JOURNAL OF PETROLEUM AND GEOTHERMAL TECHNOLOGY
ISSN: 2723-0988, e-ISSN: 2723-1496
Vol. 6 No. 1 2025

The Use of PhaseNet for Event Identification of Microearthquake Monitoring in Geothermal Field

Muhamad Firdaus Al Hakim [1)], I Putu Raditya Ambara Putra [*1)], Dina Kholish Annisa [1)], Haris Muhlisin [1)] Handika Putra Pratama [1)], Mohammad Rachmat Sule [2)], Fanata Yudha [3)]

1) Geophysical Engineering, Faculty of Mineral Technology, UPN Veteran Yogyakarta, Indonesia
2) Geophysical Engineering, Faculty of Mining and Petroleum Engineering, Institut Teknologi Bandung, Indonesia
3) Petroleum Engineering, Faculty of Mineral Technology, UPN Veteran Yogyakarta, Indonesia

* corresponding email: muhamad.firdausalhakim@upnyk.ac.id

ABSTRACT

Geothermal energy is a sustainable energy source that requires continuous microseismic monitoring to assess reservoir integrity and geomechanical behavior. Traditional phase identification methods are challenged by noisy environments and complex waveforms, especially in geothermal fields. This study explores the efficacy of PhaseNet, a deep learning neural network model, in detecting P and S wave arrival times for micro-earthquake events. The PhaseNet model was retrained using local seismic data from a geothermal field and tested for its performance in identifying seismic phases. The results were validated against a manual seismic catalog, with additional clustering and association analysis conducted using GaMMA and hypocenter locations determined with NonLinLoc. The findings demonstrate that PhaseNet, combined with GaMMA, provides robust phase detection capabilities, essential for early-stage monitoring in geothermal development.

Keywords: automatic phase detection; geothermal; microseismic

I. INTRODUCTION

Geothermal energy, as a renewable and sustainable source of power generation, has gained increasing attention in recent years due to its potential to mitigate greenhouse gas emissions and reduce dependence on fossil fuels. A major challenge in the exploitation of geothermal fields is the monitoring and management of microseismic events, which are small-scale earthquake-like events that occur as a result of fluid injection or extraction processes (Wibowo et al., 2022).

Microseismic monitoring is essential for assessing the integrity of the reservoir and understanding the underlying geomechanical processes (Folesky et al., 2016; Huang et al., 2017; Okamoto et al., 2018; Rossi et al., 2020). Currently, microseismic monitoring has been widely applied in various fields, including fault belts, volcanic activities, landslide-prone regions, mining, tunnel excavation, and geothermal and hydrocarbon reservoirs. Microseismic monitoring allows for the delineation of fracture networks by utilizing the parameters of detected microseismic events. Microseismic events in geothermal fields are often detected using seismometers or accelerometers placed strategically around the reservoir. However, accurately identifying the phases of these microseismic events is crucial for further analysis and interpretation. The arrival time of microseismic events provides important information for phase identification, source location, source mechanism analysis, and microseismic interpretation (Wang et al., 2018).

The phase picking is a challenging and important task in identifying earthquake characteristics, particularly in recognizing the S-wave phase, which is not the initial arrival phase (Lois et al., 2013; Namjesnik et al., 2021; Yu et al., 2020). To assist in determining the arrival time of the P and S wave phases, many automated picking techniques have been devised. The first example is the STA/LTA, also known as the short-term average or long-term average (Allen, 1978; Baer & Kradolfer, 1987). Although reasonably efficient, this approach is quite noise-prone, and it is exceedingly challenging to estimate the S wave's phase. The next example uses integrated Akaike Information Criterion and autoregressive (AR) modeling, which were created by Sleeman & van Eck in 1999. Deep learning with neural networks has been used for phase detection techniques during the past several years (Gentili & Michelini, 2006; Ross & Ben-Zion, 2014). Zhu & Beroza (2018) created the neural network method called PhaseNet, which was trained using a catalog of data from the Northern California Earthquake Data Center. However, the effectiveness of the automatic phase picking still needs to be tested for identification of micro-scale earthquakes.

Therefore, in this study, we use PhaseNet to detect phase arrival time from microseismic occurrences in waveform data collected at a geothermal field. Following that, we will also compared the travel time from PhaseNet picking with the available catalogue and determine its location.

II. METHODS

Data Availability

The data for this study were gathered from eight three-component seismometer stations strategically positioned across a geothermal field to monitor micro-earthquake (MEQ) activity. These stations were deployed continuously over a two-month period, capturing a comprehensive set of seismic waveforms. Each station recorded seismic activity in three orthogonal directions—vertical, north-south, and east-west—allowing for a full representation of the waveforms generated by seismic events. The three-component setup is crucial for distinguishing between different wave types and accurately identifying seismic phases, particularly P and S waves.

The waveform data were recorded at a sampling rate of 200 Hz, which provided high temporal resolution necessary for detecting the subtle and often weak signals characteristic of MEQ events. The high sampling rate is particularly advantageous in geothermal fields, where seismic signals can be noisy due to complex geological structures and fluid interactions. For ease of processing, the continuous waveform data from the first month were divided into one-hour segments. This segmentation facilitated efficient data management, reduced computational load, and allowed for more focused analysis in subsequent stages. The two-month dataset offers a representative view of seismicity within the geothermal reservoir, making it ideal for testing and validating the PhaseNet model.

Method

Initially, PhaseNet was trained using data from the Northern California Earthquake Data Center, which primarily consists of larger tectonic earthquakes. To apply this model effectively in a geothermal setting, where MEQs are often much smaller and more complex, the model required retraining with local data. The retraining process began with the first month of seismic data from the geothermal field. This dataset, consisting of 720 hours of continuous waveform recordings, was pre-processed by normalizing amplitudes and applying a bandpass filter to minimize low-frequency noise. These pre-processing steps are essential for enhancing the signal-to-noise ratio, especially in noisy environments like geothermal fields. The retraining was performed by using a supervised learning approach, where the manual picks from an existing seismic catalog were used as ground truth labels for the P and S wave arrivals. The PhaseNet model was trained to minimize the difference between its predicted arrival times and the catalog-based ground truth using a cross-entropy loss function, which helps the model improve its classification accuracy.

Following retraining, the PhaseNet model was tested on the second month's dataset. The evaluation involved comparing the P and S phase picks generated by the model with those from the seismic catalog. This comparison helped assess the model's performance, focusing on metrics such as the total number of correctly identified phases, the rate of false positives, and the accuracy of phase arrival time predictions. These metrics were vital for determining whether retraining had improved the model's capability to detect seismic phases in the geothermal setting.

Once the PhaseNet model identified potential P and S wave arrivals, the next step was to associate these phases with specific seismic events. This was accomplished using GaMMA, a Gaussian Mixture Model-based algorithm for phase association. GaMMA operates by analyzing the temporal and spatial distribution of the detected seismic phases to cluster them into groups that likely originate from the same seismic event.

GaMMA is an unsupervised machine learning algorithm , meaning it does not rely on labeled data for training. Instead, it clusters phases based on their arrival times across different stations and their spatial relationships using Bayesian Gaussian mixture model that has been widely used in a variety of research fields, including image processing (Permuter et al., 2006), speech recognition (Reynolds & Rose, 1995), and earthquake studies (Ross et al., 2020; Seydoux et al., 2020). For each detected phase, GaMMA calculates the likelihood that it is part of a seismic event by comparing its arrival time with those from other stations. To ensure that the detected phases truly correspond to an event, at least four phases—two P and two S—must be identified from different stations.

In this study, GaMMA processed the PhaseNet-generated picks and successfully associated a significant number of phases with specific events. However, approximately 93.6% of the detected phases were discarded because they could not be reliably paired with corresponding phases from other stations, emphasizing the importance of high-quality seismic data and robust event association techniques. The clustered phases were then compared with the manual seismic catalog to validate their accuracy. Events identified by both PhaseNet-GaMMA and the catalog were further analyzed to assess the consistency of the results, particularly through the use of Wadati diagrams, which plot the relationship between the arrival times of P and S waves. The close match between the catalog and the PhaseNet-GaMMA results indicated that the combination of these methods is effective for event identification.

Once the phases had been associated with specific events using GaMMA, the next step was to determine the hypocenter locations of these events by employing NonLinLoc (Lomax et al., 2000) program. NonLinLoc uses a non-linear global search approach, which allows for more accurate hypocenter determination compared to traditional linear methods, particularly in complex environments like geothermal fields.

The NonLinLoc algorithm uses the arrival times of P and S waves, along with a velocity model of the Earth, to compute the most likely location of a seismic event. In this study, a homogeneous velocity model was used for initial computations. Although this simplified velocity model may not fully capture the complexity of the geothermal field's subsurface, it was sufficient for a preliminary analysis. The NonLinLoc algorithm applies Metropolis-sampling, a Monte Carlo-based method that allows for efficient exploration of the solution space, making it computationally feasible to handle large datasets.

Hypocenters were calculated both for the events identified by PhaseNet-GaMMA and for those in the manual catalog. The results were then compared to assess the accuracy of the PhaseNet-GaMMA model in determining event locations. While PhaseNet-GaMMA identified more events than the manual catalog, the spatial distribution of hypocenters from both sources showed good agreement, indicating that the PhaseNet-GaMMA combination can reliably identify and locate micro-earthquake events in geothermal fields. However, further refinement of the velocity model and additional data quality control would likely improve the accuracy of these location estimates.

III. RESULTS AND DISCUSSION

Comparison before-after training

Before retraining, the model identified 13062 total P and S phases, but after retraining, it identified 52355 total P and S phases. Although there are more P and S wave phase identifications, further investigation is required to determine whether these phases are related to a certain event. In Figure 1, it can be seen that the histogram of the permanence model before and after training still shows the same pattern. If we compare the time difference of the results after and before retraining with the available catalogue picking data, the number of matches decreases from 636 to 511. This may be due to the fact that the micro-seismic catalogue data we used is still not sufficiently large and well quality-controlled, so we need to evaluate it further. It also means that the existing PhaseNet model is robust enough to detect the phase of micro-earthquakes in geothermal fields. Therefore, in the next test we still use the existing PhaseNet model that has not been retrained.

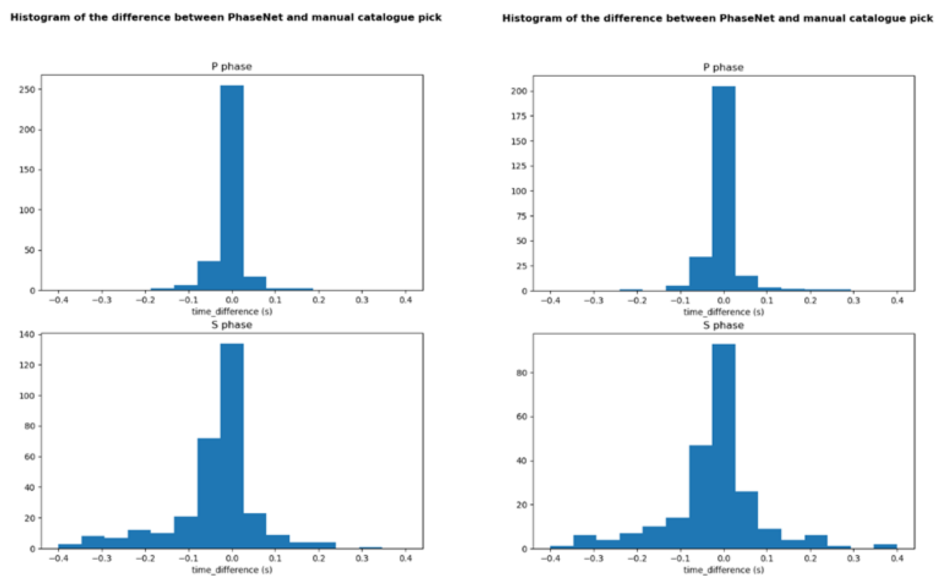

Figure 1. Histogram of the difference between the identification resolusi of the PhaseNet model that has not been retrained (left) and has been retrained (right) the manual catalogue of picks.

Result of PhaseNet-GaMMA event identification

For the processed data, the results of event identification using PhaseNet generate a total of 20313 phases with specifics for 10323 P and 9990 S. Phase identification using the PhaseNet model resulted in a total number of 20313 phases, with 10323 P and 9990 S. Then after being associated with GaMMA, almost 93.6% of the total phases were taken out because they did not have a partner. This is expected because a minimum of 4 total P and S phases are required for us to be confident that it is the same event and the hypocenter parameters can be well determined.

Figure 2 shows some identification results obtained using PhaseNet-GaMMA and the catalogue. It is clear that some phases can be recognized using PhaseNet-GaMMA but not in the manual pick catalog, and vice versa. However, for the phases that are both detected in PhaseNet-GaMMA and the catalogue, the similarities are quite good. The average picking time difference is 0.1 second and all are still below 1 second for both P and S.

We also performed a Wadati diagram analysis of the Phasenet-GaMMA results and compared them with the catalogue data (Figure 3). The Wadati diagram is indeed one of the ways we can perform quality control of the P and S phase identification results. The Vp/Vs of the catalogue data shows a value of 1.642, while the Phasenet-GaMMA shows a value of 1.552. These values still match the average Vp/Vs value in the Earth's crust of around 1.75.

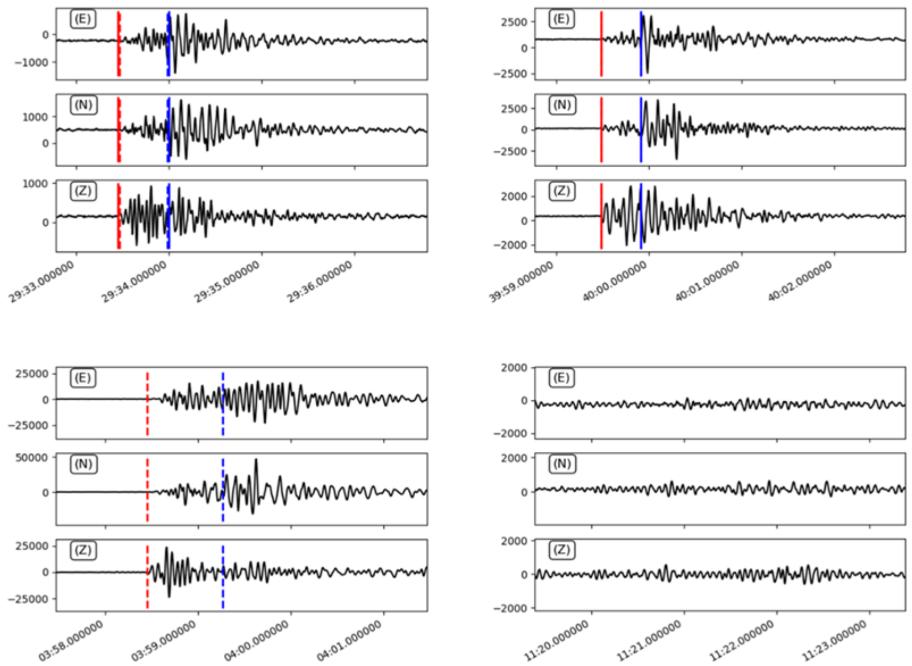

Figure 2. Some examples of the results of P phase (red) and S phase (blue) picking overlaid with waveform. The dashed-line is the result of PhaseNet and GaMMA, while the solid line is the data from the data catalog used in this study.

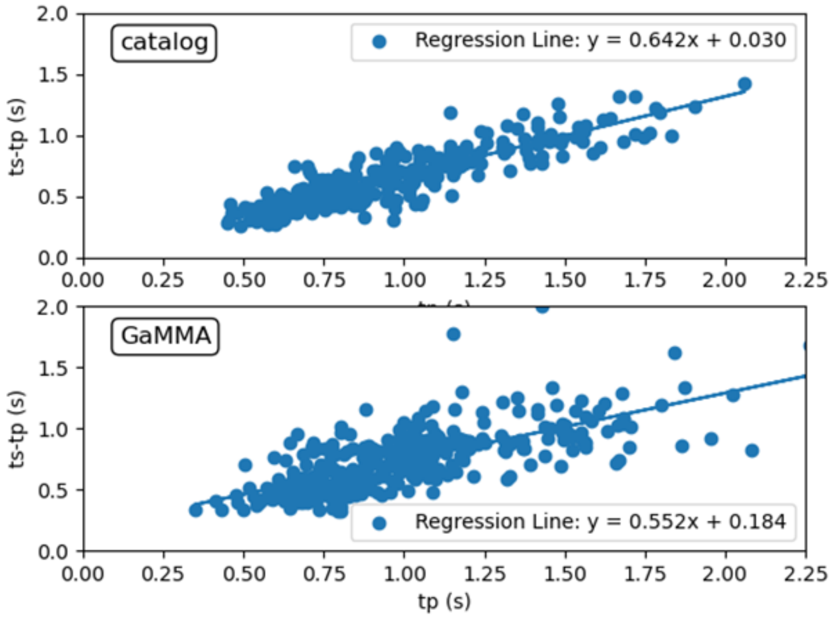

Figure 3. Wadati Diagram (ts-tp vs tp-to) plot from catalog data (first row) and identification results using PhaseNet and GaMMA (second row). Vp/Vs value is the slope of the regression plus 1, so the Vp/Vs value in the catalogue data is 1.642 and the PhaseNet-GaMMA result Vp/Vs value is 1.552.

Event Location using NonLinLoc

The hypocenter results from PhaseNet-GaMMA show poor results when compared to the hypocenter obtained from the catalogue. The location of the hypocentre is generally sparser compared to the existing catalogue. This makes sense because the approach used is very simple by including a constant (homogeneous) velocity parameter. Therefore, we then applied NonLinLoc to determine the hypocenter of the phase. We applied the hypocenter to the data from PhaseNet-GaMMA as well as the phase data available in the catalogue with the aim of finding out how far the time in the phase identification compares with the position of the hypocenter.

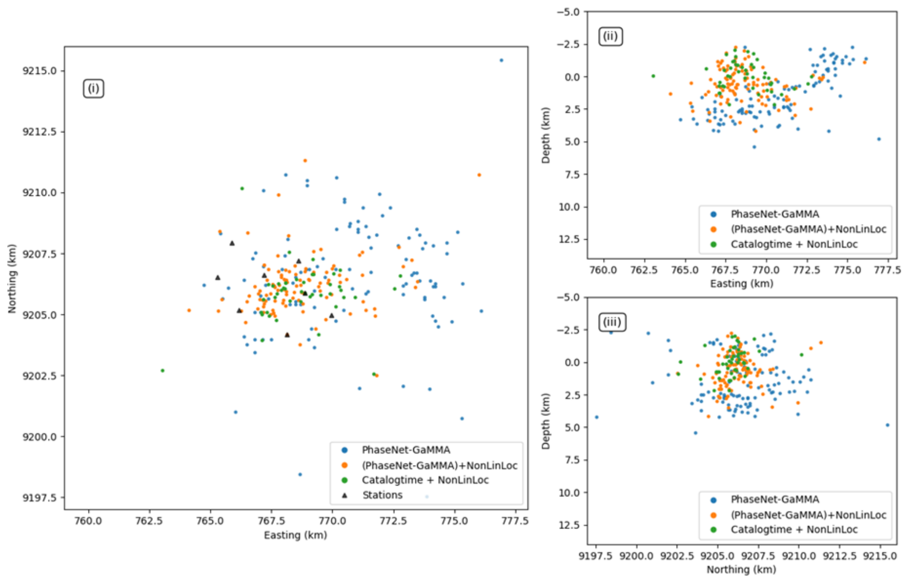

Figure 4. Comparison between the locations of the relocated catalogue data, the catalogue processed with NonLinLoc, PhaseNet-GaMMA, and also PhaseNet-GaMMA whose locations were determined with NonLinLoc.

Figure 4 shows the results of the hypocenter location determination. There are 107 events that have successfully determined the location of the hypocenter from the PhaseNet-GaMMA data while only 45 events from the catalogue data have successfully identified the location of the hypocenter. Although the numbers are different, the distribution of hypocenter locations shows agreement, both between PhaseNet-GaMMA + NonLinLoc, catalogue + NonLinLoc, and also the relocation catalogue. The only discrepancy is shown by the direct result from GaMMA. It can be seen in Figure 4 that the direct results from PhaseNet-GaMMA show the hypocenter locations are very sparse and far from the relocation catalogue.

Then we linked events that we considered were the same based on their origin time. The average difference in hypocenter locations between PhaseNet-GaMMA+NonLinLoc and Catalog+NonLinloc is 2 km. This number is not significantly different from the average difference between Catalog+NonLinLoc and the relocation catalog, which is just 1.6 km. Based on these results, we suggest that the PhaseNet-GaMMA combination performs well in recognizing micro-earthquake event phases.

IV. CONCLUSION

Results of this work indicate that the PhaseNet machine learning model, when paired with GaMMA, is sufficiently good and robust in identifying the wave phase of micro-earthquake events in geothermal fields. The difference in time between the manual pick catalog results and PhaseNet-GaMMA shows no significant results. Similarly, when we compare hypocenter locations processed using NonLinLoc. Therefore PhaseNet-GaMMA is ideal for usage in the early identification stage of geothermal fields, although manual quality checks may still be required.

ACKNOWLEDGEMENTS

We extend our heartfelt appreciation to all those who played a role in the successful completion of this study. We are especially grateful to the Geophysical Engineering Department, the Faculty of Mineral and Energy Technology, and Universitas Pembangunan Nasional "Veteran" Yogyakarta for their invaluable financial support, which made this research possible.

REFERENCES

Allen, R. V. (1978). Automatic earthquake recognition and timing from single traces. Bulletin of the Seismological Society of America, 68(5), 1521–1532. https://doi.org/10.1785/BSSA0680051521

Baer, M., & Kradolfer, U. (1987). An automatic phase picker for local and teleseismic events. Bulletin of the Seismological Society of America, 77(4), 1437–1445. https://doi.org/10.1785/BSSA0770041437

Folesky, J., Kummerow, J., Shapiro, S. A., Häring, M., & Asanuma, H. (2016). Rupture directivity of fluid-induced microseismic events: Observations from an enhanced geothermal system. Journal of Geophysical Research: Solid Earth, 121(11), 8034–8047. https://doi.org/10.1002/2016JB013078

Gentili, S., & Michelini, A. (2006). Automatic picking of P and S phases using a neural tree. Journal of Seismology, 10(1), 39–63. https://doi.org/10.1007/s10950-006-2296-6

Huang, W., Wang, R., Li, H., & Chen, Y. (2017). Unveiling the signals from extremely noisy microseismic data for high-resolution hydraulic fracturing monitoring. Scientific Reports, 7(1), 11996. https://doi.org/10.1038/s41598-017-09711-2

Lois, A., Sokos, E., Martakis, N., Paraskevopoulos, P., & Tselentis, G.-A. (2013). A new automatic S-onset detection technique: Application in local earthquake data. GEOPHYSICS, 78(1), KS1–KS11. https://doi.org/10.1190/geo2012-0050.1

Namjesnik, D., Kinscher, J., Gunzburger, Y., Poiata, N., Dominique, P., Bernard, P., & Contrucci, I. (2021). Automatic Detection and Location of Microseismic Events from Sparse Network and Its Application to Post-mining Monitoring. Pure and Applied Geophysics, 178(8), 2969–2997. https://doi.org/10.1007/s00024-021-02773-4

Okamoto, K., Yi, L., Asanuma, H., Okabe, T., Abe, Y., & Tsuzuki, M. (2018). Triggering processes of microseismic events associated with water injection in Okuaizu Geothermal Field, Japan. Earth, Planets and Space, 70(1), 15. https://doi.org/10.1186/s40623-018-0787-7

Permuter, H., Francos, J., & Jermyn, I. (2006). A study of Gaussian mixture models of color and texture features for image classification and segmentation. Pattern Recognition, 39(4), 695–706. https://doi.org/10.1016/j.patcog.2005.10.028

Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using Gaussian mixture speaker models. IEEE Transactions on Speech and Audio Processing, 3(1), 72–83. https://doi.org/10.1109/89.365379

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation (pp. 234–241). https://doi.org/10.1007/978-3-319-24574-4_28

Ross, Z. E., & Ben-Zion, Y. (2014). Automatic picking of direct P, S seismic phases and fault zone head waves. Geophysical Journal International, 199(1), 368–381. https://doi.org/10.1093/gji/ggu267

Ross, Z. E., Trugman, D. T., Azizzadenesheli, K., & Anandkumar, A. (2020). Directivity Modes of Earthquake Populations with Unsupervised Learning. Journal of Geophysical Research: Solid Earth, 125(2). https://doi.org/10.1029/2019JB018299

Rossi, C., Grigoli, F., Cesca, S., Heimann, S., Gasperini, P., Hjörleifsdóttir, V., Dahm, T., Bean, C. J., Wiemer, S., Scarabello, L., Nooshiri, N., Clinton, J. F., Obermann, A., Ágústsson, K., & Ágústsdóttir, T. (2020). Full-Waveform based methods for Microseismic Monitoring Operations: an Application to Natural and Induced Seismicity in the Hengill Geothermal Area, Iceland. Advances in Geosciences, 54, 129–136. https://doi.org/10.5194/adgeo-54-129-2020

Seydoux, L., Balestriero, R., Poli, P., Hoop, M. de, Campillo, M., & Baraniuk, R. (2020). Clustering earthquake signals and background noises in continuous seismic data with unsupervised deep learning. Nature Communications, 11(1), 3972. https://doi.org/10.1038/s41467-020-17841-x

Sleeman, R., & van Eck, T. (1999). Robust automatic P-phase picking: an on-line implementation in the analysis of broadband seismogram recordings. Physics of the Earth and Planetary Interiors, 113(1–4), 265–275. https://doi.org/10.1016/S0031-9201(99)00007-2

Wang, P., Chang, X., & Zhou, X. (2018). Estimation of the Relative Arrival Time of Microseismic Events Based on Phase-Only Correlation. Energies, 11(10), 2527. https://doi.org/10.3390/en11102527

Wibowo, D. A., Ramadhan, I., Agoes Nugroho, I., Baroek, M. C., Ganefianto, N., Azis, H., Suryantini, Sahara, D. P., & Mozef, P. W. (2022). Microseismic and Focal Mechanism Analyses for Structural Interpretation – Muara Laboh Geothermal Field. IOP Conference Series: Earth and Environmental Science, 1014(1), 012004. https://doi.org/10.1088/1755-1315/1014/1/012004

Yu, Z.-C., Yu, J., Feng, F.-F., Tan, Y.-Y., Hou, G.-T., & He, C. (2020). Arrival picking method for microseismic phases based on curve fitting. Applied Geophysics, 17(3), 453–464. https://doi.org/10.1007/s11770-020-0831-9

Zhu, W., & Beroza, G. C. (2018). PhaseNet: A Deep-Neural-Network-Based Seismic Arrival Time Picking Method. Geophysical Journal International. https://doi.org/10.1093/gji/ggy423

Zhu, W., McBrearty, I. W., Mousavi, S. M., Ellsworth, W. L., & Beroza, G. C. (2022). Earthquake Phase Association Using a Bayesian Gaussian Mixture Model. Journal of Geophysical Research: Solid Earth, 127(5). https://doi.org/10.1029/2021JB023249