reaching comparable or better scores than PhaseNet in the original paper for Northern California. Therefore, the evidence suggests that training a model to a particular region can improve phase picking precision and recall, even with limited data. It is found that a model trained with an initially high, decaying learning rate performs better than a model trained with a low learning rate using the Californian-trained model as a base. Retraining/transfer learning from a base model is limited because a low learning rate is required to avoid "catastrophically forgetting" original information and overhauling to the retraining dataset. The results strengthens the idea that 18,158 samples is sufficient to train PhaseNet from scratch in Askja and reach diminishing returns.

I suggest two ways of estimating PhaseNet's picking errors: 1) Applying Monte Carlo dropout, similar to DeepPhasePick, during testing and training. 2) Calculating the difference between the picking probability output and the target Gaussian.

Picking precision can reach over 90% for both phases at a cost to recall. The flexibility of precision-recall dependence may find use in calculating event polarisations to estimate fault angles without human biases and reduce laborious tasks. It also highlights potential uses in estimating event origins and building velocity models under Iceland from phase arrival times.

The work shown here paves the way for machine learning to locate earthquakes in Iceland. PhaseNet successfully locates event origins in QuakeMigrate, but the location uncertainty is greater than the STA/LTA method. Since PhaseNet's P-phase picking time precision for shallow events has a 42.7% increase over QuakeMigrate and PhaseNet's lack of measurable picking confidence, PhaseNet is more suited to associating phase arrival times and then applying to seismic event locating software, like NonLinLoc, which may reduce PhaseNet's perceived high location uncertainties.

Acknowledgements

Thank you to my project supervisor, Nicholas Rawlinson, for advertising the project and giving me prompt help along the way. Thank you Joseph Fone for technical advice on machine learning and seismology and to Tom Winder and Conor Bacon for help with QuakeMigrate.

6 References

[1] Bente Lilja Bye. *Volcanic eruptions: Science and risk management*. Aug. 2014. URL: https://www.science20.com/planetbye/volcanic_eruptions_science_and_risk_management-79456.

[2] Erik Sturkell, Freysteinn Sigmundsson, and Ragnar Slunga. "1983–2003 decaying rate of deflation at Askja caldera: Pressure decrease in an extensive magma plumbing system at a spreading plate boundary". In: *Bulletin of Volcanology volume* 68 (Jan. 2006), pp. 727–735. DOI: 10.1007/s00445-005-0046-1.

[3] Tom Winder. "Tectonic earthquake swarms in the Northern Volcanic Zone, Iceland". PhD thesis. University of Cambridge, May 2021, pp. 37–67. DOI: 10.17863/CAM.82505.

[4] Conor Bacon et al. *QuakeMigrate/QuakeMigrate: QuakeMigrate v1.0.1*. Version v1.0.1. Apr. 2023. DOI: 10.5281/zenodo.7813845.

[5] Tom Winder et al. "QuakeMigrate: a Modular, Open-Source Python Package for Automatic Earthquake Detection and Location". AGU Fall Meeting 2020. 2020.

[6] Weiqiang Zhu and Gregory C Beroza. "PhaseNet: a deep-neural-network-based seismic arrival-time picking method". In: *Geophysical Journal International* 216.1 (Oct. 2018), pp. 261–273. DOI: 10.1093/gji/ggy423.

[7] Seyed M. Mousavi et al. "Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking". In: *Nature Communications* 11.1 (2020), pp. 1–12.

[8] Hugo Soto and Bernd Schurr. "DeepPhasePick: a method for detecting and picking seismic phases from local earthquakes based on highly optimized convolutional and recurrent deep neural networks". In: *Geophysical Journal International* 227.2 (July 2021), pp. 1268–1294. DOI: 10.1093/gji/ggab266.

[9] Yarin Gal and Zoubin Ghahramani. *Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning*. 2016. DOI: 10.48550/arXiv.1506.02142.

[10] Chengping Chai et al. "Using a Deep Neural Network and Transfer Learning to Bridge Scales for Seismic Phase Picking". In: *Geophysical Research Letters* 47.16 (Aug. 2020), e2020GL088651. DOI: https://doi.org/10.1029/2020GL088651.

[11] Ce Jiang et al. "Comparison of the earthquake detection abilities of PhaseNet and EQTransformer with the Yangbi and Maduo earthquakes". In: *Earthquake Science* 34.5 (Oct. 2021), pp. 425–435. DOI: 10.29382/eqs-2021-0038.

[12] Sacha Lapins et al. "A Little Data Goes a Long Way: Automating Seismic Phase Arrival Picking at Nabro Volcano With Transfer Learning". In: *Journal of Geophysical Research: Solid Earth* 126.7 (July 2021). DOI: 10.1029/2021JB021910.

[13] Cornelius Lanczos. *Applied Analysis*. Dover Books on Mathematics. Dover Publications, 1988. ISBN: 9780486656564.

[14] Anthony Lomax et al. "Probabilistic Earthquake Location in 3D and Layered Models". In: *Advances in Seismic Event Location*. Ed. by Clifford H. Thurber and Nitzan Rabinowitz. Dordrecht: Springer Netherlands, 2000, pp. 101–134. DOI: 10.1007/978-94-015-9536-0_5.

[15] Tim Greenfield and Robert S. White. "Building icelandic igneous crust by repeated melt injections". In: *Journal of Geophysical Research: Solid Earth* 120.11 (Nov. 2015), pp. 7771–7788. DOI: https://doi.org/10.1002/2015JB012009.

[16] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Segmentation". In: (2015). Ed. by Nassir Navab et al., pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[17] Michael McCloskey and Neal J. Cohen. "Catastrophic Interference in Connectionist Networks: The Sequential Learning Problem". In: Psychology of Learning and Motivation 24 (1989). Ed. by Gordon H. Bower, pp. 109–165. ISSN: 0079-7421. DOI: https://doi.org/10.1016/S0079-7421(08)60536-8.

[18] Kaichao You et al. "How Does Learning Rate Decay Help Modern Neural Networks?" In: *arXiv* (Sept. 2019). DOI: 10.48550/arXiv.1908.01878.

[19] Zachary E. Ross et al. "PhaseLink: A Deep Learning Approach to Seismic Phase Association". In: *Journal of Geophysical Research: Solid Earth* 124.1 (2019), pp. 856–869. DOI: https://doi.org/10.1029/2018JB016674.

[20] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: https://www.tensorflow.org/.

7 Appendix

7.1 Code availability

7.1.1 Testing and training

Code used is available in the `.zip` file attached. The `phasenet_setup_tom.py` script gathered waveforms and separated them into training, test and validation datasets. To do this, it requires CVSG's manual picks, which can be found in the `AskjaManualPicks` directory. The script `phasenet_eqt_tester.py` was then used to apply a given PhaseNet or EQT model on a created dataset and to produce figures 7 and 12.

To train PhaseNet, the source code script `train.py` was run on the HPC supercomputer. As of 5 May 2023, the `data_reader.py` script at lines `685` to `696` implemented a 5% chance of using the waveform as noise and not labelling the manual picks. This 5% chance was removed during my training. Apart from this change, the training is the same as PhaseNet's original creators. The HPC was sent batch commands to queue and then start PhaseNet training, two example batch files are `phasenet_training_tom.wilkes3` and `phasenet_training_scratch_tom.wilkes3` for retraining and training from scratch respectively.

7.1.2 Locating with PhaseNet

QuakeMigrate's source code is available online. The scripts `phasenet.py` and `__init__.py`, created by Conor Bacon, are placed inside `/quakemigrate/signal/onsets` to allow PhaseNet in the tensorflow framework to be used and applied for earthquake locating [20]. Then the scripts `get_dike_intrusion_data.py`, `dike_intrusion_lut.py`, `dike_intrusion_trigger_pn.py`, `dike_intrusion_detect_pn.py` and `dike_intrusion_locate_pn.py` can be run given a PhaseNet model directory, which can be found in `/from_scratch_model` and `/retrained_model`.

7.1.3 Figures

Figure 3 was made using `snr_reader.py` and `ml_reader.py`. Figures 4, 7 and 12 were made by `phasenet_eqt_tester.py`. Figure 6 was made using `loss_plot.py`. Figures 8 and 11 were made using `residual_times_statistics.py`. Figure 9 was made using `ml_scoring_plot.py`. Figure 10 was created using `QM_versus_PhaseNet_pickings.py` and `QM_versus_PN_plot.py`. Figure 13 was made using `STA_LTA_and_PN_plot.py`.

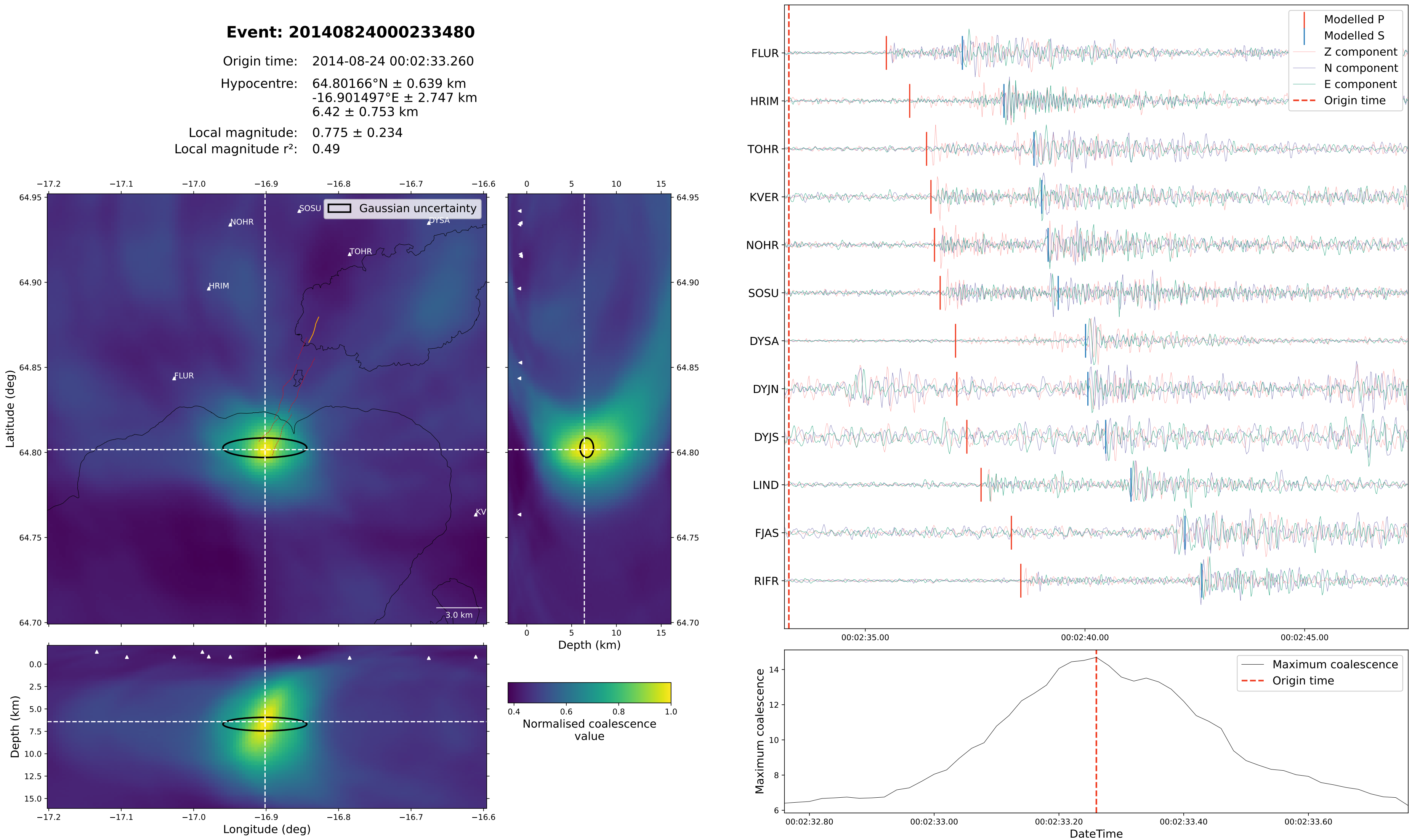

Figure 14: QuakeMigrate locate using STA/LTA onset functions. Bottom right shows the maximum coalescence (how well the onset functions align) against the earthquake origin time.

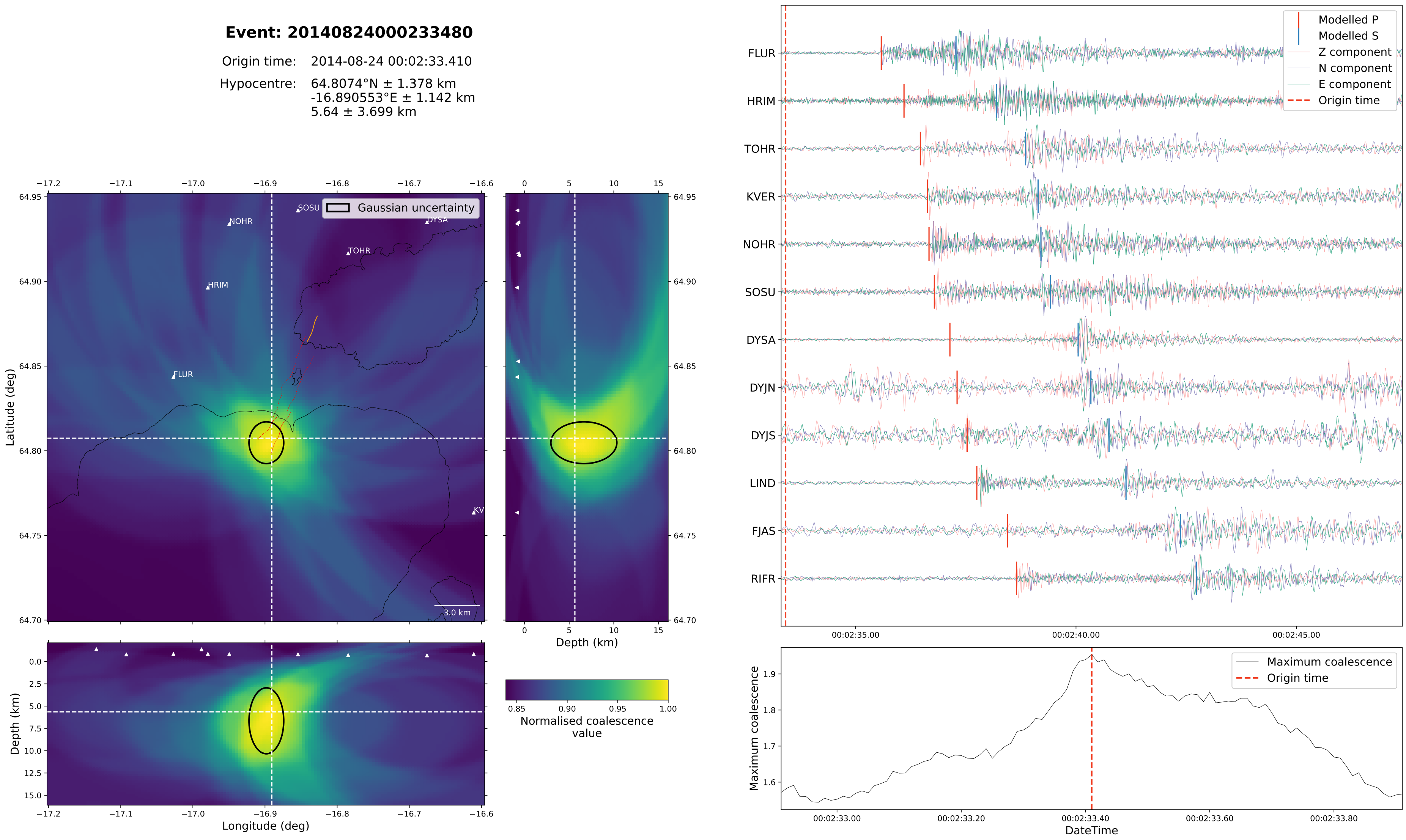

Figure 15: QuakeMigrate locate using PhaseNet onset functions.