

Deliverable

Deliverable 2.1 - Deep learning methods for microseismicity characterization

Report information	
Work package	WP2: Innovation in real-time monitoring and big-data analysis for EGS
Lead	ETH
Authors	Peidong Shi, ETH; Francesco Grigoli, UNIPi; Federica Lanza, ETH; Katinka Tuinstra, ETH; & Stefan Wiemer, ETH
Reviewers	
Approval	Stefan Wiemer
Status	Draft
Dissemination level	internal
Will the data supporting this document be made open access?	Yes* *data will be soon available via FDSN service
If No Open Access, provide reasons	
Delivery deadline	31.08.2021
Submission date	31.08.2021
Intranet path	DOCUMENTS/DELIVERABLES/Deliverable2.1.docx



Table of Contents

Summary	3
1. Automated End-to-End Seismic Catalog Builder based on Machine Learning and Waveform Migration	3
2. Data Set	7
3. Results and comparisons	8
4. Conclusions and Outlook	15
References	16

Summary

Obtaining an accurate and complete earthquake catalog is fundamental for the seismological analysis of induced microseismicity during geothermal injection operations. It is also crucial to rapidly update the earthquake catalog in real-time during the injection process for assessing the stimulation effects and mitigating the risk of the induced earthquakes. In this deliverable, we introduce a new framework to build earthquake catalogs directly from continuous streaming seismic data. The new framework combines machine learning (ML) and waveform migration techniques, and harnesses the advantages provided by both methods. The new workflow does not require phase detection nor association, instead the probabilities of P- and S-phases generated by ML models are directly migrated into the space and time domain to obtain source location and origin time. The whole framework is data driven, thus does not have any explicit parameters to tune. The automated nature and parallel efficiency make it capable for real-time monitoring.

We applied the proposed framework to the COSEISMIQ data set to monitor induced earthquakes close to a geothermal production site. Compared to conventional migration methods, we are able to produce much clearer coherence images with higher resolution and reliability especially for small seismic events. By using phase probabilities for migration, extrinsic factors affecting recorded waveforms and migration results, such as site effects, instrument response, propagation effects, and ambient noise conditions are largely removed. In addition, the proposed approach utilizes the spatial coherence across the whole array to detect and locate seismic sources; thus the uncertainties brought by outliers from individual stations are largely reduced. Applying the proposed framework to one month of continuous data collected by the COSEISMIQ network as part of the monitoring of the Hengill geothermal area in Iceland, we were able to automatically detect and locate 886 earthquakes, which is more than two times the earthquakes reported in the manual catalog (431 earthquakes). This deliverable constitutes a first report on the current work, and in the future we will continue to develop and integrate newly trained ML models to the proposed approach as well as improve efficiency and precision of the event locations. The next data application will use the data set collected during the upcoming stimulations scheduled at the Utah FORGE EGS site. This will be an opportunity to test the real-time performance of the method to monitor induced seismicity within the Adaptive traffic Light System (ATLS) framework.

1. Automated End-to-End Seismic Catalog Builder based on Machine Learning and Waveform Migration

Real-time monitoring of the induced microseismic events during geothermal stimulation operations is a key process to manage the seismic risk associated with these events. A real-time seismic monitoring system naturally calls for an automated workflow that can rapidly process the continuous incoming seismic streams and produce earthquake catalogs with minimal or no human intervention. To detect and locate earthquakes, conventional real-time seismic monitoring system involves several independent procedures, such as event detection, phase picking, phase association and event location. Among these procedures, phase picking and association are essential for location accuracy and are the most difficult part to perform especially when multiple or overlapping events occur in a short time period. Small and overlapping events are even more common in enhanced geothermal systems (EGS) where we deliberately create thousands of microseismic events of which many may show overlapping waveform trains.

Traditional seismic phase picker relies on the statistical characteristic of seismic waveforms, such as short-term-average over long-term-average (STA/LTA) and kurtosis, to identify the arrival-times of P- and S-phases (Allen, 1978; Withers et al., 1998). The performance of these statistical phase pickers highly depends on the input parameters, such as time window length and filtering frequency band. The optimal parameters often differ for different seismic events and even stations and are subject to manual adjustment in practice, which obstructs real-time monitoring. Benefiting from massive seismic data and the many existing labels therein, recent developments using ML for event detection and phase picking outperform traditional phase picking algorithms and even human experts in terms of picking efficiency and precision (Mousavi et al., 2020).

However, for accurately detecting and picking the phase arrivals, threshold parameters are still required. In addition, the following processing procedures, such as phase association, are still challenging tasks in real-time monitoring.

Waveform migration/stacking techniques (Grigoli et al., 2013; Shi et al., 2019a) are promising methods to automatically detect and locate earthquakes without the need for phase picking and association. These methods first back-project continuous seismic waveforms or characteristic functions of waveforms into the space and time domain and then stack the back-projected data to detect and locate seismic sources. During the back-projection process, seismic phases are automatically associated and the wavefield coherence from the whole array is exploited to detect and locate seismic sources. Therefore, these methods are particularly powerful at detecting and locating microseismic events or overlapping events, and have been successfully applied in induced and triggered seismicity analysis (Grigoli et al., 2018; Shi et al., 2019b).

Despite the excellent performance over small events and the automated nature, the migration-based methods still have several disadvantages that limit their practical applications. First, the computational cost of the migration-based methods is very high, making them unsuitable for real-time monitoring without the aid of a high-performance computing system. Second, for enhancing the signal-noise-ratio (SNR) of coherence images, various characteristic functions are proposed, such as STA/LTA, energy and cross-correlation coefficient of waveforms. The calculation of these characteristic functions usually involves different parameters, e.g. time window length and frequency band, which introduce additional dependencies and are not conducive to automated processing. Lastly, the recognition of seismic events from a continuous migration volume is a difficult task and still needs further research.

To fully leverage the advantages of both the ML techniques and the migration-based methods and at the same time overcome some disadvantages of both methods, we propose a new framework to perform automated seismic event detection and location, called MACHine Learning Migration Imaging (MALMI). The proposed framework (MALMI) consists of four core modules and the detailed workflow is shown in Figure 1.

- *Data pre-processing and workflow configuration module*: this module processes input station metadata and continuous seismic data of various formats and generates input data sets that ML models can handle. Global parameters, such as the input and output directories for different modules, are also configured to facilitate the automated/real-time processing.
- *Machine learning module*: a pre-trained ML model is used to generate the probabilities of the existence of earthquake, P-phase, and S-phase in continuous data (Figure 2). It is worth noting that the generated probabilities are continuous in time and no event detection nor phase picking is needed in this module. The continuous probabilities will be used as inputs for the migration module. Currently, EQTransformer (Mousavi et al., 2020) is used as the ML engine to produce event and phase probabilities. In the future, we plan to integrate additional ML models, currently in development, as alternative ML engines.
- *Array event scan module*: theoretically, the generated continuous phase probabilities can be directly used as inputs for migration location. However, in order to improve the computational efficiency, this module is developed to continuously scan the generated probabilities and automatically detect and separate different events for further migration location (Figure 3). Therefore, instead of performing migration on continuous data, only the probability data segment tailored for each locatable event will be further processed by the migration module. Each probability data segment only contains one event as long as the phase arrivals of different events do not overlap. Compared to conventional migration schemes, this module provides two key benefits: first, it reduces the computational cost dramatically; second, since this module already automatically identifies and separates different seismic events, the

migration module does not need to recognize events from the 4-dimensional migration volumes, which is often difficult.

- Migration module:* Phase probability segments of each event are fed into a migration engine to find the source location and origin time. The current implemented migration engine is LOKI (Grigoli et al., 2018), which is computationally optimized using OpenMP. Testing shows that using 10 threads from a single laptop, the computing time spent on the migration process for one event is around 10-15 seconds. The array event scan module in combination with an efficient migration engine makes the whole framework capable of real-time seismic monitoring when enough computing resources are secured (e.g. using one computing node with many cores from high-performance clusters). The computational resources required for real-time monitoring will depend on the data set to be processed and will be evaluated according to the data that will be acquired at Utah FORGE.

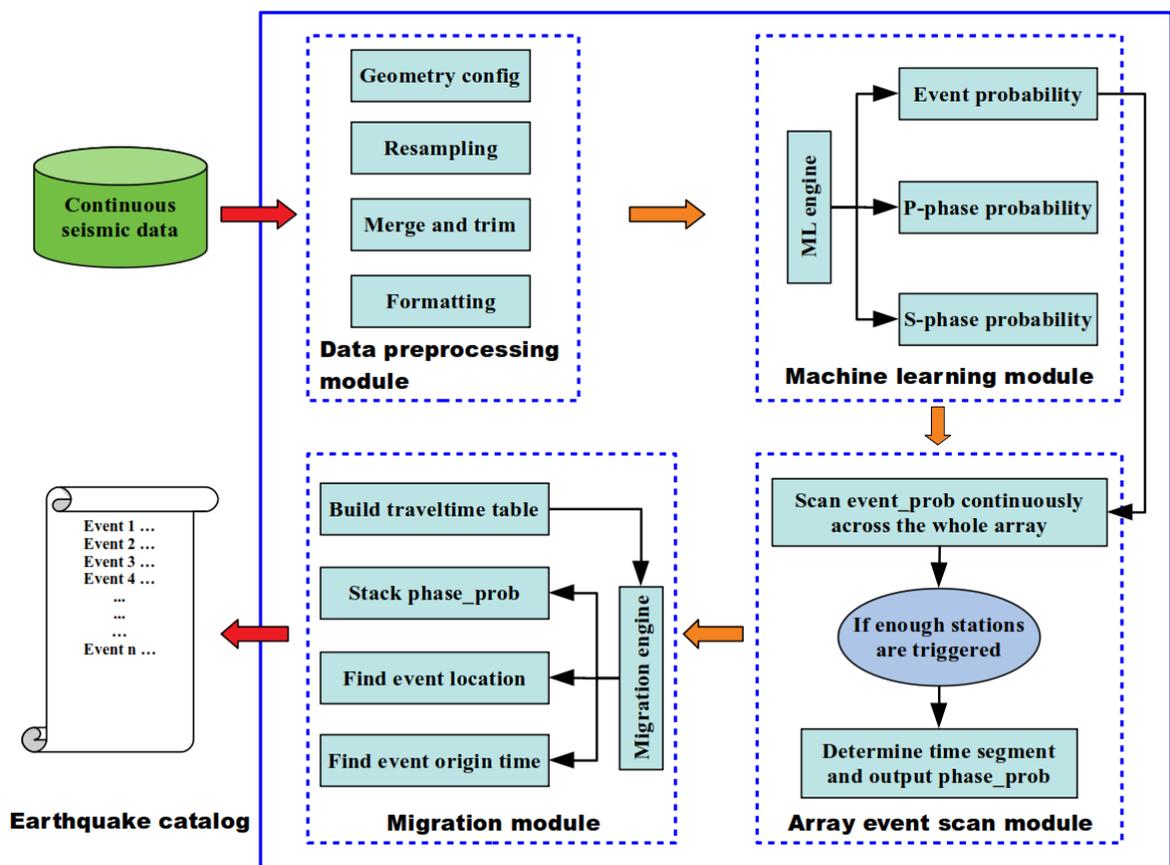


Figure 1. Schematic of the MALMI workflow. MALMI acts as an end-to-end earthquake catalog builder and is capable of real-time seismic monitoring.

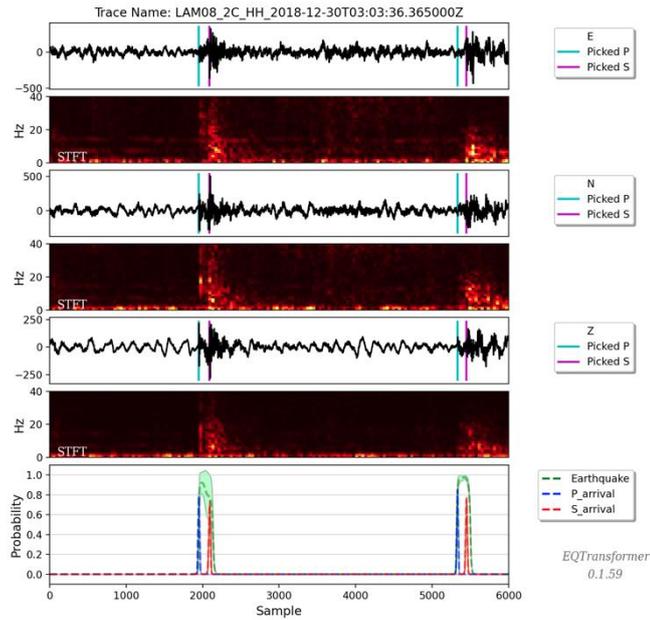


Figure 2. 60 seconds record length seismic data showing two microseismic events and the associated event, P-phase and S-phase probabilities, respectively in green, blue and red lines (bottom panel). Pick probabilities are generated by the pre-trained machine learning model EQTransformer (Mousavi et al. 2020).

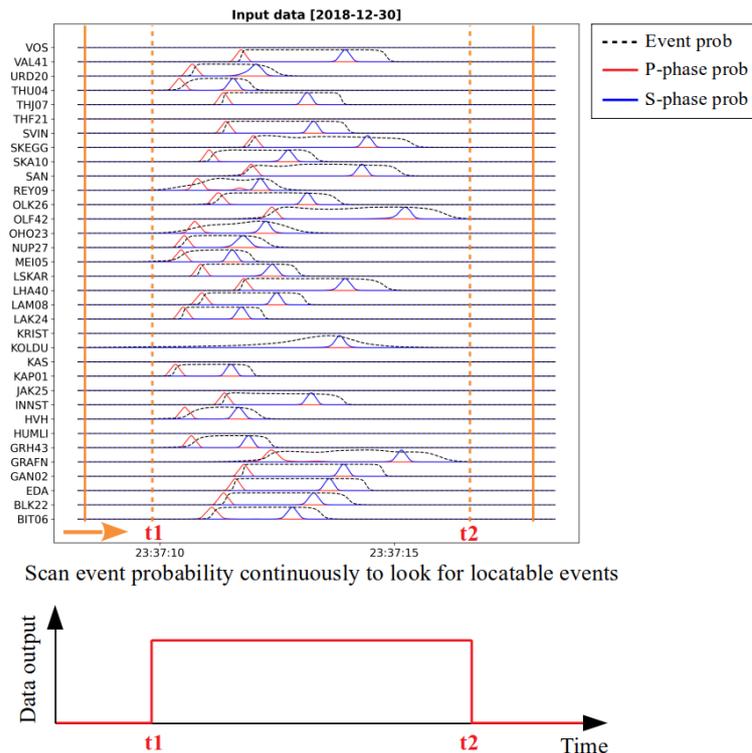


Figure 3. Schematic illustration showing the processe of the array event scan module which scans the whole array for identifying locatable seismic events and outputs probability data segments for migration.

The proposed framework acts as an end-to-end earthquake catalog builder which can directly process continuous seismic data and produce a final event catalog. Compared to conventional automated event location workflow (which includes traditional phase pickers or ML pickers) and migration techniques which require parameter testing and optimizing, the proposed approach is data driven and only involves minimal input parameters. The only required input parameters are the detection threshold and the minimum number of

triggered stations used in the array event scan module, and can be conveniently determined according to the lower limit of earthquake magnitude to be detected.

2. Data Set

We applied this approach to the data set collected within the COSEISMIQ project funded by the EU-Geothermica program. The COSEISMIQ project has been carried out in the Hengill region, a geothermal system located in Iceland. This region is located in the southern end of the Western Volcanic Zone (WVZ), at the triple junction between the WVZ, the Reykjanes Peninsula (RP), the landward extension of the Reykjanes spreading ridge, and the South Iceland Seismic Zone (SISZ) (Figure 4). Therefore, the area is characterized by a complex local geology and tectonic setting, and abundant natural seismicity. The Hengill geothermal system started to be exploited for electrical power and heat production since the late 1960s. In this region, the two largest geothermal power plants in Iceland that are currently in operation, i.e. the Nesjavellir and the Hellisheidi power stations (Figure 4). Due to its complex tectonic setting and the ongoing geothermal operations, several thousand of induced and natural seismic events occur every year. In the Hengill area, the permanent seismic network is run by the Iceland GeoSurvey (ISOR) and the Icelandic Meteorological Office (IMO) and comprises a total of 14 stations. Since November 2018 the network has been densified with 23 additional temporary stations and an array of 5 stations (Figure 5). This temporal network includes 10 broadband STS-2 (120s), 5 Guralp 6D (30s) and 8 5s-Lennartz. During the period spanning 02.11.2018-02.07.2020, ISOR reported a total of 12,157 events in the Hengill region, 69 of them with magnitudes larger than 2.0 M_L . Seismic sequences in this area are usually characterized by short inter-event times and small magnitudes below the resolution of regional networks. This makes the COSEISMIQ dataset a perfect application for testing our seismic data analysis workflow.

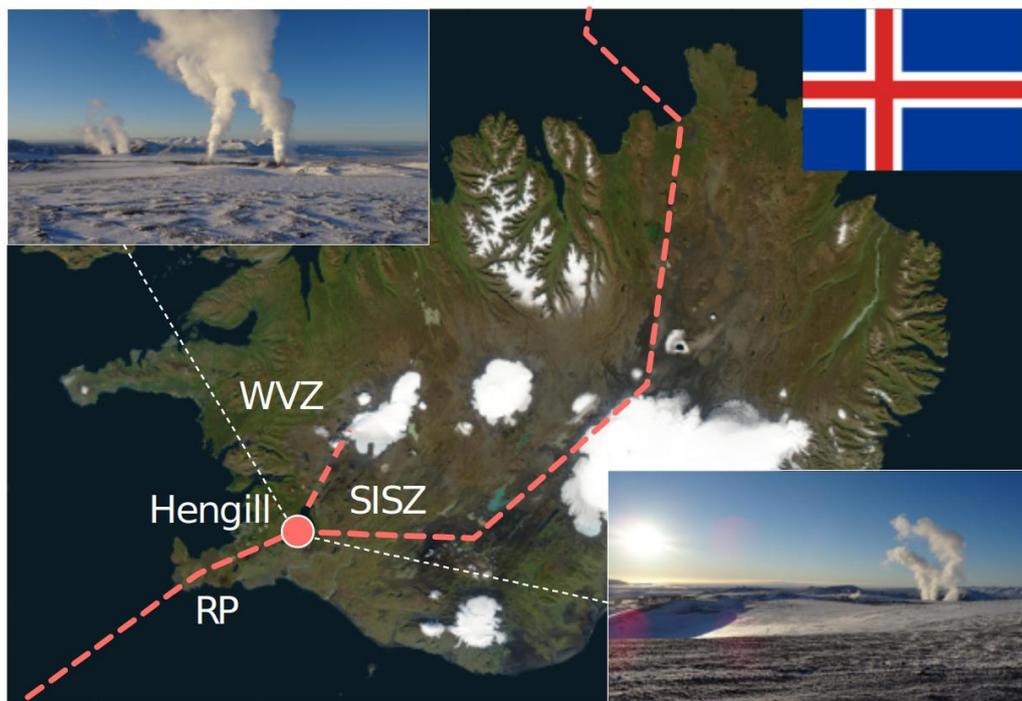


Figure 4. Map of Iceland showing the location of the Hengill geothermal area (red dot). Hengill is at the triple junction between the Reykjanes Peninsula oblique rift (RP), the Western Volcanic Zone (WVZ), and the transform-type South Iceland Seismic Zone (SISZ).

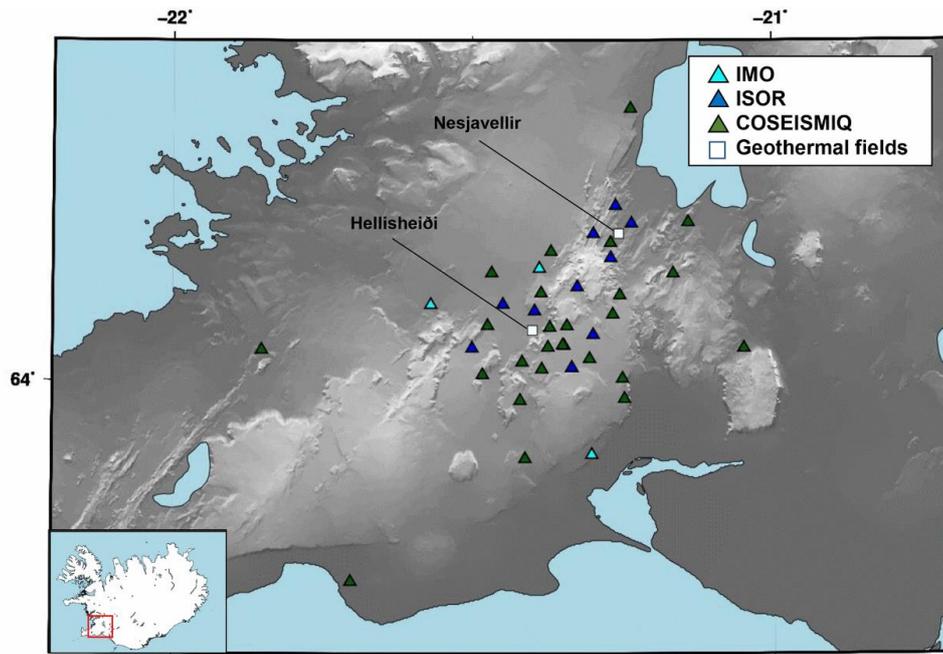


Figure 5. Seismic network around the Hengill geothermal site: IMO stations are shown as light blue triangles, ISOR's stations are indicated by dark blue triangles and dark green triangles show the temporary network deployed as part of the COSEISMIQ project. The Nesjavellir and Hellisheiði geothermal fields are marked with white squares (from Rossi et al. 2020).

3. Results and comparisons

We applied MALMI to one month of continuous data recorded by the COSEISMIQ network starting from 2018-12-01 to 2018-12-31. The migration area includes the whole COSEISMIQ array and has a spatial dimension of 100 km x 100 km x 20 km in north, east and depth directions, respectively. The spatial resolution of the migration grid is 400 meters. Since this region is seismically active, we use a large migration area and relatively coarse grid to accommodate regional earthquakes which might be spread over a broad area. In situations where earthquakes are limited to a small region, a finer migration grid can be used to improve the location accuracy without increasing the computational cost. With an event probability larger than 0.1 and a minimum number of 3 triggered stations, we were able to detect and locate 886 earthquakes. In comparison, the manual catalog contains 431 earthquakes. The temporal evolution of the earthquakes in the MALMI catalog and in the manual catalog is shown in Figure 6. Although we detected doubled the number of earthquakes, the cumulated curves of the earthquakes in the two catalogs are still highly similar. In addition, the temporal clustering features between the two catalogs are consistent, which also demonstrates the reliability of the new and more complete MALMI catalog.

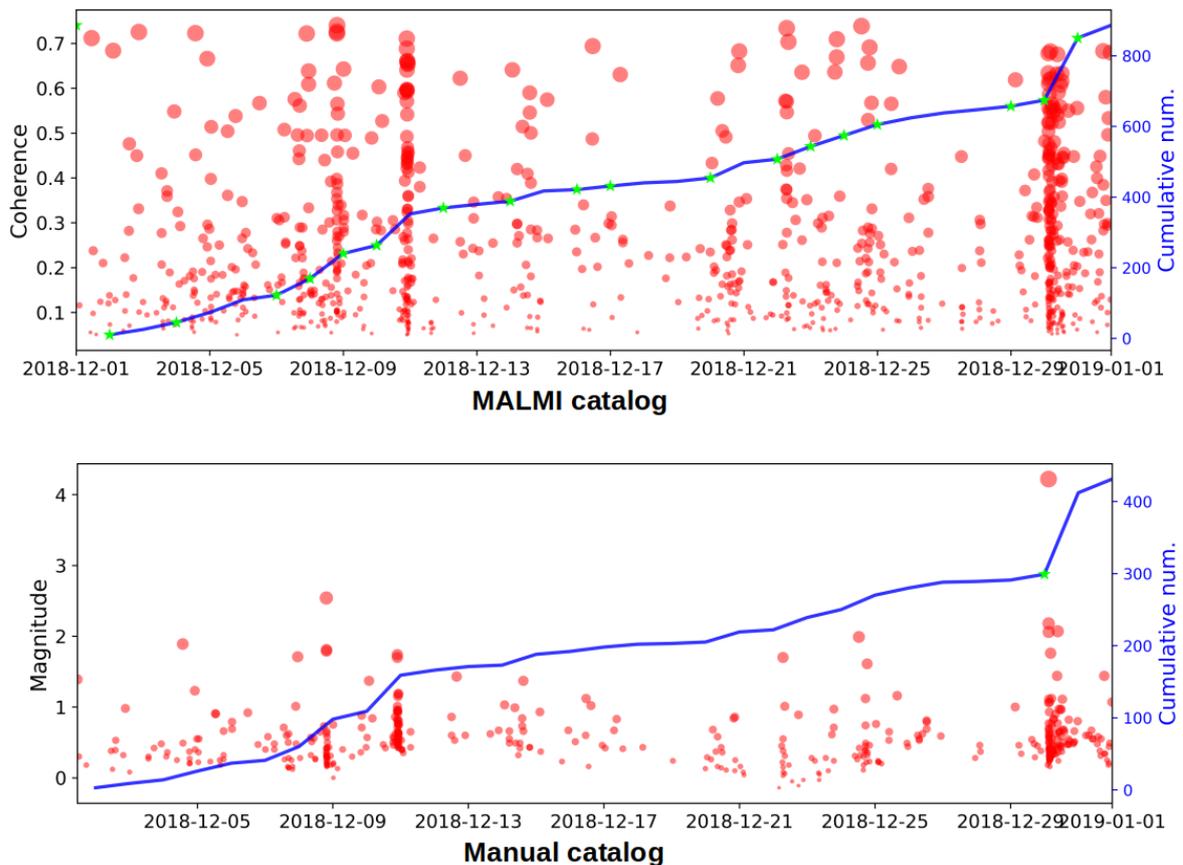


Figure 6. Temporal evolution (red dots) and cumulative number (blue line) of earthquakes in the MALMI catalog (top panel) and in the manual catalog (bottom panel). Note that in the plot of the MALMI catalog, the left Y-axis is migration coherence; whereas in the plot of the manual catalog the left Y-axis is magnitude.

The spatial distribution of the earthquakes in the MALMI catalog and in the manual catalog is shown in Figures 7-8. It is observed that the distribution of major earthquake clusters is in good agreement between the two catalogs. In addition, both depth and origin time of these clustered events are consistent across the two catalogs (Figures 7-8). Since location within the migration method is performed on a discrete grid with spacing of 400 m., due to this grid effect, a slightly scattered distribution of the earthquakes in the MALMI catalog can be observed from the map. However, relocation techniques such as the real-time double-difference relocation module SCRTDD developed by the Swiss Seismological Service (SED) and embedded in the SeisComp3 real time monitoring framework can be further applied to the obtained MALMI catalog to produce a more clustered relocation catalog. Nevertheless, we demonstrate that MALMI has a superior efficiency and location precision with respect to the current conventional migration techniques.

Furthermore, the MALMI catalog contains more earthquakes than the manual catalog. It can be observed from Figures 7-8 that most of the additional detected and located earthquakes belong to or are close to the same major earthquake clusters. This demonstrates that the newly detected earthquakes in the MALMI catalog are most likely to be microseismic events below the noise level or beyond the manual picking ability. There is also a small amount of scattered distributed earthquakes detected in the MALMI catalog which might represent smaller background earthquakes.

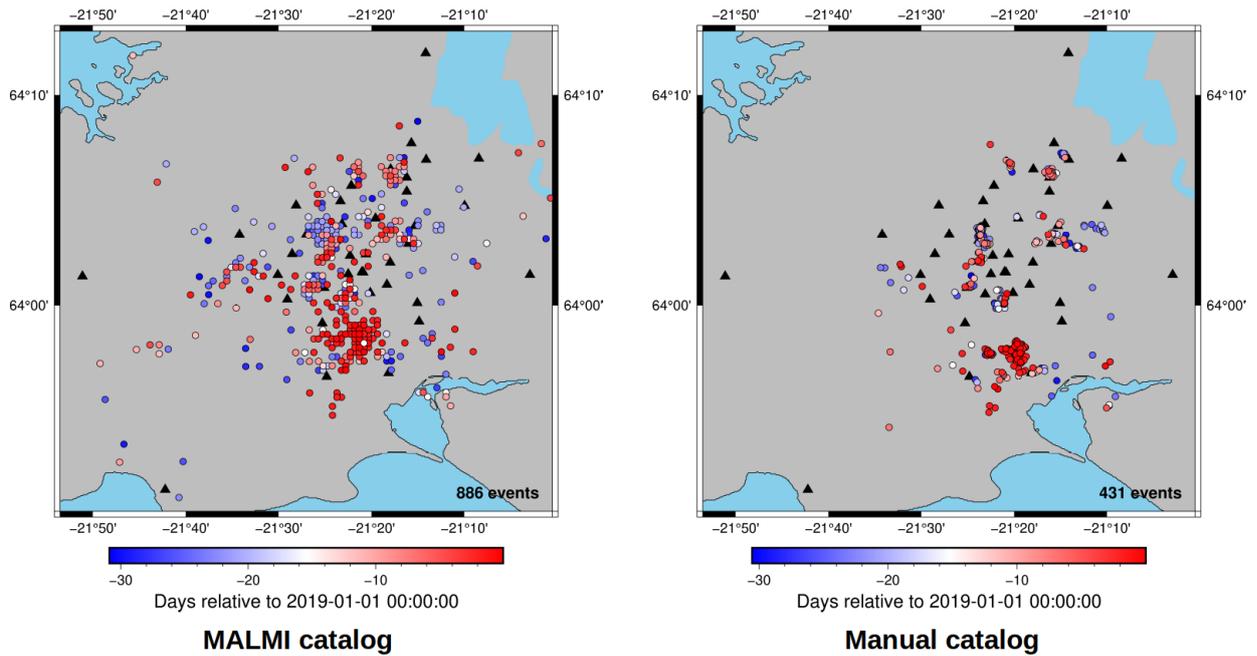


Figure 7. Spatial distribution of the earthquakes in the MALMI catalog (left) and the manual catalog (right). The earthquakes are color-coded according to times (in days) relative to 2019-01-01 00:00:00 (UTC).

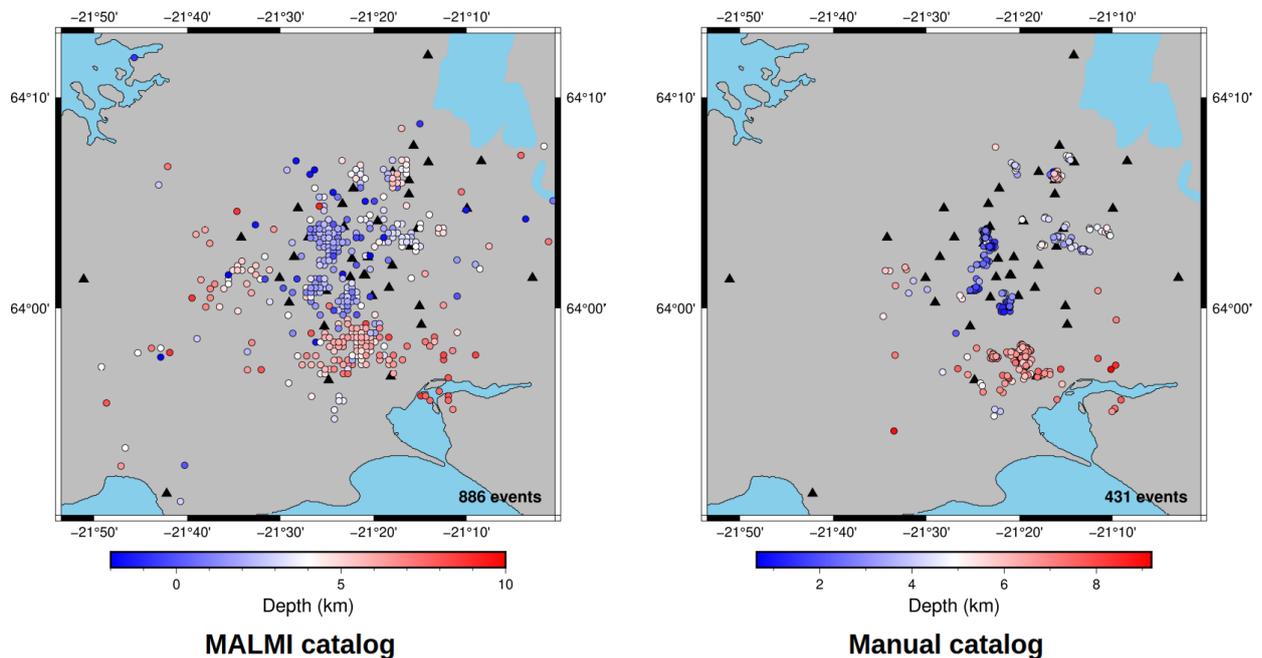


Figure 8. Spatial distribution of the earthquakes in the MALMI catalog (left) and the manual catalog (right). The earthquakes are color-coded according to the depth of earthquakes.

To further evaluate the performance of our proposed approach, we extract and analyze the migration profiles of two events selected from the obtained catalog. In detail, we chose one small event with a local magnitude of 0.37 (Figures 9-12) and one larger event with local magnitude of 2.3 (Figures 13-16). We compare MALMI with conventional migration techniques based on STA/LTA characteristic functions (Grigoli et al., 2018). For migration based on STA/LTA, the short time window and long time window are 0.1 seconds and 0.2 seconds respectively for the P-phase and 0.15 seconds and 0.3 seconds respectively for the S-phase. The filtering

frequency band for seismic data is set from 2 to 30 Hz. In comparison, MALMI does not require any input parameters related to time window and frequency band.

A migration data volume is automatically generated for each event data segment over the whole migration area. In general, a migration data volume consists of a 4-dimensional data set spanning 3D in space and 1D in time. However, as described in the previous section, the time range for a particular event is automatically determined by the array event scan module and only the maximum stacked coherence is kept for each spatial grid point during the determined time range. Therefore, the migration data volume shrinks to 3D. Figures 10-11 and Figures 14-15 show the profiles along different directions of the generated migration volume for the small and the larger event, respectively. Figure 12 and Figure 16 show the isosurfaces along 80% and 50% of the maximum stacked coherence in the migration volume for the M_L 0.37 and M_L 2.3 event, respectively. The isosurfaces can be used to depict the uncertainties of the event location and are indicators of the energy focusing during the migration process.

As observed in Figure 9 and Figure 13, the phase probabilities provide a better estimation of the arrivals for both P- and S-phases compared to the STA/LTA. The estimates of STA/LTA for P-phase arrivals are systematically later than the true arrival times. In addition, the estimates of STA/LTA for S-phase arrivals are often problematic and sometimes show elevated energy over the entire trace especially for smaller seismic events. The migration profiles and the isosurfaces in the migration volume (Figures 10-12) of the smaller event (M_L 0.37) highlight how MALMI can reach a much higher imaging resolution for locating microseismic events of very low magnitude ($M_L < 1$) compared to the conventional migration techniques. For seismic events with larger magnitude ($M_L > 2$), both MALMI and conventional migration techniques produce clear and high-resolution coherence images (Figures 14-16).

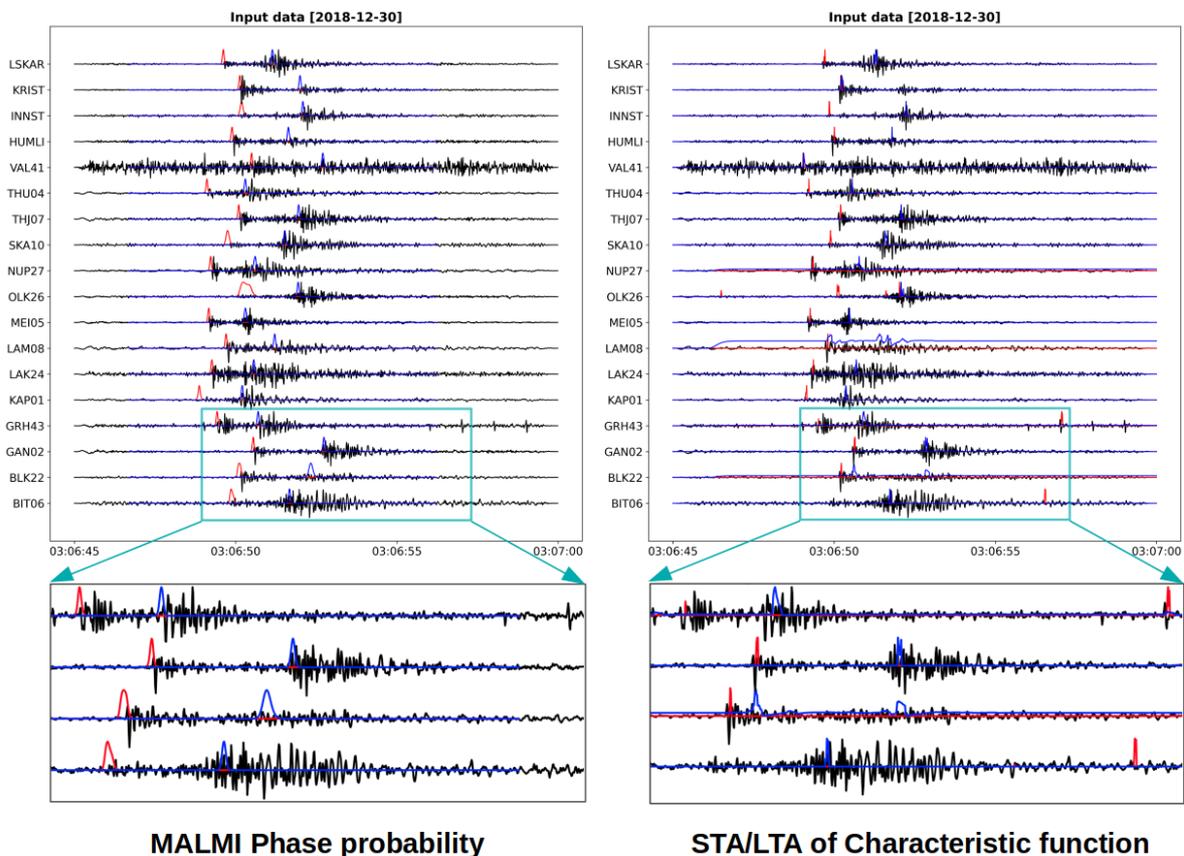


Figure 9. Z component seismic data for the M_L 0.37 event recorded by stations within the COSEISMIQ network. In the left panel the seismic data are overlaid with P- (red) and S-phase (blue) probabilities generated by MALMI, and in the right panel data are overlaid with STA/LTA calculated from vertical component data (red) and horizontal component data (blue) in the conventional migration method.

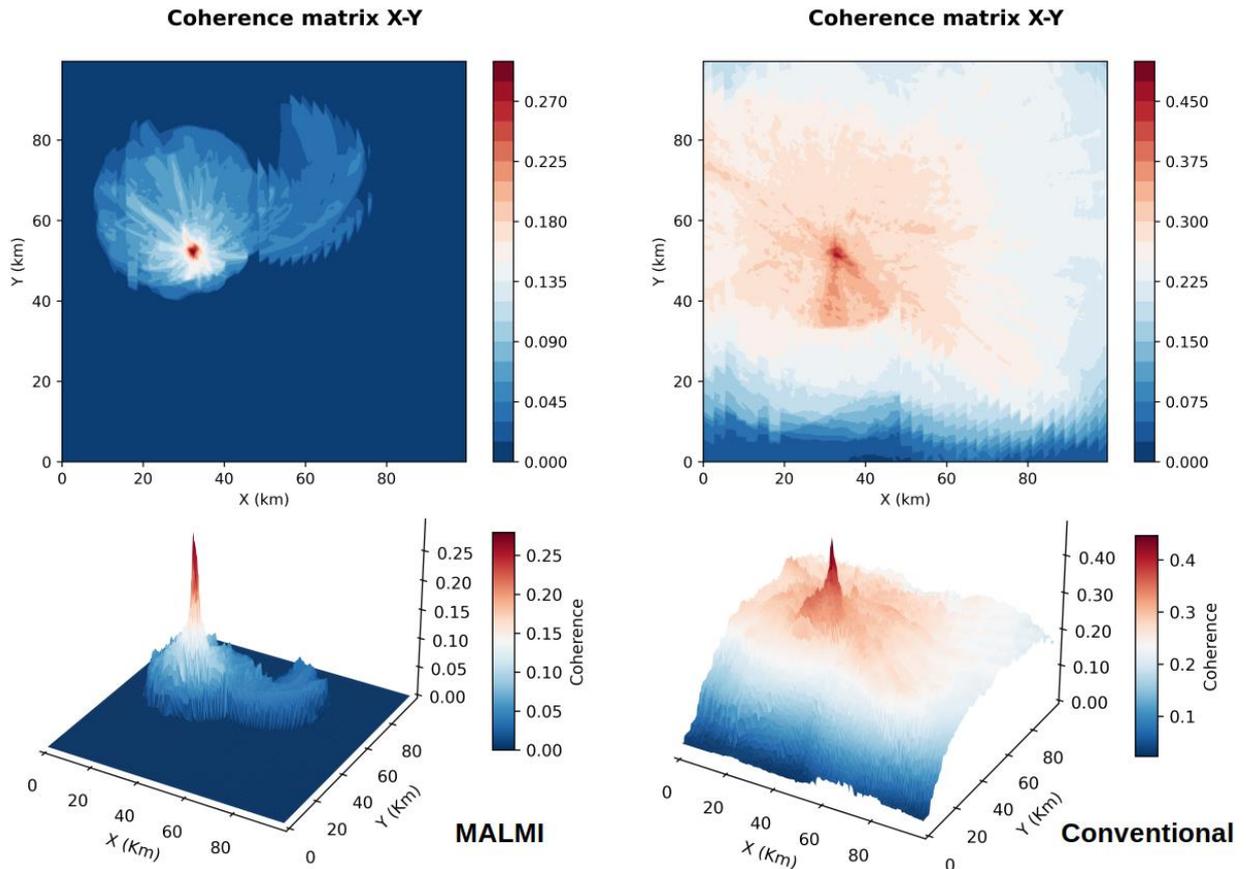


Figure 10. Migration profiles (north-east profile) projected along the vertical direction for the MALMI (left) and conventional migration (right) for the M_L 0.37 event. For each method, the top panel shows a plane view, whereas the bottom panel shows a 3D surface map.

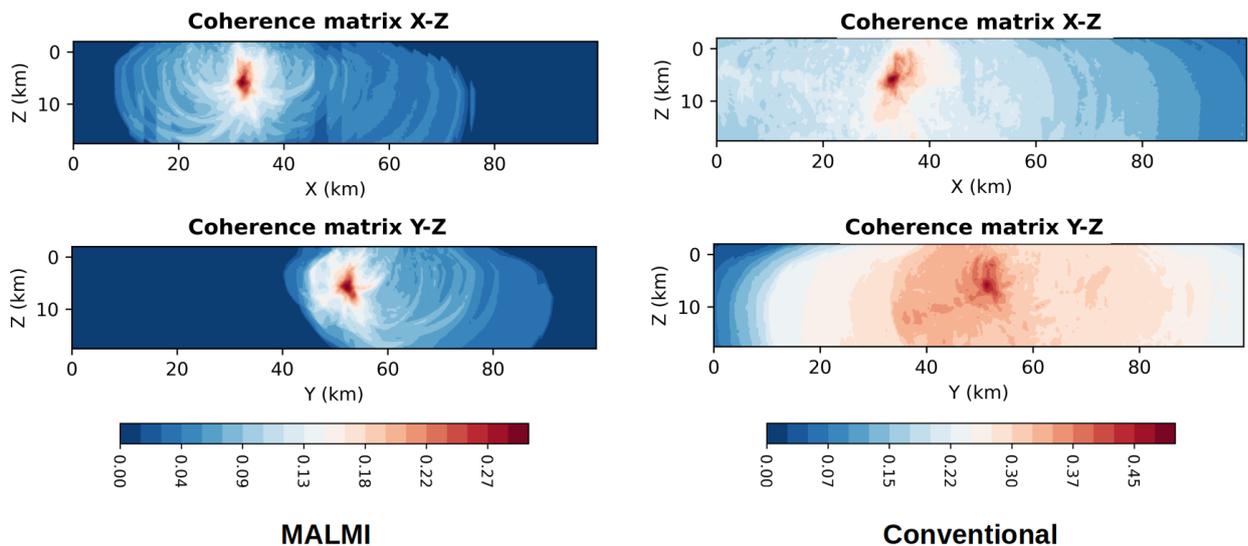


Figure 11. Migration profiles projected along the horizontal directions for the MALMI (left) and conventional migration (right) for the M_L 0.37 event.

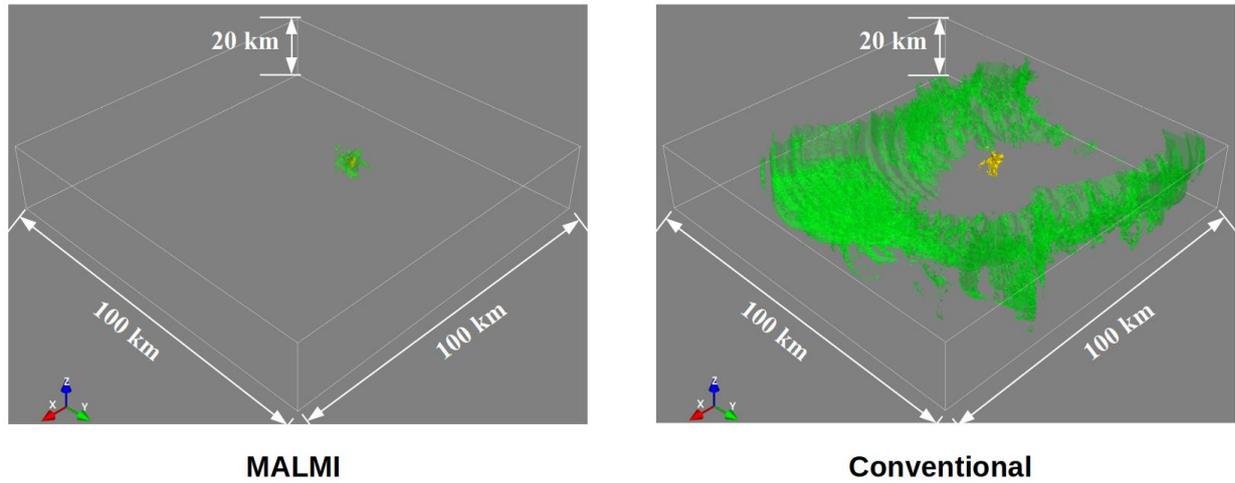


Figure 12. Isosurfaces in the migration volume for the M_L 0.37 event. The green isosurface represents the coherence surface of 50% of the maximum coherence, while the yellow isosurface represents the coherence surface of 80% of the maximum coherence.

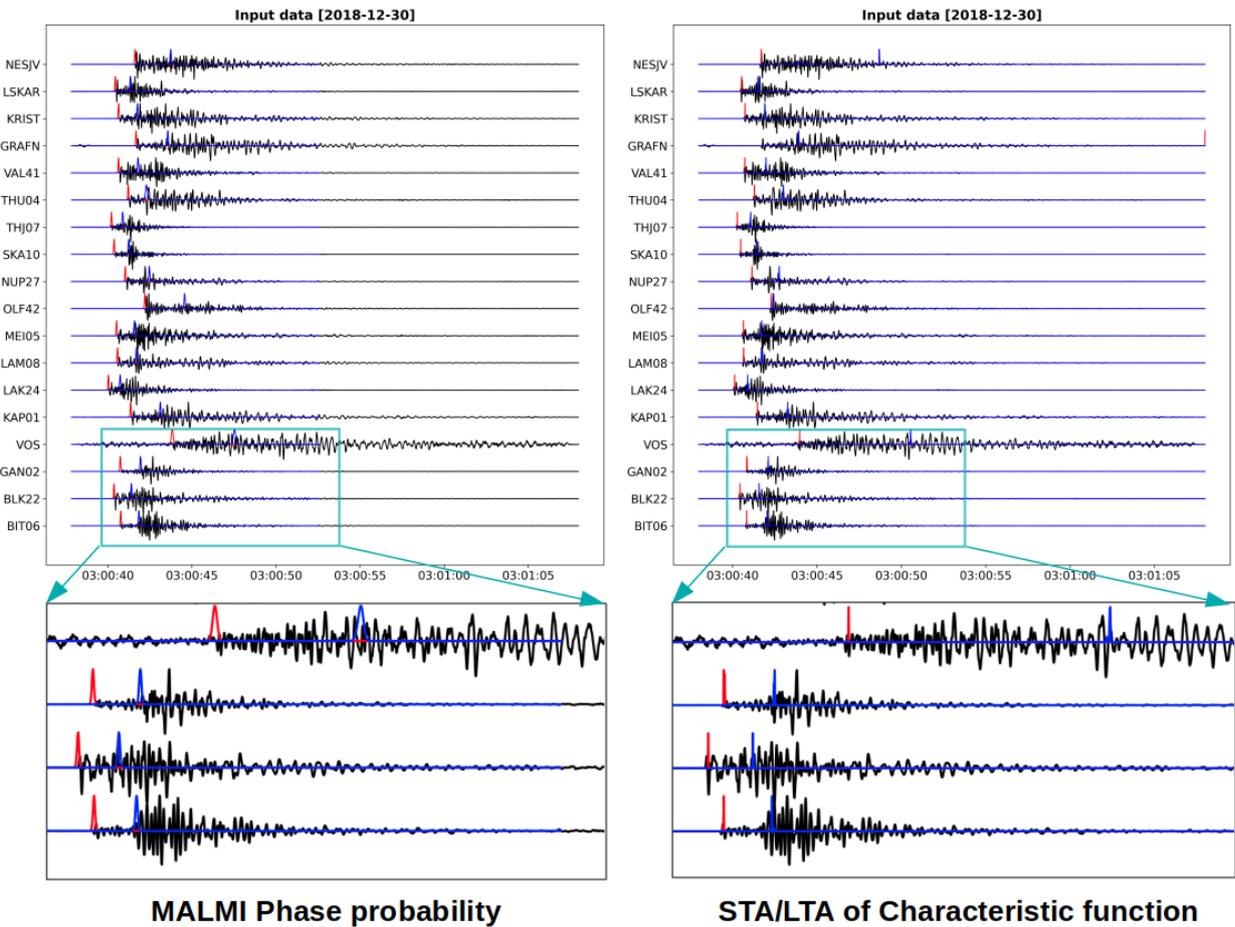


Figure 13. Z component seismic data for the M_L 2.3 event recorded by stations within the COSEISMIQ network. In the left panel the seismic data are overlaid with P- (red) and S-phase (blue) probabilities generated by MALMI, and in the right panel data are overlaid with STA/LTA calculated from vertical component data (red) and horizontal component data (blue) in the conventional migration method.

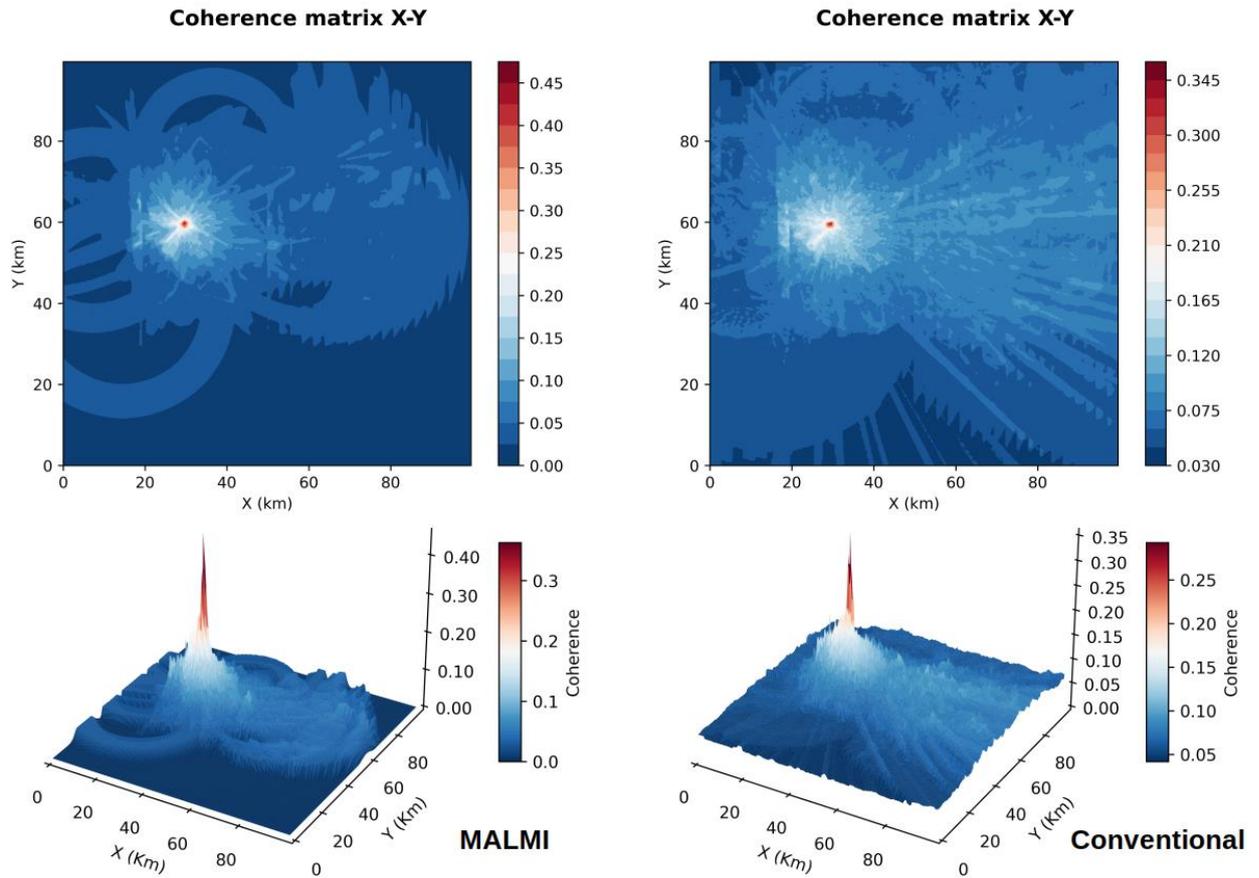


Figure 14. Migration profiles (north-east profile) projected along the vertical direction for the MALMI (left) and conventional migration (right) for the M_L 2.3 event. For each method, the top panel shows a plane view, whereas the bottom panel shows a 3D surface map.

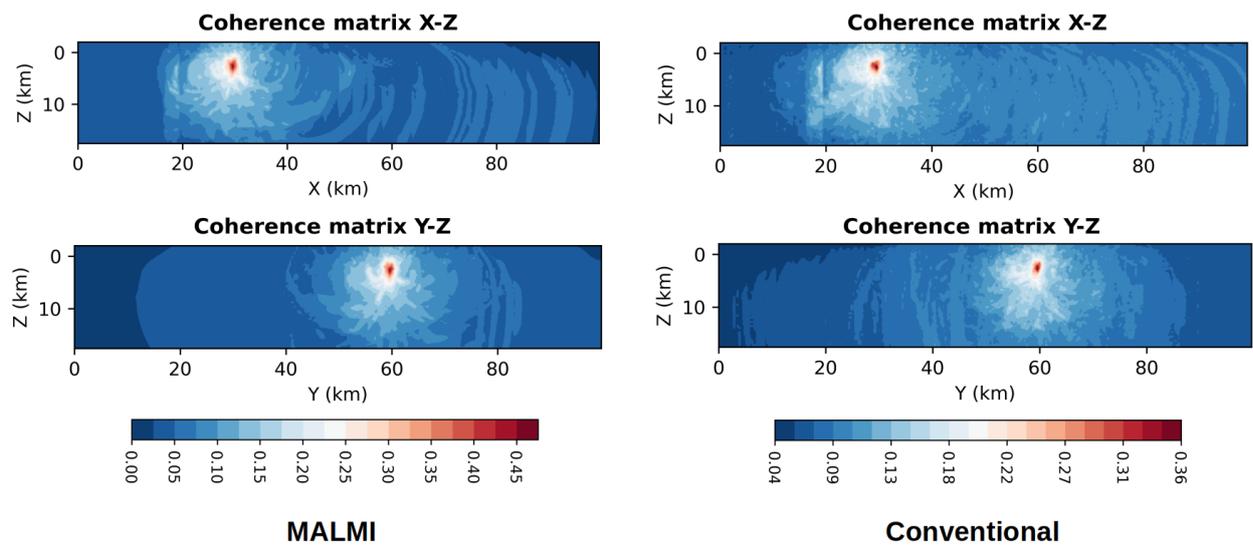


Figure 15. Migration profiles projected along the horizontal directions for the MALMI (left) and conventional migration (right) for the M_L 2.3 event.

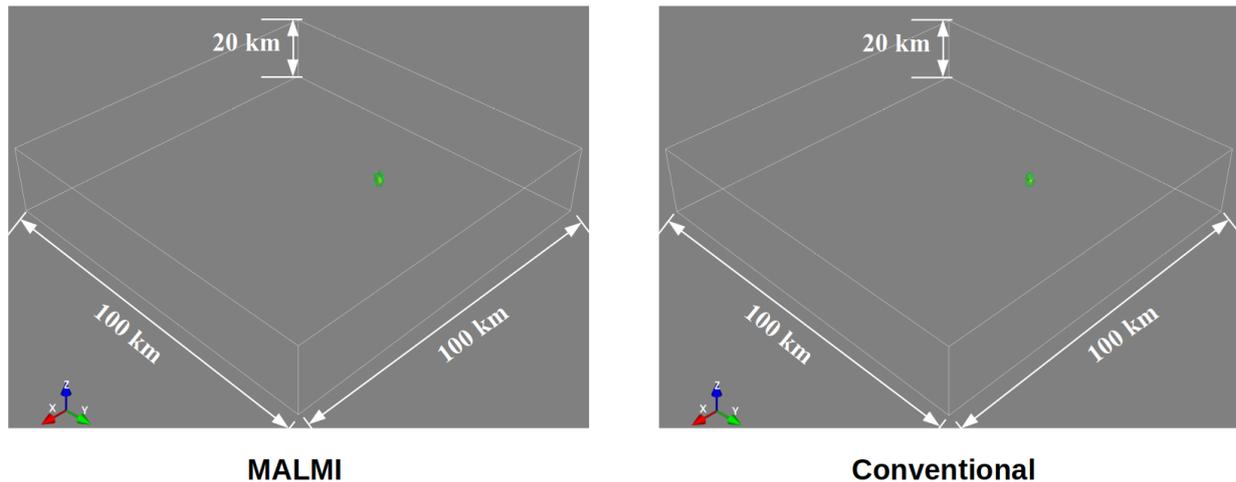


Figure 16. Isosurfaces in the migration volume for the M_L 2.3 event. The green isosurface represents the coherence surface of 50% of the maximum coherency, while the yellow isosurface represents the coherence surface of 80% of the maximum coherency.

4. Conclusions and Outlook

In this deliverable, we presented a new integrated framework to automatically detect and locate earthquakes from continuous seismic data. This framework integrates ML and migration techniques and is data-driven and suitable for real-time monitoring as it is computationally efficient. The proposed framework is able to detect earthquakes of smaller magnitudes, thus allows us to lower the magnitude of completeness and obtain more complete earthquake for further advanced seismological analysis, such as b-values and tomography studies as well as spatial-temporal evolution analysis of earthquakes. The latter bear critical knowledge to understand fluid-flow processes and possible hydraulic connections during EGS stimulations. Finally, the automated nature and the efficient processing ability make it capable of providing earthquake catalogs in a near real-time manner and thus contribute to harvest the geothermal energy more efficiently and more safely in EGS contexts.

The developed framework is freely available at <https://github.com/speedshi/MALMI>. The framework involves two independent software packages: EQTransformer and LOKI. These two packages have been substantially modified from the original version to be integrated into the proposed framework. The new version of these two packages developed under this deliverable can be accessed at <https://github.com/speedshi/EQTransformer> and <https://github.com/speedshi/LOKI>. A publication describing the proposed approach is in preparation. We will apply this framework to the EGS stimulation data at Utah FORGE once the data set is acquired this coming December 2021, and further assess the location performance and real-time processing ability.

We will continue to develop and integrate the proposed framework by integrating more ML models and improving the current efficiency and precision of the array event scan module. At the same time, we are training a convolutional neural network to identify seismic events and simultaneously retrieve their locations and origin times from a 4D or 3D migration volume. With such a trained ML model embedded in the framework, we will be able to provide to the EGS community a robust and efficient tool for microseismicity characterization.

References

- Allen, R. (1982). Automatic phase pickers: Their present use and future prospects. *Bulletin of the Seismological Society of America*, 72(6B), S225-S242.
- 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.
- Grigoli, F., Scarabello, L., Böse, M., Weber, B., Wiemer, S., & Clinton, J. F. (2018). Pick- and waveform-based techniques for real-time detection of induced seismicity. *Geophysical Journal International*, 213(2), 868-884.
- Mousavi, S. M., Ellsworth, W. L., Zhu, W., Chuang, L. Y., & Beroza, G. C. (2020). Earthquake transformer—an attentive deep-learning model for simultaneous earthquake detection and phase picking. *Nature communications*, 11(1), 1-12.
- Rossi, C., Grigoli, F., Cesca, S., Heimann, S., Gasperini, P., Hjörleifsdóttir, V., et al. (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.
- Shi, P., Angus, D., Rost, S., Nowacki, A., & Yuan, S. (2019a). Automated seismic waveform location using multichannel coherency migration (MCM)—I: theory. *Geophysical Journal International*, 216(3), 1842-1866.
- Shi, P., Nowacki, A., Rost, S., & Angus, D. (2019b). Automated seismic waveform location using Multichannel Coherency Migration (MCM)—II. Application to induced and volcano-tectonic seismicity. *Geophysical Journal International*, 216(3), 1608-1632.
- Withers, M., Aster, R., Young, C., Beiriger, J., Harris, M., Moore, S., & Trujillo, J. (1998). A comparison of select trigger algorithms for automated global seismic phase and event detection. *Bulletin of the Seismological Society of America*, 88(1), 95-106.