Deliverable

Deliverable 2.3 – Integrating DAS and geophone chain in real-time monitoring systems

Report information

Work package	WP2 Innovation in real-time monitoring and big-data analysis for EGS
Lead	ETH, LBNL, GES
Authors	Federica Lanza, Katinka Tuinstra, Juan Porras, Francesco Grigoli, Donald Vasco, Ben Dyer, Dimitrios Karvounis
Reviewers	Claudia Finger
Approval	Stefan Wiemer
Status	/¥š¨
Dissemination level	Internal
Will the data supporting this document be made open access?	Yes
If No Open Access, provide reasons	-
Delivery deadline	30.11.2022
Submission date	05.12.2022
Intranet path	[DOCUMENTS/DELIVERABLES/Deliverable2.3.docx]

Table of Contents

Summary

In this deliverable we report on new microseismic monitoring methods that make use of both downhole geophone chains and emerging technologies, such as DAS (Distributed Acoustic Sensing). In the processing of microseismic data the first step is the detection of seismic events, which is then followed by the determination of the hypocenter's parameters of the recorded seismicity. Here, first we introduce a semblance-based event detection method that fully exploits the characteristics of the DAS data and promises real-time applicability potential. We then propose three event location workflows, two of them also make use of DAS data but different network configuration scenarios (one or multiple wells); whereas the third location algorithm is tailored to be used with downhole geophone arrays and efficiently exploit the full waveform field. The detection and location workflows presented here aim at advancing the current state-of-the-art of traditional microseismic monitoring systems by pushing the new developments towards real-time performances, whose challenges are also discussed.

This deliverable constitutes a first report and summary of the current work. More detailed information of the methods is available through a published master thesis (Porras, 2022), two submitted or in preparation publications (Tuinstra et al., submitted to GJI; Vasco et al., in prep.) and a code repository. In the future, we plan to further improve the real-time performances of the presented algorithms, as well as to provide an end-to-end (from waveform to catalog) workflow in SeisComP, an open-access seismological software for real-time data acquisition and processing, developed by the GEOFON Program at Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences and gempa GmbH.

1. Introduction

In several industries from petroleum to geothermal industries, the goal of microseismic data processing is to transform continuous wavefield observations into precise and accurate estimates of event locations, magnitudes and other source characteristics. Knowledge of such parameters is key for surveillance and diagnostic characterization of induced seismicity both in the context of development of the reservoir connectivity and for safety protocols and risk governance.

Typical microseismic monitoring systems at deep geothermal projects use both a permanent network of sensors deployed at the surface and/or in shallow boreholes and deep downhole geophone arrays as close as possible to the stimulation region. Recently, new types of sensors (e.g., Distributed Acoustic Sensing - DAS) are becoming available and together with new automated processing techniques, seismic processing is rapidly changing, moving towards near real-time performances of an increased data volume recorded by multiple types of sensors.

In this deliverable we present the DEEP contributions to such seismic analysis revolution and introduce new methodologies tailored to increase effectiveness and real-time capabilities of the processing steps for event detection, classification and location for downhole geophones and DAS data. In particular, we introduce: 1) a semblance-based method for fast detection of microseismicity in DAS record streams (section 2.1; Porras, 2022), 2) a rapid earthquake location algorithm for geophone chains based on the full waveform travel-times (Vasco et al., in prep.); and 3) a modified version of HADES (Grigoli et al., 2021), a clustered event relocation algorithm based on a distance geometry problem. The new HADES can be used with DAS data and when azimuthal coverage is limited, e.g., when only one monitoring well is available (Tuinstra et al., submitted to GJI). For cases when more than one well is used for monitoring, we show that the Hamiltonian Monte Carlo algorithm is more suitable.

Here, we report the initial results of each approach on datasets collected during the pilot stimulation (in 2019) and most recent injection experiment (in 2022) at the Utah FORGE (Frontier Observatory for Research in Geothermal Energy) site. As a proof of concept for the methods, we also used synthetic datasets and data collected at the Bedretto Underground Laboratory for Geonergy and Geosciences (Hertrich et al. 2021), in Switzerland. We then conclude with current challenges and some outlook for future development of real-time microseismic monitoring systems.

2.   Detection workflows

2.1.   Semblance-based microseismic event detector for DAS

Event detection is a critical aspect of the monitoring workflow and it is especially challenging when dealing with induced seismicity datasets as they consist of numerous weak overlapping events (Grigoli et al 2018b). These challenges have exponentially increased in the last decade when seismology entered the big data era and modern instruments, such as fiber optic sensors, contribute to generating extremely large datasets to analyse. Fiber-optic cables (e.g., conventional telecommunication or built-for-purpose cables) can be turned into a dense array of geophones that samples seismic wavefields continuously for several kilometers. Distributed Acoustic Sensing (DAS) is particularly interesting for microseismic monitoring of geothermal systems since it does not have the same temperature limitations as standard electronic equipment. The sensing fiber can therefore be installed at high-temperature reservoir conditions and in the same well that is being stimulated. Because of these advantages, the distance between the detecting sensor and the induced seismicity can be minimized, maximizing the detection capability. Typical DAS acquisition samples the wavefield at about 1 m spacing and sampling frequencies of 1 kHz or higher. Unfortunately, standard seismological techniques (e.g., STA/LTA) are not capable of exploiting this high spatial density of sensors, hence they are ineffective in processing this kind of data.

Here we propose a semblance-based seismic event detection method that instead 1) fully exploits the high spatial sampling of DAS data, 2) it is sensitive with a low missed event rate and, at the same time, robust with a low false event rate, and 3) it is fast enough to be potentially applied in near-real time. We apply our detector to a DAS microseismic monitoring dataset collected in 2019 at the FORGE geothermal experiment site in Utah, USA, as part of a pilot stimulation. At FORGE, the presence, in addition to the DAS fiber, of a co-located 12-level geophone string in the monitoring well, also allowed for a detailed comparison of performances between the two sensor types and traditional techniques (see also Deliverable 2.2). From an initial spectral analysis, it is revealed that DAS has a lower sensitivity and a lower signal to noise ratio (SNR) than the geophones. Thus, a pre-processing step, aimed at denoising the DAS raw data should be performed before the detection process. As shown below, such denoising techniques help to increase the SNR of DAS to levels as good as the geophones.

2.1.1.   Denoising

We developed an automatic workflow to increase the DAS signal to noise ratio. The workflow is calibrated on the FORGE dataset collected during the 2019 pilot stimulation. The workflow starts with the removal of the mean and linear trends of each trace of the array. The linear and mean filtering step is essential, as it removes most of the interrogator imprint (laser drift, reference loop vibration, etc.). We then define a bandpass filter between 10 to 300 Hz as the energy signal is generally concentrated in this frequency range. Further coherent noise is removed by applying a frequency-wavenumber (FK) filter along the k=0 wavenumber. Figure 2.1 shows an example of a low magnitude event (maximum amplitude 44 nε/s) before and after denoising, with its corresponding FK spectrum. The energy signal of the raw data is not clearly visible either in time domain and in the FK spectrum. However, after denoising, the coherent k=0 noise is attenuated as well as the high frequency random noise, allowing the spatial coherency of the filtered wavefield to be visible in both time and frequency domains.
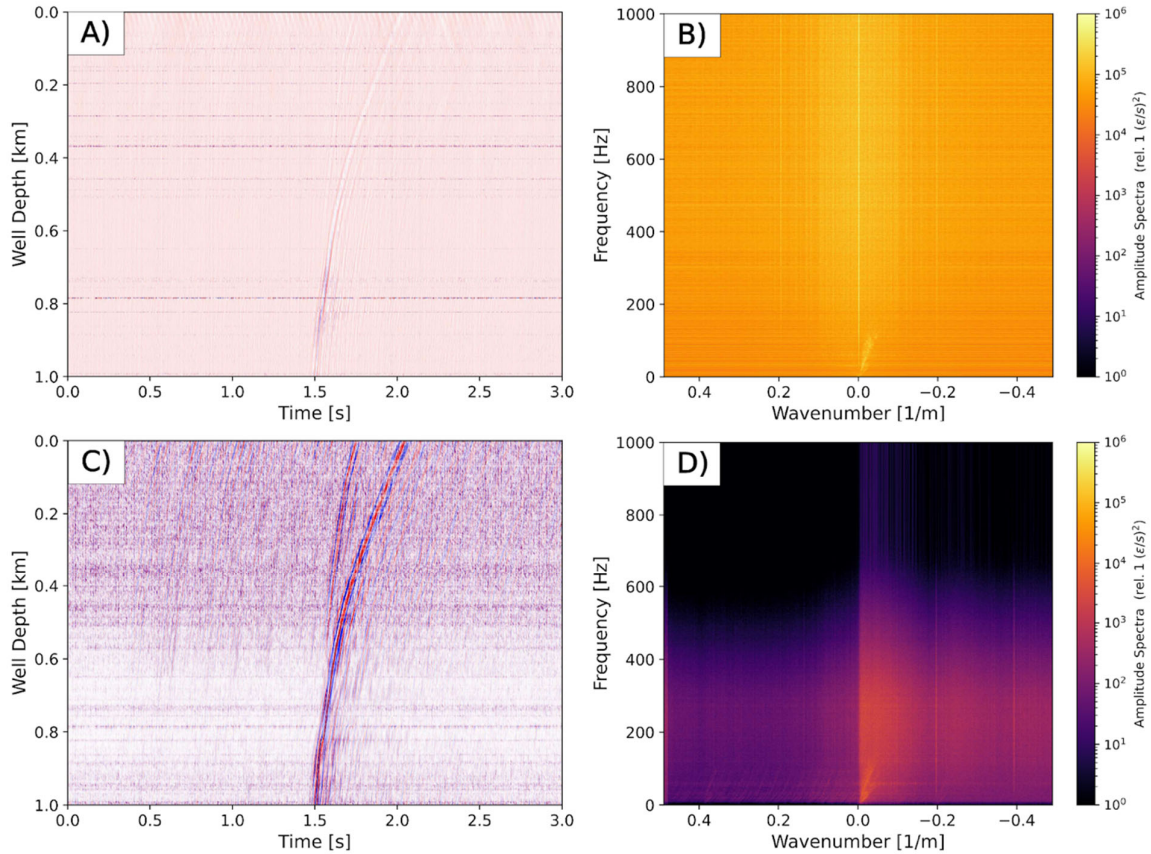
Figure 2.1. Example of the denoising workflow of DAS data A) Raw data, B) FK spectrum of raw data. C) Filtered data and D) FK spectrum of filtered data.

2.1.2.   Methodology

In the following, we describe the core principles of the proposed semblance-based DAS detector. For further details, we refer the readers to Porras, 2022. Borrowing from reflection seismology and the theory of the semblance function (Neidell and Taner, 1971), the method detects seismic events by measuring the waveform coherence along several trial hyperbolic trajectories (Figure 2.2). The hyperbolic trajectories that follow the seismic wavefield will have high semblance coefficients (close to 1) indicating summation of the signal (event); whereas coefficients of nearly zero will be equivalent to the summation of random noise (no event).
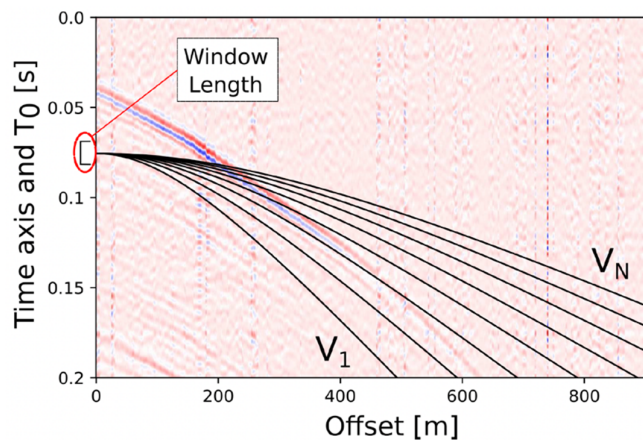
Figure 2.2. Sketch of the computation of the Semblance panel. Trial hyperbolic trajectories along which the measurement of the coherency takes place. Curvatures depend on the velocity value (V_1-V_N).

As seen in Figure 2.2, the geometry of the trial hyperbolas strongly depends on the velocity. To overcome such dependency, as the velocity is not known, here for each trial hyperbola, the curvature and position of the vertex is changed according to equation 1:

$$\frac{y^2}{b^2} - \frac{(x+a)^2}{c^2} = 1 \qquad [1]$$

Where the coefficients a, b and c represent the lateral search, time search and curvature coefficient, respectively. The method returns a time series of coherence values and, if these values are higher than a determined threshold, a seismic event is declared. In detail, for the discrimination of real versus false events, the algorithm is designed to implement a dynamic threshold function, defined as the mean of the coherency vector plus 2.5 standard deviations in a time window defined as the 40% of the total duration of the input data. After the declaration of a real event, the discrimination function is halted for 1 second to make sure S-waves arrivals do not trigger a separate event.

Initial tests showed that the best performances of the algorithm are obtained when the derivative of the STA/LTA stacking function is used as input data, rather than the raw filtered waveforms (Figure 2.3a; Table 1). This is because the derivative of the STA/LTA traces, which are non-negative stacking functions represented by "gaussian-shaped" pulses in correspondence of the P and S waves arrivals, become zero-mean traces with uniform polarities, and thus they allow to remove the effect of the radiation pattern of the seismic source (Figure 2.3b).
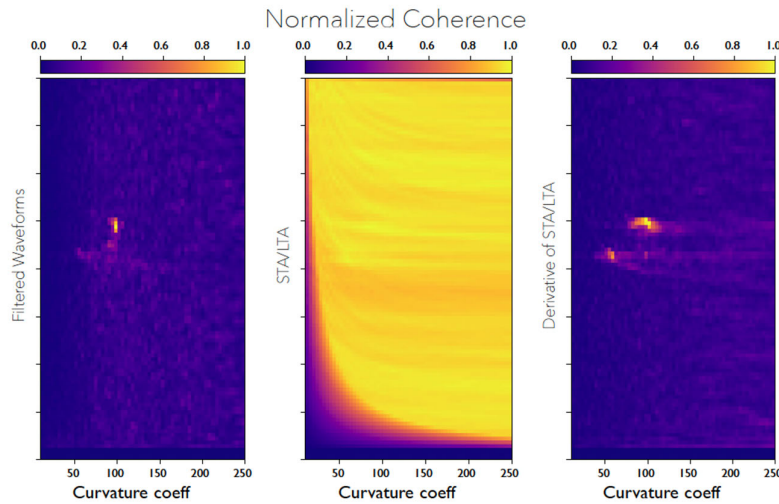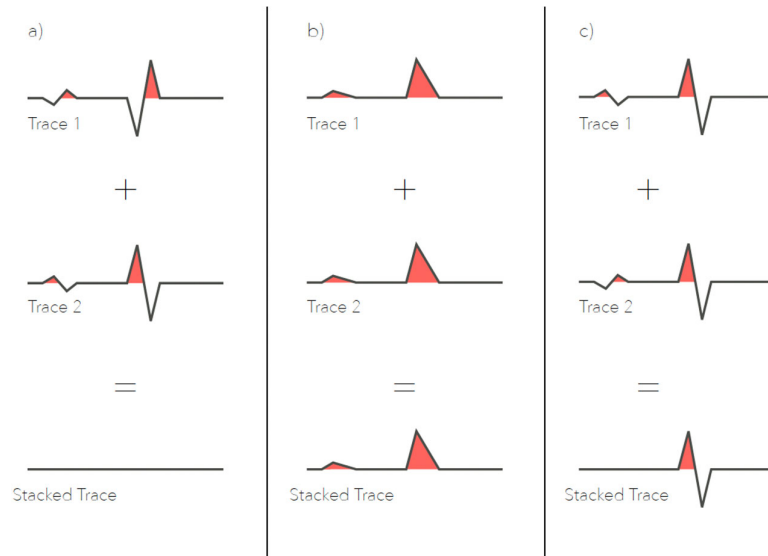
Figure 2.3. Top: Coherence images for a low SNR microseismic event and for different processed input data: (left panel) Filtered waveforms, (central panel) STA/LTA (similar results are obtained for the envelope) and (right panel) Derivative of the STA/LTA. Bottom: A) Stacked traces of opposite polarity are attenuated. B) Stacked STA/LTA traces is not efficient in noise attenuation. C) Stacked traces after the application of the STA/LTA derivative. The energy signal is well stacked since polarities are homogenized.
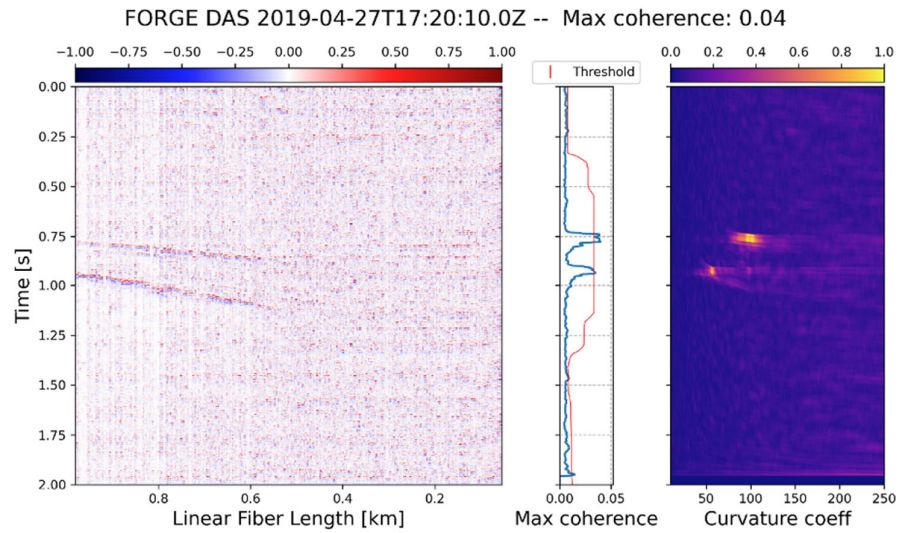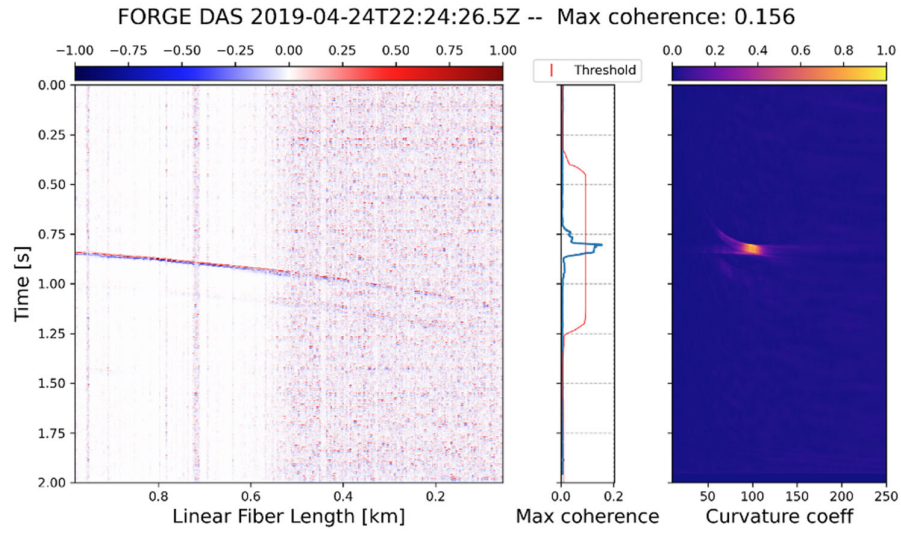
Apart from the data (i.e., derivative of the STA/LTA traces), other inputs required by the detector include:1) the width of the time window for which the waveform coherency measurement takes place, 2) the range of curvature coefficients which defines the family of hyperbolas to be used, 3) the sample step, 4) the curvature step and 5) the width of the lateral search. The latter is optional and currently not intended to be used for real-time applications, as it considerably slows-down the algorithm. Similarly, to reduce memory requirements and processing time, we recommend to downsample, albeit without altering or losing useful information contained in the data.

2.1.3.   Results: Application to the FORGE stimulation of April 2019

We applied the detector to a test sample of 110 events recorded by the DAS cable installed at FORGE site (well 78-32) during the 2019 stimulation. The computing time (serial code) is approximately 2 seconds per each second of data measured in a laptop with an Intel quad-core i7 processor and 16 GB of Random Access Memory (RAM). These numbers are based on downsampled data and indicate the method has the potential for real-time applications with the adequate optimization. Table 2 summarizes the performance of the algorithm when using three different data inputs: 1) raw data; 2) filtered data; and the 3) the derivative of the STA/LTA. Figures 2.4a, b, and c, show the results of the detection algorithm for three example events. The first event (Figure 2.4a) is a perforation shot, displaying high amplitude with a clear wavefield and maximum in the coherency panel. The second event (Figure 2.4b) is a weak (low SNR) microseismic event, but it is still well detected by the algorithm. The last event is extremely noisy (Figure 2.4c), and the algorithm is not able to detect any maximum and it is therefore classified as false detection. Overall, when we apply the derivative of the STA/LTA to the input data, the rate of detected events is equal to 97% (107 events of the 110 events considered as the ground truth). A visual inspection of the missing 3 events shows that the events are indeed not visible in time nor in frequency domains in the DAS data, and can be slightly identified in the geophone array, which has higher sensitivity than DAS. Therefore, those events are below the SNR of the denoised DAS data and not detectable by the algorithm. We conclude that this new semblance-based DAS detector is able to correctly identify all the events that are effectively visible in the data.

Table 1. Results of the algorithm application with different versions of the input data.

Data	Events	Detected events (and %)	Missed events (and %)
Raw	110	22(20%)	88 (80%)
Filtered	110	72 (65%)	38 (35%)
STA/LTA'	110	107 (97%)	3 (3%)

FORGE DAS 2019-04-24T22:24:26.5Z -- Max coherence: 0.156

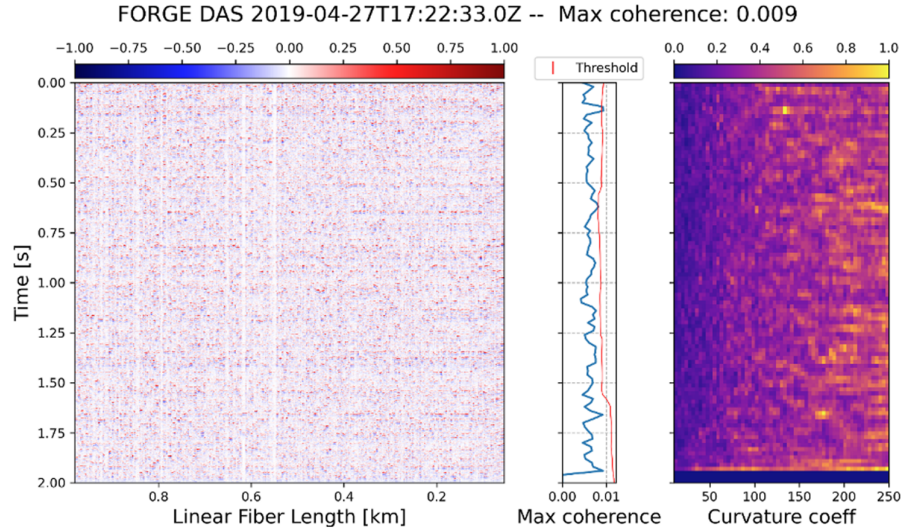FORGE DAS 2019-04-27T17:20:10.0Z -- Max coherence: 0.04

Figure 2.4. Example of the semblance-based detector performances on three test events of different signal-to-noise ratio: a) Perforation shot, characterized by a well-defined P-wave arrival at ~0.8 s. On the right, the coherence panel shows a well-defined maximum in correspondence of the event. b) Microseismic event, with the P-wave arrival at ~0.76 s and the S-wave at ~0.9 s. On the right, the coherence panel shows a well-defined maximum in correspondence of P and S phases, c) Noise, the coherency panel that does not show any well-defined maxima. At ~0.2 s the coherency vector overcomes the threshold but it is classified as a false detection

3.   Location workflows

3.1.   Modified HADES: Relative Cluster Location algorithm for DAS data

The next step after event detection is the determination of the event hypocenter parameters. By knowing the event location, we can derive further information such as source mechanisms and slip characteristics, thus allowing for a better interpretation (e.g., in terms of source, spatial and temporal migration patterns) of the recorded microseismicity. In underground borehole settings, location constraints are imposed by the limited availability of deep, nearby wellbores leading to an acquisition geometry that is rarely close to ideal. In many cases, only a single vertical observation well is available for monitoring.

We here propose a new methodology to locate clustered microseismic events with a fiber-optic cable deployed downhole, in non-ideal network configurations. The method builds on the work of Grigoli et al. (2021), that proposed an efficient method based on Distance Geometry Problem (DGP) solvers: HADES (earHquake locAtion via Distance gEometry Solvers) for locating clustered seismicity with sparse surface networks of one or two stations. The following paragraphs can be found also in the manuscript "Locating clustered seismicity using Distance Geometry Solvers: applications with sparse seismic networks and borehole Distributed Acoustic Sensing" by Tuinstra et al. submitted to GJI.

3.1.1.   Methodology

As described in Grigoli et al. (2021), HADES is a relative location method in which seismic events within a cluster are located with respect to the position of other seismic events (one or many). To find event locations in a relative frame, HADES uses the inter-event distance between two events computed from the double difference in arrival times of S- and P- phases (t_s and t_p) and measured at a seismic station (equation 2):

$$\|r_{e1e2}^{S}\| = k_v \left| \left(t_s^{e1} - t_p^{e1} \right) - \left(t_s^{e2} - t_p^{e2} \right) \right| \qquad [2]$$

Where r^s is the inter-event distance between event $e1$ and event $e2$, and $k_v = \frac{v_p v_s}{(v_p - v_s)}$. In Grigoli et al. (2021) a minimum of four master events with known locations is required to then find the absolute location and rotation of the events in the cluster. Any rotational error of the cluster is minimized in the horizontal (latitude-longitude) plane, whereas rotational

errors in depth remain unresolved. Advantage of this method lies in the ability to provide the correct shape of clustered seismicity, particularly in ill-posed event-receiver configurations. These are two conditions, clustered microseismicity and poor azimuthal coverage, that often characterize downhole monitoring settings. Thus, leveraging on HADES properties, we first modified the method to allow for the handling of large arrays, such as DAS, by considering multiple random combinations of DAS channels, rather than the original two station condition. In addition, to avoid the master event dependency, which can lead to improper uncertainty propagation in the solution, we reduced the number of required master events from four to one. Once a single master event is selected, it represents the start of the coordinate system, and four other events are then located relative to this single event to construct a relative coordinate system as shown in Figure 3.1.
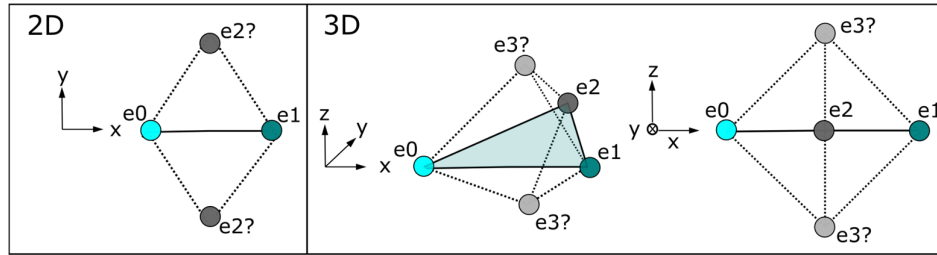

Figure 3.1. Construction of a relative coordinate frame to locate the events using only one master event. The left panel shows the 2D case: the master event is shown in cyan (e0) and another event in the event cloud is located with respect to this first event (e1). Then, by calculating the distances to e0 and e1, a third event is added to the system forming a plane (e2). Note that its azimuth with respect to the first plane will remain ambiguous. The left panel shows the 3D case: the fourth reference event (e3) will have an ambiguous location with respect to the first plane. Thus, we can say that these 4 events form an arbitrarily oriented coordinate system that can be used to locate in a relative sense the remaining events in the seismic cluster.

The correct rotation of the cluster is then found by using quaternions. Whereas complex numbers can describe rotations in two-dimensional space, quaternions describe rotations around arbitrary axes in higher-dimensional space, specifically using a four-dimensional representation. Quaternions are used in e.g., computer graphic simulations and robotics, as they can uniquely describe a rotation and are computationally more efficient than rotation matrices (Shoemake, 1985). The quaternion q is computed using the three components of an axis unit vector v and the rotation angle θ:

$$q(\theta) = [sin\left(\frac{\theta}{2}\right), \; v_0 cos\left(\frac{\theta}{2}\right), \; v_1 cos\left(\frac{\theta}{2}\right), \; v_2 cos\left(\frac{\theta}{2}\right)] \tag{3}$$

To orient a seismic cluster, the rotational origin of the seismic cluster is the chosen master event. The three rotation optimization axes are then a) a vertical axis from the origin pointing upwards to the surface, b) an orthogonal axis from the origin pointing to one selected station on the horizontally projected plane, and c) the cross product of the first two axes. The choice of station that governs axis b can be interchanged. The three unit vectors (v_a, v_b, v_c) are computed for these three axes. The optimal rotation angle for each of the three axes $\theta_1, \theta_2, \theta_3$ is the one that maximises the rectilinearity of the absolute distance between the receiver and the events, with respect to the differential travel-time $t_s - t_p$ at each receiver. Ideally the distribution of $t_s - t_p$ is linearly increasing with increasing distance to the receivers. A cluster in which the differential traveltime to distance ratio decreases is non-physical and thus inversely rotated. Non-linearity of the $t_s - t_p$ versus distance can be caused by rotation errors or large uncertainties in the traveltime picks. The rectilinearity measure is a form of dimensionality reduction of the traveltime and distance data, and is computed by Principal Component Analysis (PCA) (Jolliffe, 2002). The inverse of the rectilinearity can be used as a misfit function that has to be minimized. This misfit function prevents calculating synthetic traveltimes for e.g., a more conventional L_2 misfit between observed and computed traveltimes, bypassing the use of a potentially incorrect velocity model. The rectilinearity is obtained by taking a data matrix X with the sorted and demeaned distances r and traveltime differences D of all N events on the columns,

$$X = \begin{bmatrix} r_1 & D_1 \\ \vdots & \vdots \\ r_N & D_N \end{bmatrix} \tag{4}$$

From the transpose of X we can compute the covariance matrix:

$$C = cov(X^T)$$
[5]

The rectilinearity is the ratio between the maximum and minimum normalized eigenvalues (λ_{max} and λ_{min}) of the data covariance matrix:

$$R = \frac{\lambda_{max}}{\lambda_{min}}$$
[6]

A perfect linear relationship between differential traveltimes and distance to the receiver produces vanishing eigenvalues and infinite rectilinearity. The optimiser uses a brute force grid search over all θ, with a 5° resolution. However, in some angular situations, the solution to the rotation problem may be non-unique i.e., multiple rotations of the cluster will result in equivalent distance-traveltime distributions. This can occur in a situation with a single vertical borehole, where lateral rotations of events could result in equivalent traveltimes. Another possibility is an incorrect cluster shape due to large picking errors or velocity model estimations. In the latter situation, if the final automated output is still rotated, this can be fine-tuned using a manual rotation tool. With this tool, the user can set manual rotations around the three axes, and fine-tune up to an accuracy of 1° until the rectilinearity is maximized, and the traveltime of the events in the cluster aligns well visually with the distance to the station.

The required inputs for the modified HADES consist of (1) a minimum of one event with a known absolute location (the master event), and (2) P- and S- arrival time picks. The master event could be e.g., a check shot, or the largest event in the sequence that is also recorded by the surface array, and that can therefore be located with the lower uncertainty.

3.1.2.   Benchmark test on the 2019 Ridgecrest seismic sequence, California

We first performed a benchmark test of the new version of HADES using the Ridgecrest 2019 earthquake data. This is the same dataset used in Grigoli et al. (2021). As shown in Grigoli et al. (2021), they were not able to recover the shape of the cluster if less than 15 master events were used. Figure 3.2 shows our results after automated cluster location and rotation and an additional manual rotational as the final fine-tuning step. Although some events have a relatively large error, and some are even located above the surface, the L-shape of the cluster is now retained with only one master event and two stations at 50 km distance. The improvement is notable. Further details and comparison figures can be found in Tuinstra et al., submitted to GJI.
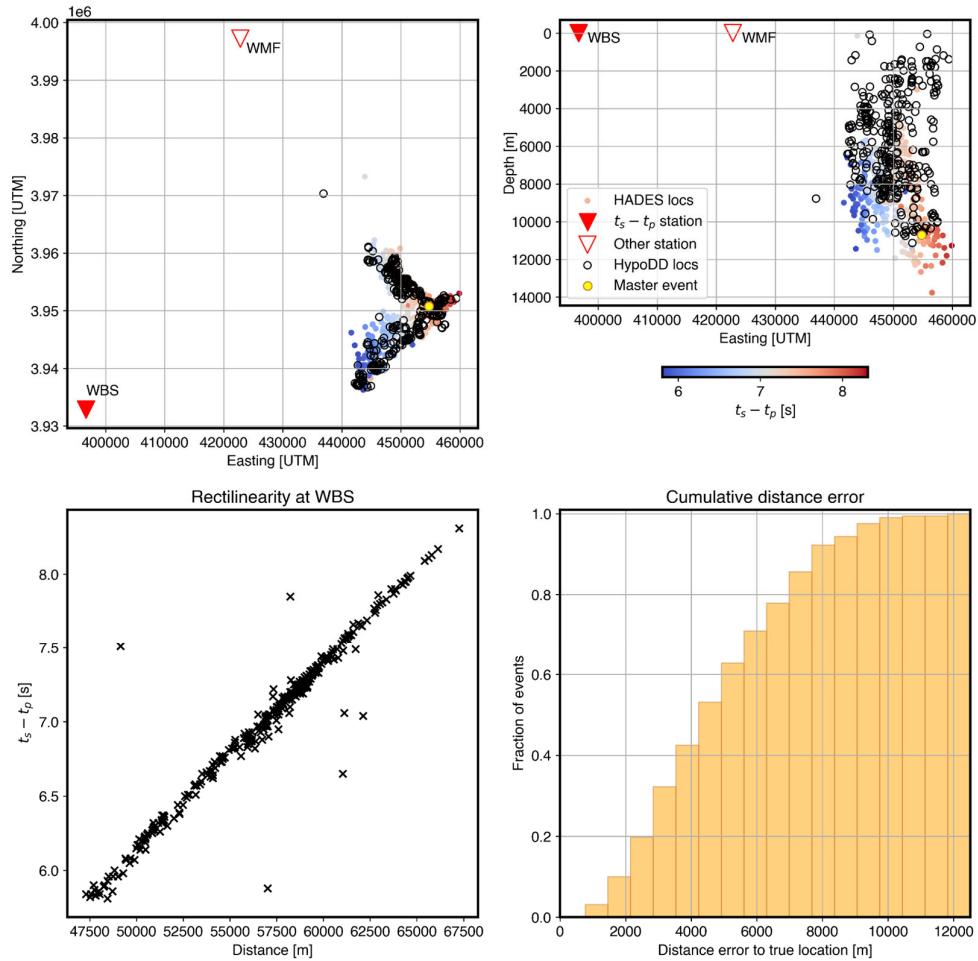
Figure 3.2. Benchmarking the single-master event version of HADES with the Ridgecrest aftershock sequence of 2019.

3.1.3.   Results: Application to the FORGE stimulation of April 2019

During the 2019 stimulation phase, Schlumberger published a catalogue with over 400 microseismic events across the three phases of stimulation in well 58-32. The events were located using the 12-level geophone string. Due to the challenging configuration between the monitoring well and the events, the location uncertainties are high and of the order of 400 m (Moore et al., 2019) The microseismicity at the toe of well 58-32 was also recorded by a co-located DAS fiber. Lellouch et al. (2020), detected with DAS a subset of 110 events in the time window of 27 April 2019, 5 p.m. (UTC) and 28 April 2019, 5:10 p.m. (UTC), of which they were able to locate 45 events. Of these 32 events were suitable for picking and locating with our updated version of HADES. The location process is conducted in 150 iterations with random channel pairs, with a velocity of $V_p = 6100 \ m/s$ and $V_s = 3200 \ m/s$ which reflects the average velocity along the ray paths taken from the 1D velocity model. The location spread in Figure 3.3 shows a circular pattern in all three planes. This is indicative of a high rotational uncertainty, in which the rotation optimiser finds a maximal rectilinearity in many directions. The boundaries of the cloud however coincide with the boundaries of the Schlumberger catalogue, implying that the cluster shape has the correct dimensions. More tests on synthetics and other seismicity clusters are reported in the submitted paper (Tuinstra et al.) as well as in D2.2.
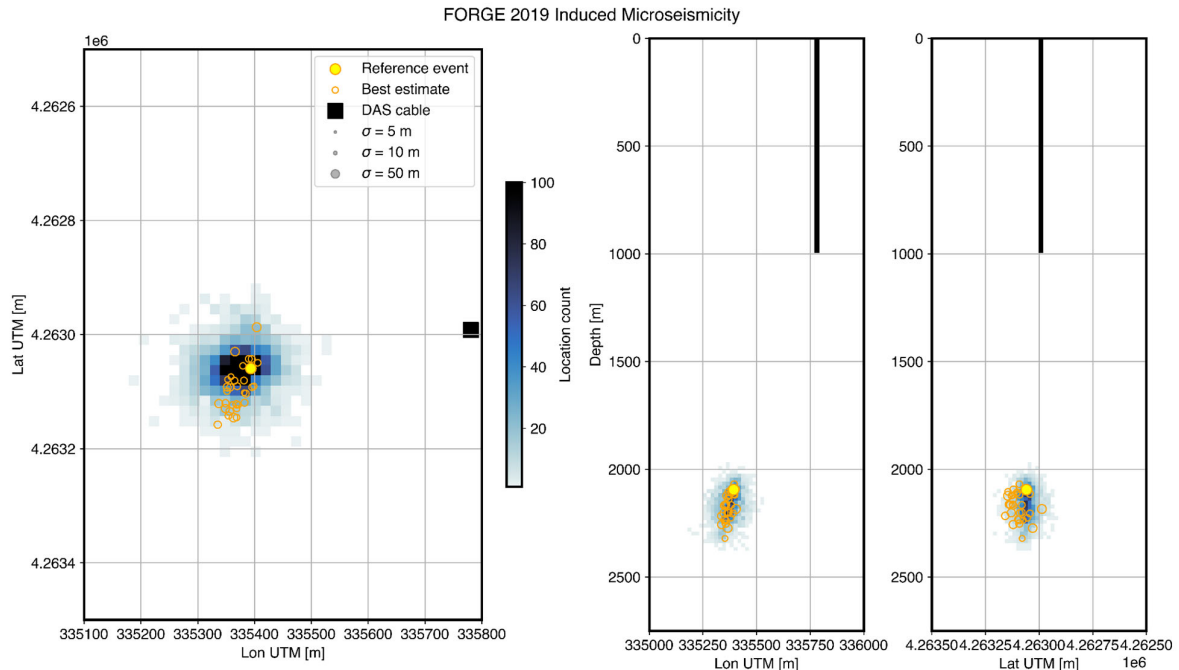
Figure 3.3. HADES results for the microseismicity recorded with the DAS in the 2019 FORGE stimulation in the horizontal (left), longitude-depth (top right) and latitude-depth view (bottom right). The location count histogram indicates the amount of times an event was located at a location in 150 iterations using different station combinations.

We conclude that the new version of HADES now has capabilities to reconstruct the shape and relative location of a seismicity cluster using only P- and S- wave arrival times, one a-priori located event and large arrays (large number of channels in the case of DAS). Thus, the method is able to overcome the azimuthal uncertainty inherent to the straight-cable geometry and downhole-deployment of DAS. Because the computation times of HADES range from seconds to tens of seconds per iteration depending on the number of events in the cluster, we show that the method has the potential, if included in the monitoring workflow, to track in time the cluster shape.

3.2.   Hamiltonian Monte Carlo (HMC) location for DAS

HADES's strength lies in its ability to relatively locate clustered events and reconstruct the correct cluster shape, however it is not designed to obtain absolute location for individual events. For individual locations and when more than one DAS-instrumented monitoring well is available, we tested a Hamiltonian Monte Carlo (HMC) sampler, based on Fichtner et al. (2018). The method samples the misfit between observed and simulated data by obtaining a series of random samples. This sampling is guided by energy conservation principles, which can be seen as a ball rolling through the misfit function landscape with a given starting velocity, and this guides the exploration of the misfit function in such a way that it can overcome being stuck in local minima and find the global minimum. The process is sped up by multiple parallel processes of simulated tempering and annealing to different 'temperatures', which means that the misfit function is being expanded and contracted for more efficient misfit exploration of the landscape and escaping local minima. The final outcome is a posterior probability distribution for the wave velocities (V_p and V_s), as well as the source coordinates (x,y,z). The algorithm takes as input the station locations, P and S-wave first arrival picks, and a homogeneous velocity model.

We first applied the method to a set of synthetic events calculated according to the Utah FORGE geothermal laboratory conditions and network configuration of the 2022 stimulation. The synthetic travel-times are computed with a JULIA-based Eikonal solver using three forward models: 1) a homogeneous velocity model (Vp = 5200 m/s, Vs = 2740 m/s), 2) a 1D velocity model based on Zhang and Pankow (2021), and 3) a linearly interpolated 3D model based on the SPAC model by Zhang and Pankow (2021). Because the sampler assumes a homogeneous velocity, the complication of forward modelled data with a 1D and 3D model should test the robustness of the sampler. Results are shown in Figure 3.4. While the 1D and 3D model do not produce a completely linear moveout, we can still obtain good locations with a linear fit,

because the events and lower section of the fibre lie in the relatively homogeneous granite layer. The high depth resolution of the DAS fibres is reflected in a lower variability in depth (~200 m) than horizontally (~400 m). We then put the sampler to an additional test by adding 10% and 20% random noise to the 3D model arrival time picks to verify if a good fit for the lower portion of the wells is still possible. The average location remains relatively unchanged; however the ground truth location is not sampled anymore at 20% noise. This shows that with various subsurface velocities and <20% noise conditions, the HMC sample is able to retrieve well- constrained locations.
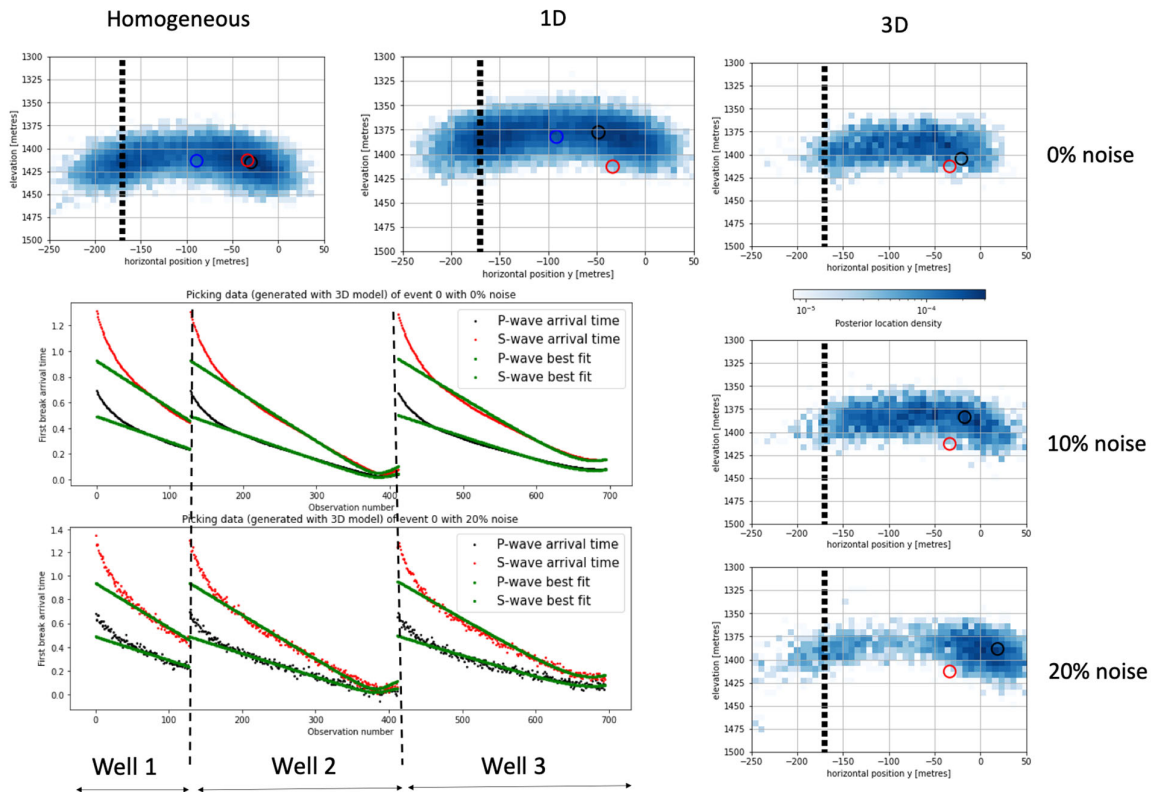

Figure 3.4. Homogeneous, 1D and 3D velocity model results, of which 3D also shows the performance with increasing noise level. The black location is the maximum likelihood output location from the HMC locator, and the red circle indicates the true location. In the homogeneous model the average location was also tested, which performed worse than the maximum likelihood. The picks on the 3D model data with the HMC fit in green are shown in the bottom left figure.

Next, we applied the method to a real case on events recorded by DAS at the Bedretto Underground Lab (see Figure 3.5) on November 23, 2022. Here three boreholes (in black) are instrumented with DAS, and a seismic event has been recorded by both the DAS and acoustic emission sensors. The acoustic emission sensors were already located and catalogued (marked "catalogue a" in Figure 3.5). The HMC method yielded a maximum likelihood position within 20 m of the catalogued location, with a location below 100 m. Looking at the picks, the model performed well in the first and last borehole, and performed more poorly on the middle borehole. This could be due to cycle skipping in the picking, or due to model heterogeneities that are not captured by the homogeneous model fitter. This could also contribute to the location error. Overall, the method performs well when multiple DAS-instrumented boreholes are present, which is promising for future datasets with multiple instrumented wells.
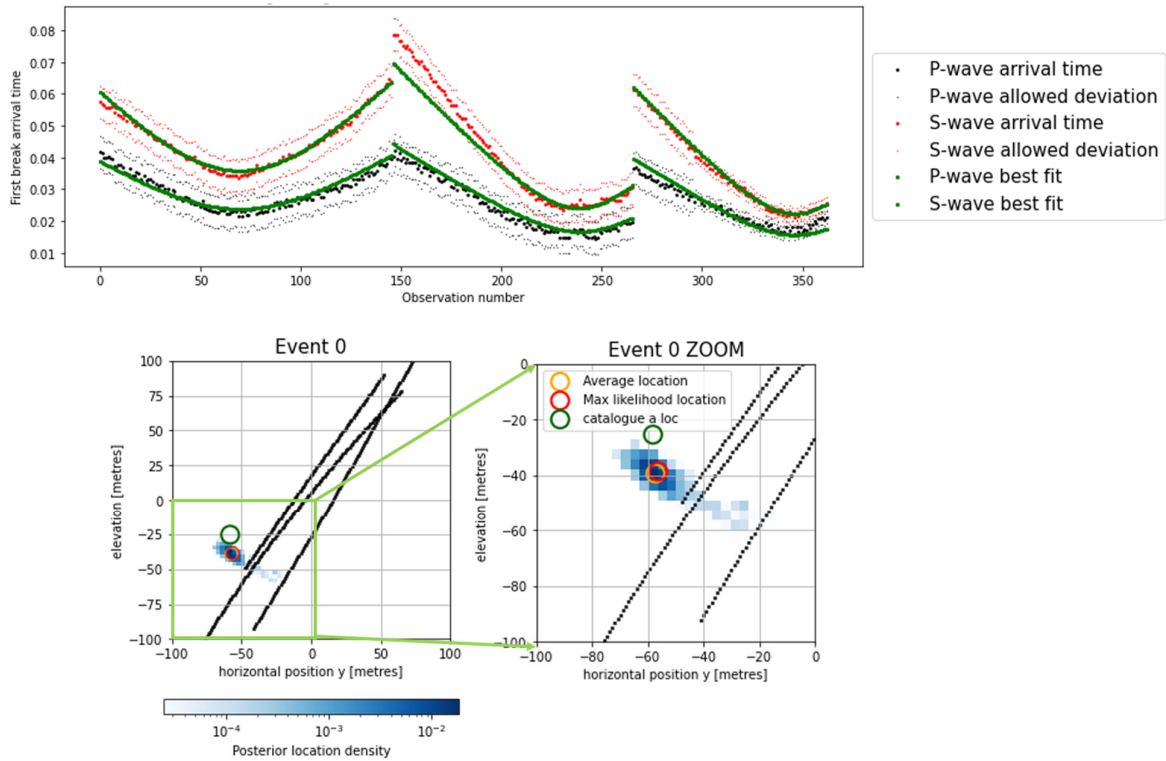
Figure 3.5. Results for HMC with Bedretto data. Top: picking data in black and red with the allowed covariance, and the model fit in green. Bottom: location results of the event with the DAS-instrumented wells in black.

3.3. Rapid microseismic event locations using waveform arrival times

Waveform-based travel times can be more accurate in a medium with strongly-varying properties, such as around boreholes, around abrupt interfaces, and in the vicinity of fault zones. Geothermal fields are particularly heterogeneous regions and algorithms for event locations due to stimulation or injection and production at such fields might benefit from improved travel time estimates (Angus et al. 2014). Unfortunately, numerical methods for waveform modeling are typically very computer intensive and a single simulation can require an hour or more of CPU time. Thus, location algorithms that scale as the number of events become impractical as this number exceeds several hundred or several thousand. Here we describe an algorithm that scales as the number of seismic sensors used to locate the events, and compare locations based upon waveform travel times with conventional locations using eikonal-based travel times. For this initial work we use a uniform velocity model, with a compressional velocity of 5800 m/s. In follow-on work we will consider the influence of sharp dipping boundaries.

3.3.1. Methodology

In this section we outline the procedure for locating events using either eikonal-based or waveform-based arrival times. The method is simply a table-lookup technique that may be interpreted as back-propagation from the station locations into the region of interest. The use of waveform-based travel times can potentially improve the locations in geologically complicated settings and around heterogeneous structures such as boreholes. The procedure is described in some detail in Vasco et al. (2019). We will illustrate the approach using the receiver locations from a stimulation at the FORGE site in central Utah. The various wells and seismic sensors positioned around the stimulation are shown in Figure 3.6.
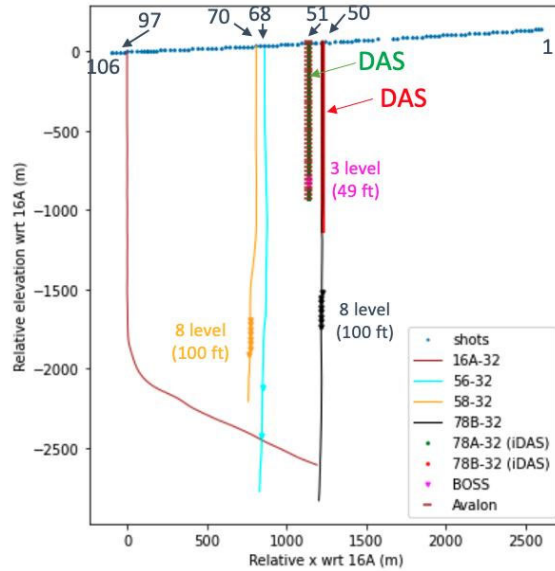
Figure 3.6. Wells and seismic sensor locations surrounding the stimulation in well 16A-32 (magenta curve).

Step 1: Calculate receiver travel time fields

All location algorithms assume that a method exists for computing seismic arrival times to any point within a volume under study. That is, a field $T(x_s, x_o)$ is available via some numerical method, providing the travel time from a source at point x_s to the location of an observing instrument at x_o. As in Nelson and Vidale (1990), the calculation may invoke the finite-difference solution of the eikonal equation (Vidale 1988, Podvin and Lecomte 1991, Zelt and Barton 1998, Sethian 1999). Here the eikonal-based travel time fields are calculated using the approach described by Zelt and Barton (1998). Our algorithm requires the travel time fields for waves propagating from every receiver point. As an example, we consider a horizontal slice through the travel time field for a compressional wave propagating from sensor location 1 (Figure 3.7).
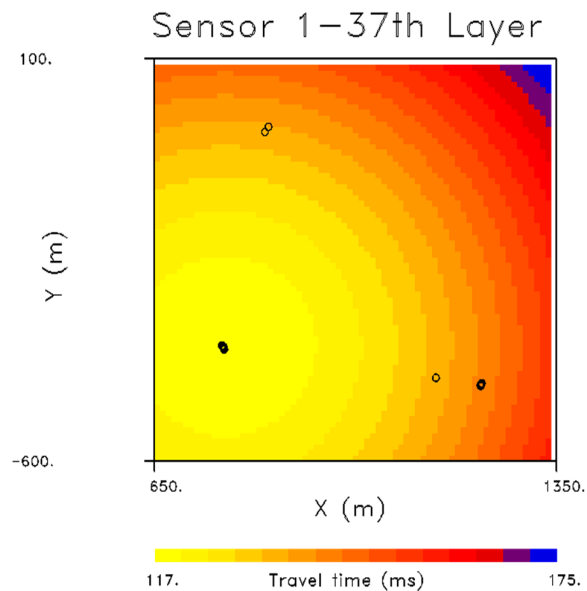

Figure 3.7. Travel time field for an elastic wave propagating from the location of sensor 1. The travel times were computed using the eikonal equation-based approach of Zelt and Barton (1998). The open circles denote the sensor locations, which are also displayed in Figure 3.6.

Greater fidelity, particularly in complicated geological settings, can be obtained by solving the full set of equations governing wave propagation in an elastic or poroelastic medium (Virieux 1986, Masson et al. 2006, Masson and Pride 2011). In Figure 3.8a we show a snapshot of the wavefield obtained for a source at the location of sensor 1. The panel illustrates the complexities of wavefield modeling with boundary reflections from the edges of the model. These arrivals can be reduced significantly using attenuation methods, but they do not influence the first arrival estimates, which are obtained by an automatic picking technique applied to the numerical waveforms. The result travel time calculations in Figure 3.8b are similar to the eikonal estimates shown in Figure 3.7.
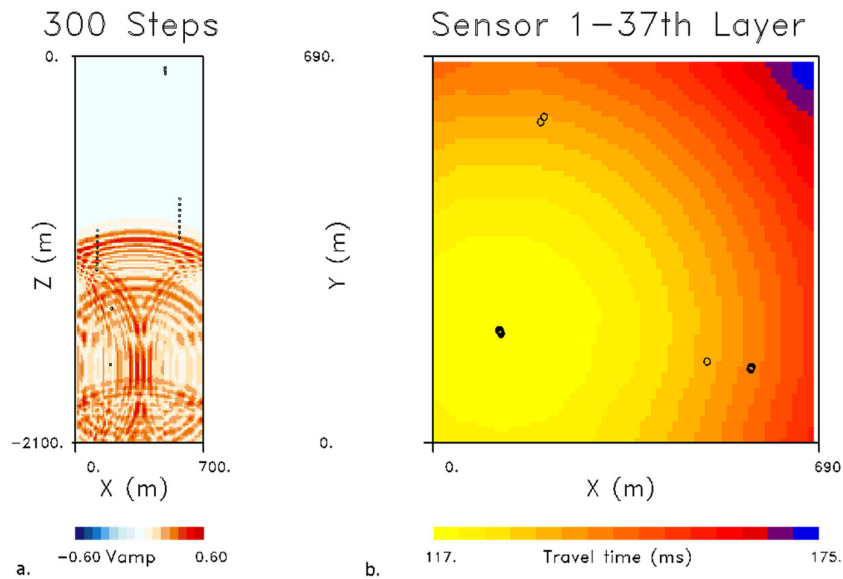

Figure 3.8. (a.) Snapshot after 300 time steps of the explicit finite-difference scheme of Masson and Pride (2011). (b.) Horizontal slice through the travel time field obtained by applying a picking algorithm to the wavefields calculated for each grid point of the finite difference mesh. The open circles denote the sensor locations (see Figure 3.7).

The travel time fields are computed for each of the 21 receivers that were available during the FORGE stimulation of April 2022. We focused on the 3rd stage of the experiment, as the full complement of instruments were operating at this time. In order to reduce the memory requirements and computation time to a manageable level for our given workstation, we adopted a grid spacing of 5 m in all three directions for both the eikonal and wavefield calculations. The eikonal computations were completed in a matter of minutes. The wavefield computations for all 21 sensors took much longer, about 1.5 days. The wavefield travel times were produced by a simple threshold approach whereby the arrival time was defined where the amplitude exceeded a fixed fraction of the amplitude of the trace maximum. If the velocity model is fixed then this calculation is complete. If the locations are part of a tomographic imaging scheme and the velocity model is updated then the calculation of the travel times fields for the sensors must be repeated for each update.

Step 2: Time reversed or back-propagation from the receivers

With the sensor travel time fields in hand, the location procedure involves simulating a wave starting from the receiver at the time of the observed arrival but propagating backward in time. That is, we propagate back through the grid, subtracting the travel time to each node from the observed arrival time. This is done for each sensor. Thus, we have N travel times to each node in the grid. For error-free travel times and perfect modeling, the N travel times should converge at the event location to the origin time of the event. Given the presence of errors, there will be some disagreement between the grid block arrival times. However, if there are no large systematic errors, the scatter of the arrival times should be minimized at the source location.

Step 3: Estimate the dispersion of the travel times for each grid node

To find the location we calculate the mean arrival time for the travel time fields for all nodes and the scatter of the residuals with respect to the mean. At the event location the scatter should be a minimum. The mean of the absolute values of the residuals is used as a robust estimator, as it is not so strongly influenced by the presence of a few outliers. An example of the variation of the residual fields for two horizontal slides through the modeling grid are shown in Figure 3.9 for the second event that was detected during the 3rd stage of the simulation. The left panel corresponds to residuals associated with the eikonal equation while the panel on the right corresponds to waveform-based travel times.
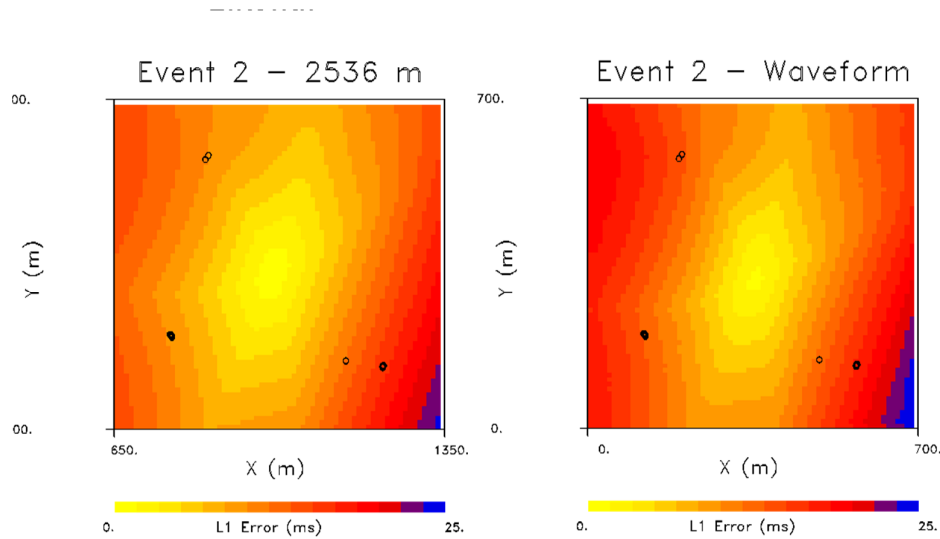

Figure 3.9. (Left panel) Mean absolute residuals associated with event 2 of the FORGE stage 3 stimulation. The residuals are with respect to travel times computed using the eikonal equation. (Right panel) Mean absolute residuals for a horizontal slice that cuts through the modeling grid at the estimated depth of the event. These residuals are associated with waveform-based travel times.

Not only do the residuals provide estimates of the event location and origin time. They also produce uncertainties of the event location, given estimates of the uncertainties on the travel times. That is, one can use the three-dimensional surface of the error contour as a measure of the location uncertainty. As shown in Figure 3.9, the contours can take non-quadratic forms due to the discontinuous nature of the absolute value used to calculate the total error. As an additional example, consider the last located microseismic event in the sequence, number 1526. The error contours for the eikonal- and waveform-based residuals are shown in Figure 3.10. There were fewer picks for this event resulting in greater uncertainty in the location. Both sets of residuals display the increased uncertainty. Note that there is less uncertainty for the waveform approach in both cases, as is evident from an examination of the contours in Figures 3.9 and 3.10.
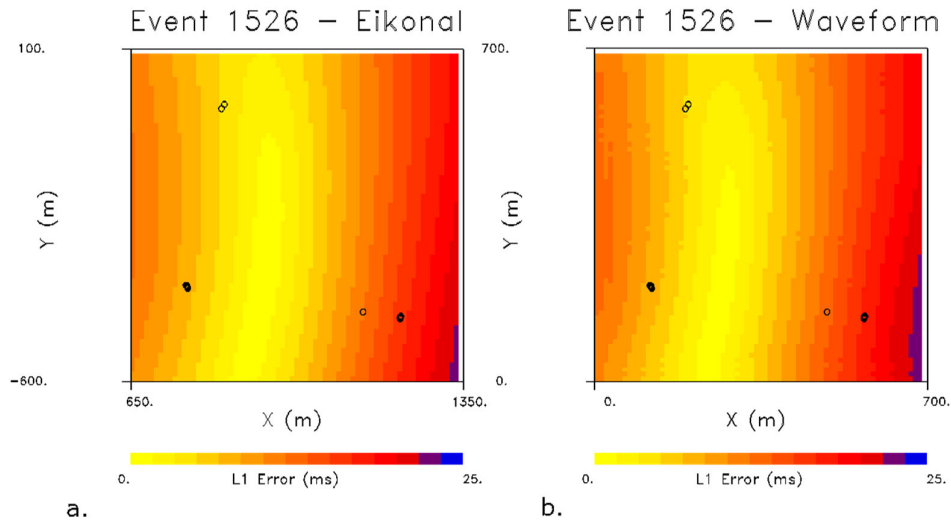
Figure 3.10. Mean absolution error variations within the model grid for event 1526. Both the eikonal-based and the waveform-based error estimate display the greater uncertainty in the north-south location of this event.

In many cases the uncertainty in depth can be large due to the geometry of the array of sensors, particularly for surface-based instruments. For the borehole-based arrays used to monitor seismicity at the FORGE site the depth resolution is found to be somewhat worse than the lateral uncertainty (Figure 3.11) due to the distribution of most stations above the event locations.
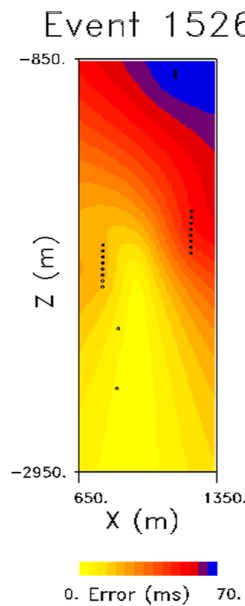

Figure 3.11. Vertical cross-section in east-west distance and depth, with the slice through the north-south event location.

Last step: Determine the point at which the dispersion is a minimum

We can find the location of the event using a simple grid search to find the minimum. However, our spatial resolution is then limited by the grid spacing, in this case 5 meters. We can use interpolation and a quadratic fit to try and fine-tune the location to sub-grid resolution. The discontinuities of the absolute value function can introduce some difficulties, but close enough to the minimum a quadratic approximation might be acceptable.
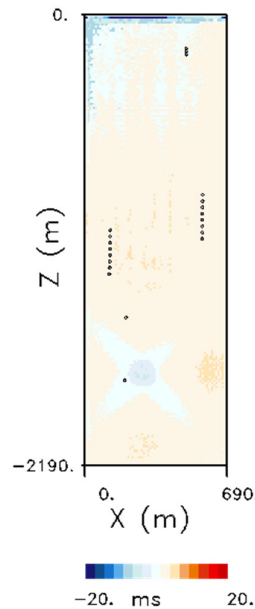
Figure 3.12. Travel time differences between eikonal-based arrivals and arrivals picked from numerical waveforms.

As a final note on the location procedure, before discussing the full set of relocated events, we considered the travel time differences between the eikonal-based travel times and those calculated using waveform calculation and automatic picking. Though our finite-difference code was designed for elastic wave propagation in a poroelastic medium, we were able to adapt it for an elastic medium by severely reducing the porosity to 0.01% and making the fluid density and elastic properties essentially identical to the solid frame properties. Thus, the travel times calculated using the eikonal equation and from picking travel times from numerical waveforms are fairly close in value (Figure 3.12). The greatest deviation is near the top boundary, where a boundary condition in the modeling led to some differences. However, this region was outside of the grid of seismic sensors and did not influence the location procedure. Note that other solvers, appropriate for special situations, such as an acoustic finite-difference code, or a semi-analytic layered code may be much faster than our poroelastic calculations.

3.3.2. Results: Application to the FORGE stimulation of April 2022

After testing the location algorithm and conducting trials with various sized grids and boundary conditions, we applied the technique to the available location data from 1526 events that were identified and located after Stage 3 of the April 2022 stimulation at the FORGE geothermal test site. The downhole monitoring network for the 2022 stimulation was operated by Geo-Energie Suisse. They also estimated the seismic phases and identified phases for the 1526 events deemed reliable from Stage 3 of the stimulation at FORGE (Figure 3.13). Their location algorithm utilized both compressional and shear phases where they were available. The method seeks the minimum of the sum of the squares of the residuals, searching over a grid with nodes spaced every 2 feet.
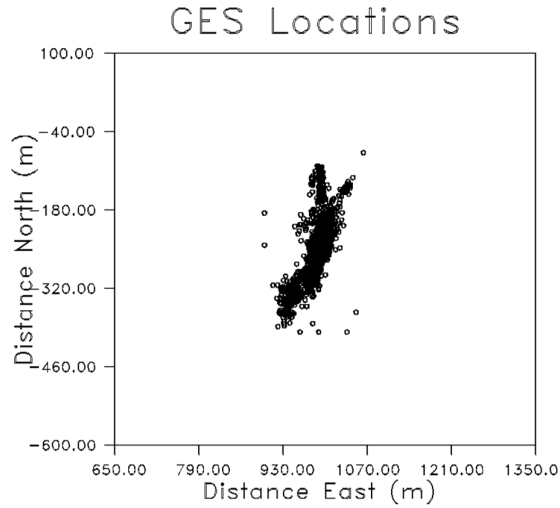
Figure 3.13. Epicenters of the 1526 events identified by Geo-Energie Suisse for the April 2022 Stage 3 stimulation.

As described above, we constructed the 21 travel time grids for all of the active sensors in the monitoring network shown in Figure 3.6. Both the eikonal equation and the numerical finite-difference waveform simulations were used to compute the arrival times. Once the travel time grids for the 21 sensors were computed it was a simple manner to backpropagate the arrival times for each event in order to derive the location and origin time. The resulting hypocenters for the events located using the eikonal equation travel times are plotted in Figure 3.14.
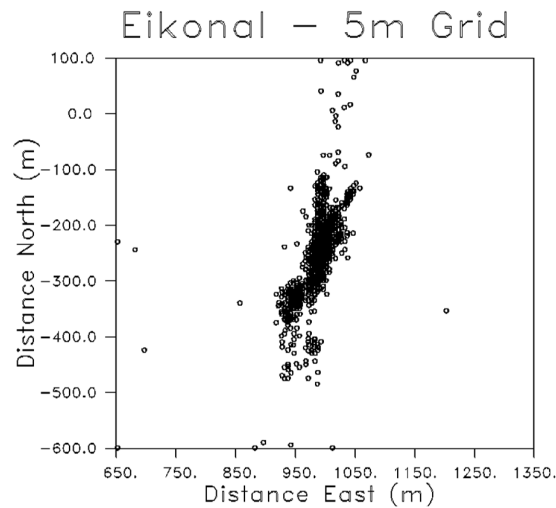

Figure 3.14. Event locations produced using the algorithm described above with eikonal-based arrival times.

The overall distribution is similar to the locations determined by Geo-Energie Suisse, with a central cloud of events trending roughly north-south. There are two linear north-south and north-easterly offshoots from the northern end of the cloud, as well as a secondary cloud at the south-eastern edge of the main body. Using the waveform-based travel times in the location procedure also produces a similar distribution of events (Figure 3.15). For the uniform velocity model used in all three location algorithms we would not expect significantly different locations for the different realizations. The eikonal equation is valid for smoothly-varying velocity models and high frequency seismic waves with sharp onsets.
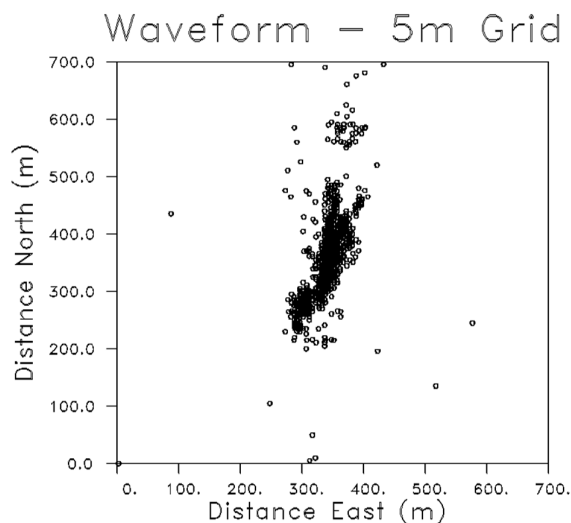
Figure 3.15. Locations based upon waveform-based arrival times.

The evolution of seismicity with time is of particular interest in order to gain insight into fracture development and fluid flow during the stimulation. Therefore, we considered short time increments, with most intervals containing 100 events. The initial and final intervals contain fewer events. We plot the eikonal and waveform results side-by-side in all four plots of the 1526 event sequence. The initial 50 events in Figure 3.16 imply the opening of a fracture oriented to the north east from the injection well. This is followed by seismicity in a northeast oriented linear zone that crosses the stimulation well. Both approaches also display a sparse set of events oriented along a north-south line. There are considerable differences in the location of these sparse events, casting doubt on their validity. The next set of four snapshots (Figure 3.17) are dominated by events in the northeast oriented linear zone crossing the well. Again, there are a sparse series of events in a north-south lineation. However, they do not seem well constrained. In Figure 3.18 the next sequence of four snapshots appear to define the development of a north-south oriented feature that heads off to the north of the stimulation well 16A. There is an offshoot to the northeast that appears to extend from the northern the north-south zone of events. The events to the south of the stimulation well appear to be much more diffuse in comparison to the north trending linear feature to the north.

The rapid location method appears to be a practical approach, and even with a uniform velocity model, the method seems sufficiently precise, in comparison to a conventional eikonal-based methodology. Once the travel time fields are computed for each sensor, the approach is as fast as any table-lookup approach. Thus, if the travel time fields are pre-computed prior to the experiment, the approach could be used for near real-time event locations in a full three-dimensional setting allowing for general variations in material properties.

In an extension of this study, we will incorporate more complex velocity structures, as often instruments (e. g., surface deployments) are located across substantial velocity interfaces. The FORGE site, for example contains a substantial velocity change across a dipping interface between overlying valley sediments and a deeper granitic body, and although the geophone chains, for the most part are below this interface, however there are near-surface instruments including a DAS cable the record waves crossing the interface. Furthermore, there are surface deployments that record waves crossing shallow sediments with relatively low velocities. A general three-dimensional velocity model is currently under development, using travel time data from a seismic reflection survey.

As shown in Vasco et al. (2019), we have the capability of incorporating shear arrivals for the eikonal-based times. But have not yet incorporated a shear arrival time estimation procedure necessary for picking secondary arrivals from the finite-difference simulations. This issue can be easily overcome by incorporating current phase pickers, such as those provided by machine-learning approaches. (see also D2.1, D2.2). That will be a subject for the research in the coming year. Incorporating secondary arrivals should better constrain events, particular those where there are few P arrivals identified, reducing the scatter in our locations in Figures 3.14 and 3.15. Then our approach will be comparable to the previous least-squares approach shown in Figure 3.13.
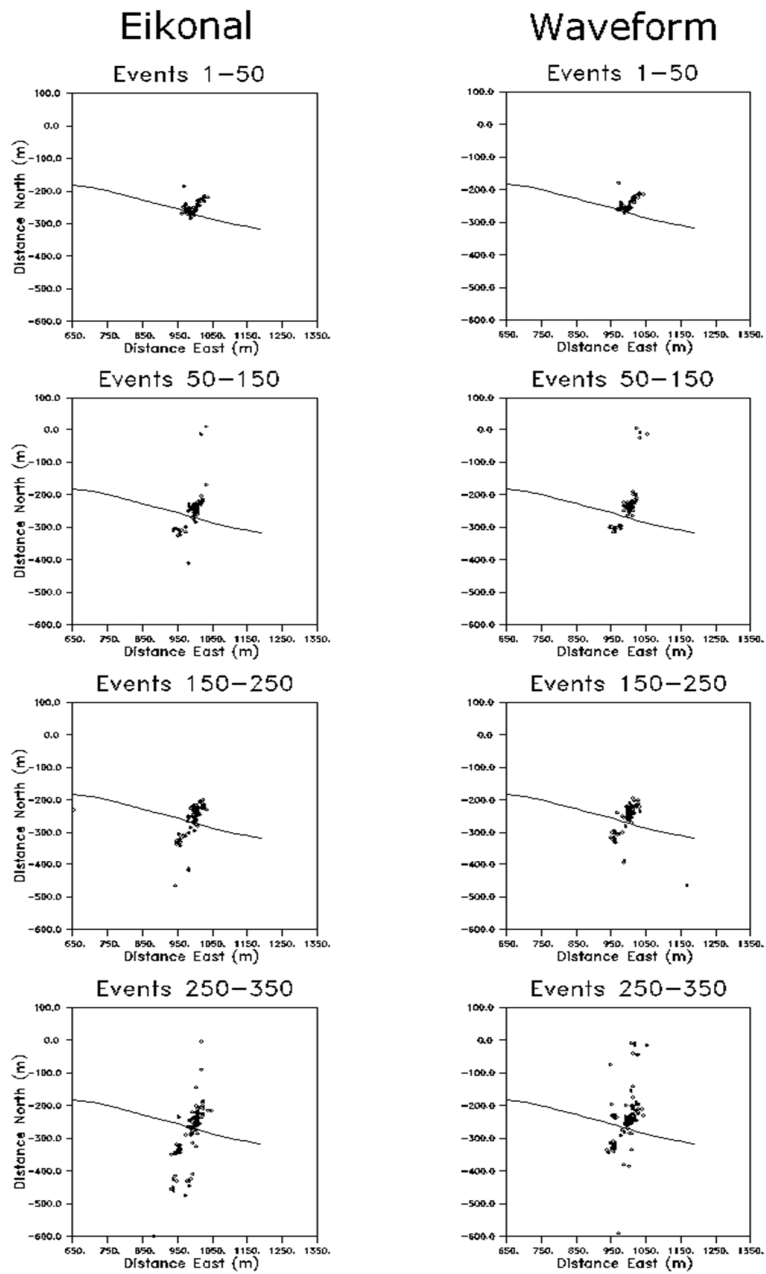
Figure 3.16. First four snapshots of seismicity during Stage 3. The black curve denotes the track of well 16A.
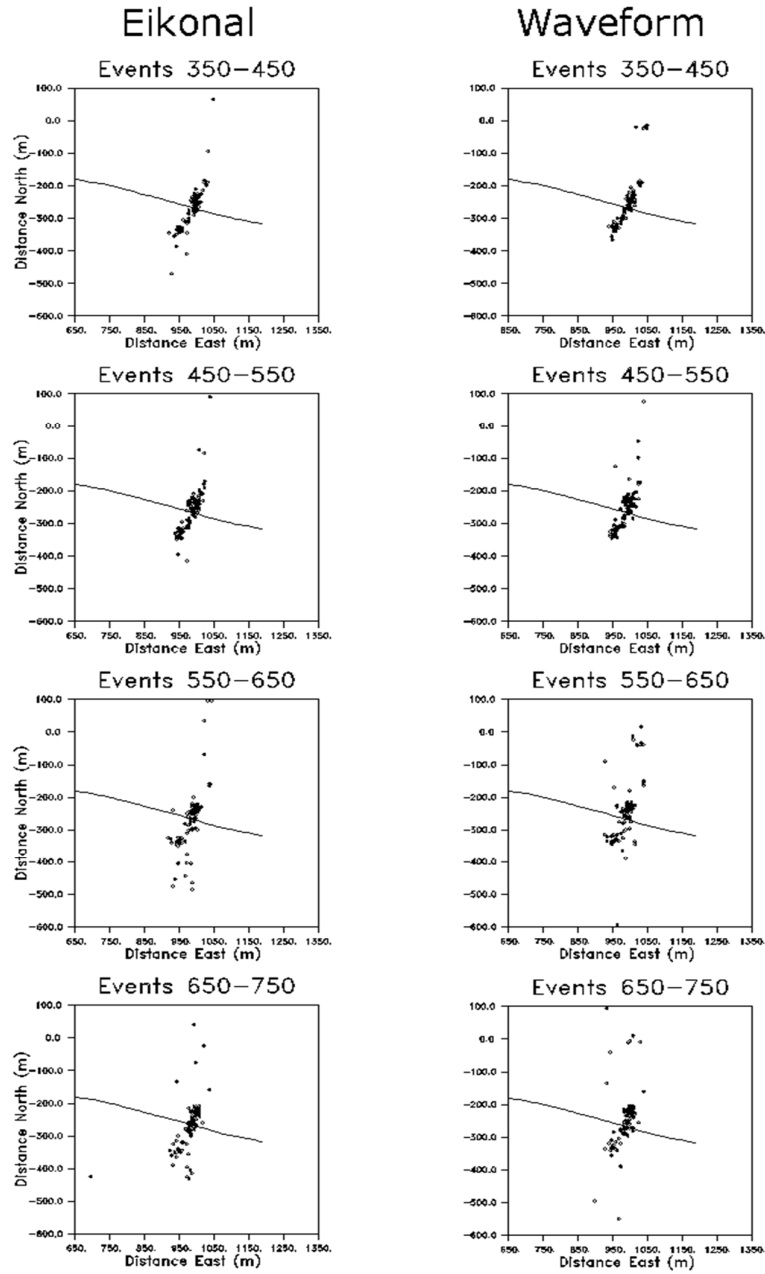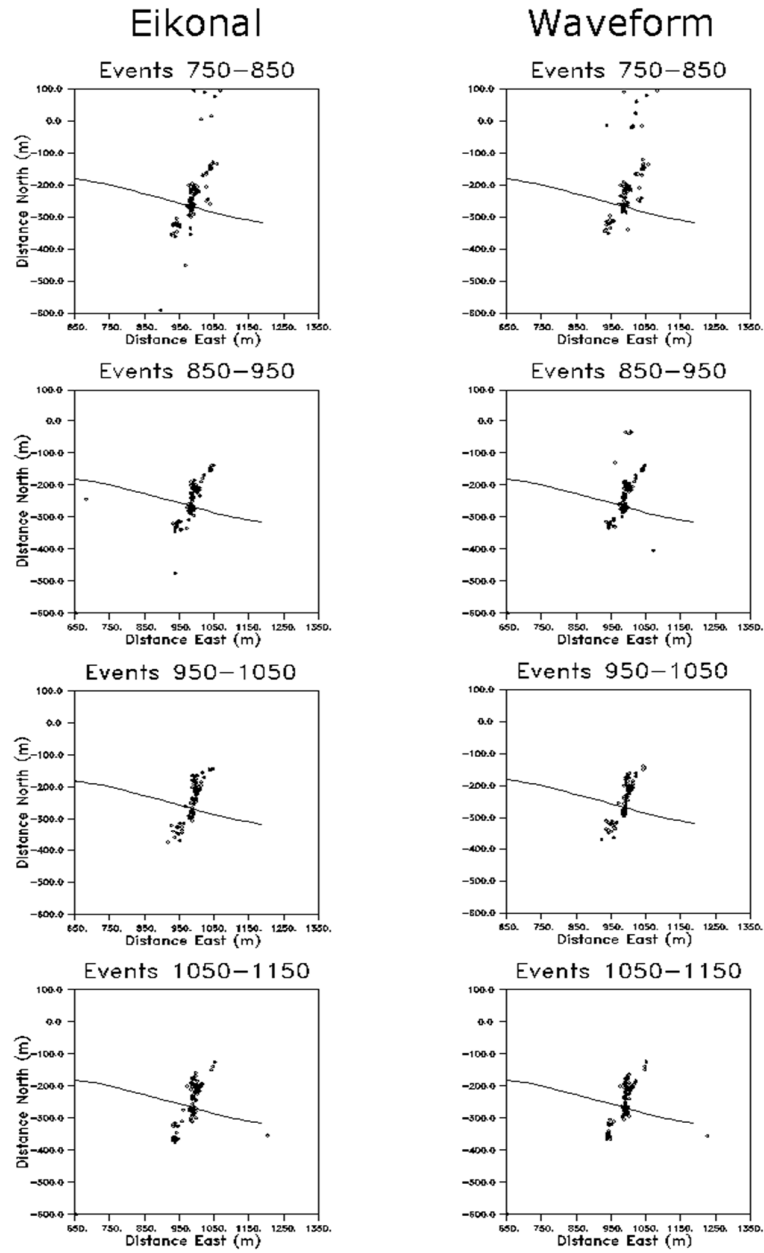
Figure 3.17. Next set of four snapshots

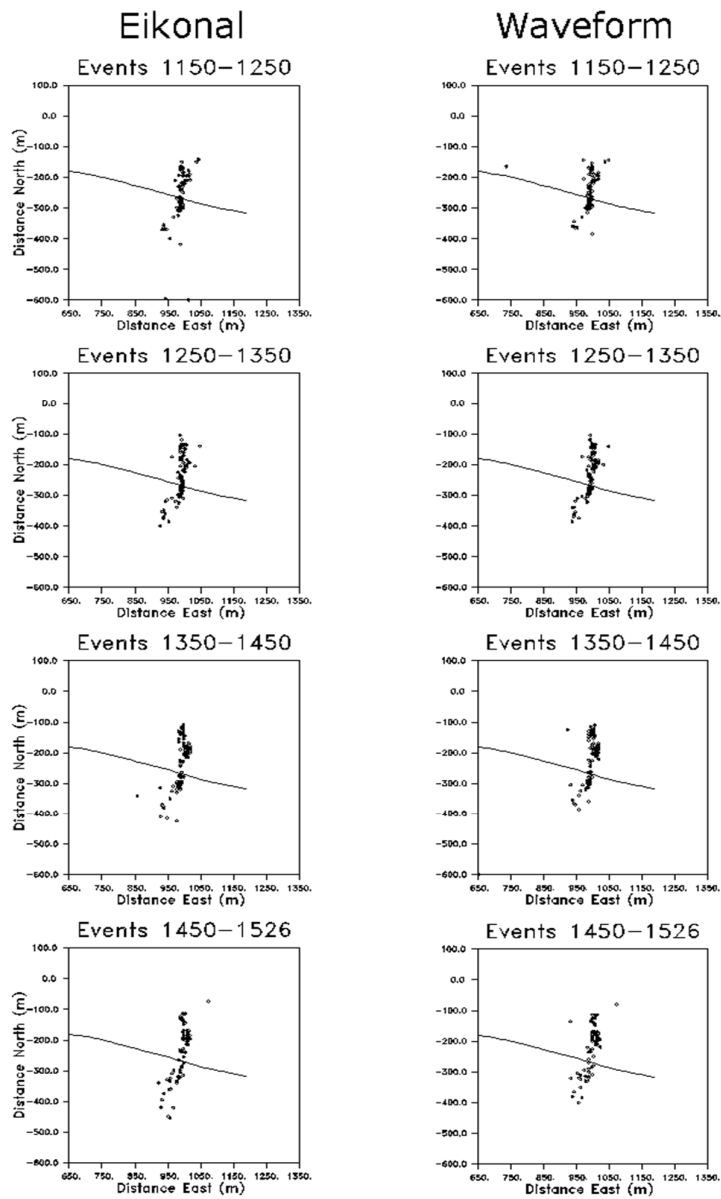**Figure 3.18**. Set of four intervals of events in sequence.

Figure 3.19. Final four sequences of events

4.   Conclusion and Outlook

We have shown new methodologies targeted at improving the monitoring of micro seismicity in borehole settings. The methods presented here constitute a fundamental first step towards the integration in real-time processing of both DAS data streams and downhole geophone deployments. Although we showed the potential for real-time applications, in the next chapters we highlight the current challenges that still need to be addressed in order to achieve full real-time performances and integration.

4.1.   Challenges and future directions for real-time integration of DAS

One of the challenges of DAS that still hinder a fully real-time operational DAS deployment in continuous monitoring settings is data transfer and storage. DAS datasets range in the order of 10^1-10^2 Terabytes (TB) per experiment (Quinteros et al., 2021), with several TB usually acquired per day or per week depending on acquisition settings. This leads to limitations in the capacity of permanent waveform archival as well as real-time data transfer from the acquisition

computer to the computer running the analyses. Lessons from the FORGE 2022 stimulation show that, even without considering the large amount of data collected by DAS, the high number of detections on the geophone streams and file size created issues/backlogs in the real-time processing. Moreover, the current lack of standardised installation requirements and rapid data processing routines for DAS contribute to challenge the real-time integration of DAS in the monitoring workflow. The new methodologies brought forward in this deliverable are the first steps directed at closing the gap between geophone and DAS processing routines, by providing the standards and computationally efficient solutions required for future real-time integration. Proposed directions to enhance the real-time potential of DAS include the optimization of the algorithms (e.g., parallelization, GPU acceleration) and a dynamic data downsampling scheme that, while allowing for fast data transfer and waveform archival, would not degrade key wavefield information.

We note here that DAS technology should always be paired with downhole geophones, as it has been demonstrated that geochains have inherently higher signal to noise ratio than DAS and they provide tri-axial measurements, rather than a single uni-axial measurement along the fiber direction. However, because DAS has a higher spatial sampling, can remain in place undisturbed behind casing, and can acquire data even when other instruments are not deployed or present mechanical failures due to e.g., high temperature, constitutes all advantages to keep exploring its suitability for monitoring.

4.2.   Challenges and future directions for downhole geophones monitoring systems

Contrary to DAS, geophone chains are already equipped with telemetry options and real-time capabilities, as processing of data in near-real time is routinely performed during EGS (or other type of) stimulations. However, telemetry is rarely integrated with surface geophone deployments and typical downhole microseismic processing in one-well network configurations often need to resort to ad-hoc, manual processing of the geophone streams using polarization analysis techniques (hodograms), due to the lack of azimuthal information. Polarization analysis is used to determine the direction of a wave's approach to the receivers, thus it can provide estimates of the source azimuth and thereby help constrain earthquake locations. To automate this step and integrate geophone streams with surface networks, we are currently developing an end-to-end (from waveforms to catalog) monitoring system with SeisComP, an open-access seismological software for real-time monitoring.

In detail, in order to integrate vertical geophone chains into the classical SeisComP workflow (scautoloc/scanloc processing flow), it is necessary to use array analysis techniques, and in particular polarization analysis to calculate azimuths and incidence angles. In this way (1) P-wave polarization can be exploited to improve detection of smaller events below conventional thresholds, and (2) the back azimuth information can be included in the hypocenter location inversion of travel-time observations allowing for a refined accuracy in the final locations. We plan to leverage the existing capabilities of LAMBDA, a plugin module developed by gempa GmbH, that already handles surface arrays, and adapt it to vertical borehole arrays, and the high sampling rates typically used in EGS monitoring.

LAMBDA is already embedded in the SeisComP environment: it performs detections and retrieve beam parameters (e.g., arrival-times, back azimuth and slowness) from a beam trace that then can be carried out to the traditional SeisComP workflow (e.g., beam picks from LAMBDA are processed together with scanloc picks to obtain a higher- precision location. Scanloc can also assign pick priorities, for example to the calculated beam picks from the geophone arrays.

We anticipate to have the full system working by the end of the project and results of playback scenarios of the 2022 FORGE stimulation will be included in the second part of this deliverable and in milestone M2.5.

Liability Claim

Reference List

Angus, D. A., Aljaafari, A., Usher, P., and Verdon, J. P. (2014). Seismic waveforms and velocity model heterogeneity: Towards a full-waveform microseismic location algorithm, Journal of Applied Geophysics, 111, 228-233, http://dx.doi.org/10.1016/j.jappgeo.2014.10.013.

Fichtner, A., Zunino, A., & Gebraad, L. (2018). Hamiltonian Monte Carlo solution of tomographic inverse problems. *Geophys. J. Int.*, 216:doi:10.1093/gji/ggy496.

Grigoli, F., Ellsworth, W. L., Zhang, M., Mousavi, M., Cesca, S., Satriano, C., Beroza, G. C., & Wiemer, S. (2021). Relative earthquake location procedure for clustered seismicity with a single station, Geophysical Journal International, 225(1), 608-626.

Hertrich, M., Brixel, B., Broeker, K. et al. (2021). Characterization, Hydraulic Stimulation, and Fluid Circulation Experiments in the Bedretto Underground Laboratory for Geosciences and Geoenergies, Paper presented at the 55th U.S. Rock Mechanics/Geomechanics Symposium.

Jolliffe, I. (2002). Principal component analysis for special types of data. Springer, 2002.

Lellouch, A., Lindsey, N. J., Ellsworth, W. L., & Biondi, B. L. (2020). Comparison between Distributed Acoustic Sensing and Geophones: Downhole Microseismic Monitoring of the FORGE Geothermal Experiment, Seismol. Res. Lett. 91, 3256–3268.

Lellouch, A., Schultz, R., Lindsey, N. J., Biondi, B. L., & Ellsworth, W. L. (2021). Low-magnitude seismicity with a downhole distributed acoustic sensing array—Examples from the FORGE geothermal experiment, Journal of Geophysical Research: Solid Earth, 126(1), e2020JB020462.

Masson, Y. J., & Pride, S. R. (2011). Seismic attenuation due to patchy saturation, Journal of Geophysical Research, 116, 1-17, doi:10.1029/2010JB007983.

Masson, Y. J., Pride, S. R., and Nihei, K. T. (2006). Finite difference modeling of Biot's poroelastic equations at seismic frequencies, Journal of Geophysical Research, 111, 1-13, doi:10.1029/2006JB004366.

Moore, J., McLennan, J., Pankow, K., Simmons, S. Podgorney, R., Wannamaker, P., Jones, C., Rickard, W., & Xing, P. (2020). The Utah Frontier Observatory for Research in Geothermal Energy (FORGE): A Laboratory for characterizing, creating and sustaining Enhanced Geothermal Systems, Proc. of 45th Workshop on Geothermal Reservoir Engineering Stanford University, 1–10.

Neidell, N. S., & Taner, M. T. (1971). Semblance and other coherency measures for multichannel data, Geophysics, 36 (3): 482–497. doi: https://doi.org/10.1190/1.1440186.

Nelson, G. D., & Vidale, J. (1990). Earthquake locations by 3-D finite-difference travel times, Bulletin of the Seismological Society of America, 80, 395-410.

Pankow, K., Mesimeri, M., Mclennan, J., Wannamaker, P., & Moore, J. (2020). Seismic monitoring at the Utah frontier observatory for research in geothermal energy, Proc. of 45th Workshop on Geothermal Reservoir Engineering Stanford University, 1–9.

Podvin, P., & I. Lecomte (1991). Finite-difference computation of travel times in very contrasted velocity models: A massively parallel approach and its associated tools, Geophysical Journal International, 105, 271-284.

Porras, J. (2022). A semblance-based Microseismic Event Detector for DAS data. Master Thesis 2022. University of Pisa.

Quinteros, J., Carter, J. A., Schaeffer, J., Trabant, C., Pedersen, H. A. (2021). Exploring Approaches for Large Data in Seismology: User and Data Repository Perspectives. Seismological Research Letters, 92 (3): 1531–1540. doi: https://doi.org/10.1785/0220200390

Sethian, J. A. (1999). Level Set Methods and Fast Marching Methods, Cambridge University Press, Cambridge.

Shoemake, K. (1985). Animating rotation with quaternion curves. In Proceedings of the 12th annual conference on Computer graphics and interactive techniques, pages 245–254, 1985

Tuinstra, K. Locating clustered seismicity using Distance Geometry Solvers: applications with sparse seismic networks and borehole Distributed Acoustic Sensing, submitted to GJI.

Vasco, D. W., Nakagawa, S., Petrov, P., and Newman, G. (2019). Rapid estimation of earthquake locations using waveform traveltimes, Geophysical Journal International, 217, 1727-1741, doi: 10.1093/gji/ggz114

Zelt, C. A., and Barton, P. J. (1998). 3D seismic refraction tomography: A comparison of two methods applied to data from the Faeroe Basin, Journal of Geophysical Research, 103, 7187-7210.

Zhang, H., & Pankow, K. L. (2021). High-resolution Bayesian spatial autocorrelation (SPAC) quasi-3-D Vs model of Utah FORGE site with a dense geophone array, Geophysical Journal International, 225(3), 1605-1615.

Wannamaker, P., Simmons, S., Miller, J., Hardwick, C., Erickson, B., Bowman, S., Kirby, S., Feigl, K. & Moore, J. (2020). Geophysical Activities over the Utah FORGE Site at the Outset of Project Phase 3, Proc. of 45th Workshop on Geothermal Reservoir Engineering Stanford University, 1–14.