

Data-Driven Geothermal Microseismic Clustering Using Machine Learning

Ryoya Tsuboi

Tohoku University: Tohoku Daigaku

Yusuke Mukuhira

mukuhira@tohoku.ac.jp

Tohoku University: Tohoku Daigaku <https://orcid.org/0000-0002-9356-8302>

Kyosuke Okamoto

AIST FREA: Sangyo Gijutsu Sogo Kenkyujo Fukushima Saisei Kano Energy Kenkyujo

Hiroshi Asanuma

AIST FREA: Sangyo Gijutsu Sogo Kenkyujo Fukushima Saisei Kano Energy Kenkyujo

Research Article

Keywords: Geothermal reservoirs, Fracture network, Microseismicity, Clustering, Data driven, Unsupervised machine learning

Posted Date: October 8th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10914083/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Data-Driven Geothermal Microseismic Clustering Using Machine Learning

Author #1: Ryoya Tsuboi, Graduate School of Environmental Science, Tohoku

University, tsuboi.ryoya.t8@dc.tohoku.ac.jp

Author #2: Yusuke Mukuhira, Institute of Fluid Science, Tohoku University,

mukuhira@tohoku.ac.jp

Author #3: Kyosuke Okamoto, National Institute of Advanced Industrial Science and

Technology, okamoto.kyosuke@aist.go.jp

Author #4: Hiroshi Asanuma, National Institute of Advanced Industrial Science and

Technology, h.asanuma@aist.go.jp

Corresponding author: Yusuke Mukuhira

12 **Abstract**

13 Accurate determination of microseismic hypocenters in geothermal reservoirs enables
14 reliable characterization of subsurface fracture systems. Hypocenter relocation
15 techniques include microseismic clustering, implicitly or explicitly, in which events with
16 high waveform similarity are grouped within a distance threshold and assumed to
17 originate from common structures. Conventional clustering methods mainly rely on
18 waveform similarity and can effectively capture seismic source characteristics; however,
19 they often require extensive trial-and-error procedures, including station selection,
20 frequency-band tuning, and waveform-similarity and distance thresholds. Microseismic
21 clustering results might inadequately capture the temporal and spatial patterns of
22 microseismic activity. As an alternative approach, this study proposes a data-driven
23 microseismic clustering framework based on unsupervised machine learning. In addition
24 to waveform similarity quantified by normalized cross correlation, spatial parameters
25 such as inter-event distance and angular information relative to the injection point, as well
26 as temporal proximity between events, were incorporated as explanatory variables. The
27 proposed framework was applied to microseismic events recorded during hydraulic

stimulation at the Basel Enhanced Geothermal System (EGS) site in Switzerland, for which previously reported conventional clustering results are available. The results show that by properly integrating spatial and temporal information with waveform similarity, the proposed method successfully extracted microseismic cluster structures consistent with the geothermal reservoir's stress field, comparable to those identified by conventional approaches. The proposed method effectively suppressed the clustering of temporally distant events into the same cluster. Furthermore, rather than relying solely on inter-event distance as a measure of spatial similarity, a spherical coordinate system centered on injection point was introduced, and angular information was incorporated into the clustering framework. This approach reduced spatial dispersion in cluster distributions and yielded clustering results that are more physically interpretable than those obtained with conventional methods. These findings suggest that the proposed data-driven clustering framework offers an effective and flexible approach for high-resolution fracture characterization in geothermal reservoirs and will be incorporated into the relocation process.

44 **Keywords**

45 Geothermal reservoirs; Fracture network; Microseismicity; Clustering; Data driven;

46 Unsupervised machine learning

47

48 **1 Introduction**

49 Geothermal is a renewable energy resource that extracts heat stored in the
50 subsurface and, when properly utilized, enables long-term, stable energy production
51 without reliance on depletable fossil fuels. The realization of geothermal power
52 generation requires a geothermal reservoir with three essential components: a sufficiently
53 high-temperature heat source, fluids that transport thermal energy, and fracture network
54 that permits fluid flow and works as a heat exchanger. An Enhanced Geothermal System
55 (EGS) compensates for missing factors through an engineering approach, such as fluid
56 injection to enhance fracture permeability by inducing shear slip along existing fractures
57 (Tester et al. 2006). In the meantime, microseismicity generated by shear slip along
58 fractures provides crucial information about the reservoir, including the locations of
59 permeable fractures, their dimensions and orientations, the fracture system therein, and
60 the shape of the reservoir (Evans et al. 2004; Majer et al. 2007). It is therefore crucial to
61 characterize fracture networks within geothermal reservoirs to improve geothermal
62 resource productivity and accurately evaluate reservoir properties (Darisma et al. 2024).

63 Microseismic monitoring has been widely adopted as a primary observational

tool for this purpose. Microseismic events occurring in geothermal reservoirs are generally interpreted as resulting from shear slip along pre-existing fractures (Pine and Batchelor 1984). Accordingly, the spatial distribution of microseismic hypocenters provides subsurface fracture planes, and this approach has been widely applied in geothermal development and reservoir evaluation studies (Baria et al. 1999; Phillips 2000; Moriya et al. 2002; Mukuhira et al. 2013; Kwiatak et al. 2014, 2019; Kim et al. 2018; Bentz et al. 2019; Darisma et al. 2024; Feng et al. 2026). However, estimated hypocenter locations include uncertainties associated with heterogeneous subsurface velocity structures, seismic wave propagation paths, and the state of the reservoir fluid, as well as potential bias in the picking times of P- and S-wave arrivals due to human factors. To extract fracture network information with higher accuracy, improving the precision of hypocenter relocation is therefore indispensable.

In hypocenter relocation, relative P- and S-wave arrival times among a group of microseismic events are used to determine their relative locations (Waldhauser and Ellsworth 2000; Trugman and Shearer 2017; McBrearty and Beroza 2025). Defining the group of microseismic events can be considered a clustering problem. Microseismic

clustering is commonly, though implicitly, employed in hypocenter relocation to classify microseismic events with high spatial and waveform similarity into the same group. Events belonging to a common cluster are assumed to originate from shear slip on the same fracture, suggesting that their source mechanisms and seismic wave propagation paths are similar. Based on this assumption, the relative location of events within a cluster can be estimated with significantly higher accuracy, enabling more reliable characterization of subsurface fracture geometries (Waldhauser and Ellsworth 2002; Ross et al. 2020; Wilding et al. 2023; Sawaki et al. 2025a, b).

In statistical seismology, a number of methods were introduced to subdivide the seismic sequence (declustering) into mainshock-aftershock sequences, seismic swarms, and background seismicity (Gardner and Knopoff 1974; Reasenberg 1985). Results of such declustering sometimes provide spatial clustering that corresponds to the seismic activity from along specific faults. Recently, declustering has also employed machine learning approaches (Zaliapin and Ben-Zion 2020; Aden-Antoniów et al. 2022). While statistical seismology focuses primarily on separating temporal-spatial seismic components, the clustering problem in geothermal application specifically addresses the

identification and delineation of the existing fracture system within the microseismic analysis for geothermal reservoir. In the case of multistage injection, the microseismicity clustering was performed based on the injection stage (Kwiatek et al. 2019) as often seen in microseismicity monitoring in fracturing for shale formation. However, to identify the fracture system in geothermal reservoirs, beyond detection of macroscopic structure or seismic cloud, more precise and rigor clustering is required.

One conventional microseismic clustering method is multiplet clustering analysis based on waveform similarity (Moriya et al. 2002, 2003; Dyer et al. 2010). A multiplet is a group of microseismic events with mutually similar waveforms. It is commonly interpreted as representing repeated shear failure on the same fracture or asperity of a fault plane (e.g., Geller and Mueller, 1980). In this approach, quantitative waveform similarity is evaluated for all possible event pairs, and those exceeding a predefined threshold are classified into the same cluster with other thresholds such as inter event distance. Waveform similarity is most commonly quantified using the waveform cross correlation coefficient, which measures the similarity between two discrete-time waveforms while accounting for time lags.

112 Despite its effectiveness in capturing detailed characteristics of microseismic
113 event populations, multiplet clustering analysis has several inherent limitations. Selecting
114 appropriate seismic stations and frequency bands often requires extensive trial and error.
115 In addition, obtaining interpretable clustering results often requires repeatedly computing
116 waveform similarity for all event pairs, which might be computationally expensive for
117 recent growing data set. Furthermore, because this method evaluates event similarity
118 solely based on waveform characteristics, it may not adequately reflect the temporal and
119 spatial occurrence patterns of microseismic events. Time series microseismic nucleation
120 in each cluster sometime exhibit intermittent seismic activity from clusters, which
121 sometime is difficult interpret at the injection condition (Asanuma et al. 2008)

122 In this study, we tested a data-driven microseismic clustering framework based
123 on machine learning clustering techniques and to establish a process for obtaining
124 clustering results that are physically interpretable as subsurface fracture systems. While
125 traditional multiplet clustering relies on a single explanatory variable, waveform
126 similarity, we introduce additional explanatory variables representing spatial and
127 temporal relationships between events. By evaluating inter-event similarity from multiple

perspectives, this study seeks to achieve more robust, physically meaningful
microseismic clustering results. This study tests performance of several proposed
clustering methods on standard seismic catalog (absolute location) and refined seismic
catalog (relocated) and explores its efficacy as preprocess before the relocation analysis.

2 Data and Methods

2.1 Basel geothermal field and microseismic data

In this study, we analyze microseismic data recorded during the EGS
development project at Basel, Switzerland. Basel is located in northern Switzerland at the
southern end of the Rhine Graben, near the borders with Germany and France. The Rhine
Graben region has a long history of active geothermal utilization, including well-known
sites such as Soultz-sous-Forêts in France (Baria et al., 1999). It is recognized as one of
the most promising geothermal regions in Europe. At the Basel site, active geothermal
development was undertaken to establish a geothermal-based cogeneration system
(Häring et al. 2008).

In December 2006, the hydraulic stimulation was carried out at this site. During

144 this stimulation, continuous water injection was performed at the injection well Basel1
145 for approximately seven days; the injection well was drilled to a depth of 4,740 m in the
146 granitic basement. The maximum injection rate reached 3,000 L/min, and the total
147 injection volume amounted to approximately 11,000 m³. Microseismic activity continued
148 intermittently even after the termination of the hydraulic stimulation (Häring et al. 2008;
149 Mukuhira et al. 2013).

150 During the stimulation period, microseismic events were monitored using the
151 local seismic network shown in Figure 1, yielding approximately 14,850 events detection
152 (Asanuma et al. 2008; Dyer et al. 2008, 2010). Asanuma et al. (2007) first determined
153 absolute hypocenter location with manual picks using single event determination method
154 (SED). Hypocenter distribution of around 3000 events shows overall NNW-SSW trend
155 (Figure 1), but still it is cloud like distribution, where finding fracture network is very
156 challenging from this hypocenter location. We used this microseismic catalog based on
157 SED for the input of machine learning clustering as the data before the relocation

158 Asanuma et al. (2008) detected differential P- and S-wave arrival times by
159 comparing waveforms of events exhibiting high waveform similarity and performed

relative hypocenter relocation using the double-difference method (Waldhauser and Ellsworth 2000). As a result, the hypocenters of 2,849 events were successfully relocated. The spatial distribution of the relocated hypocenters is shown in Figure 2. In addition, Asanuma et al. (2008) applied multiplet clustering analysis and classified the microseismic events into 100 multiplet clusters, as shown in Figure 2. Many of the clusters exhibit linear or planar geometries resembling fracture structures. Assuming each cluster as representing a fracture plane, principal component analysis was applied to estimate the normal vectors of the fracture planes as third principal component. Their normal vectors were then projected onto the lower hemisphere, as shown in Figure 2. This projection shows a concentration of poles in the northeast–southwest direction, indicating that many clusters are oriented NNW–SSE. Furthermore, the presence of numerous points plotted near the circumference of the projection suggests that many clusters have steep, nearly vertical dips. These characteristics are consistent with the known geological and *in-situ* stress conditions of the Basel geothermal field and existing fractures obtained from borehole logging (Valley and Evans 2019; Ziegler and Evans 2020), indicating that the clustering results reasonably reflect the fracture system of the study area. We set this

physically interpretable result as a benchmark for clustering results. We also used this relocated catalog too as input of machine learning clustering.

2.2 Unsupervised machine learning methods for microseismic clustering

In this study, we employed following three representative and easy-implemented unsupervised machine learning clustering methods (K-means, and Agglomerative Hierarchical Clustering). Then, we compare the results of clustering from each methods by introducing appropriate pre-processing to explore the feasibility of this approach.

2.2.1 K-means Clustering

The K-means algorithm (e.g., Lloyd, 1982) is a widely used unsupervised clustering method that partitions a dataset into k predefined clusters. The objective of this method is to minimize the sum of distances from each data point to its assigned cluster's centroid. Clustering is achieved through an iterative procedure consisting of two steps: assigning data points to their nearest cluster centroid and subsequently updating the cluster centroids. However, K-means clustering requires the number of clusters k to be

specified in advance, and the choice of k can strongly influence the clustering results. In addition, because the method assigns all data points to clusters, outliers or noise may lead to inappropriate or distorted cluster structures. Another important characteristic of the K-means algorithm is its sensitivity to the initialization of cluster centroids. K-means clustering (MacQueen, 1967) implemented in MATLAB (MathWorks Inc.) was used in this analysis. K-means relies on only single parameters, and this does not fit the objective of this study. But still, we try this method to evaluate how unsupervised machine learning clustering method works on microseismic clustering problems.

2.2.2 Agglomerative Hierarchical Clustering (AHC)

Agglomerative hierarchical clustering is a hierarchical clustering method that treats each data point as an individual cluster and progressively merges clusters based on a dissimilarity measure (distance) (Ward, 1963). The cluster-merging process can be represented as a dendrogram, which enables visualization of the hierarchical structure of clusters and the sequence of merging steps. A threshold on the constructed dendrogram determines the final clusters at an appropriate level. The threshold is commonly

determined either by specifying the desired number of clusters in advance or by identifying a significant increase in dissimilarity in the dendrogram. Thus, this method allows the clustering structure to be determined without fixing the number of clusters beforehand. The clustering by AHC was performed using the MATLAB (MathWorks Inc.) function with a user-defined similarity metric.

3 Analysis and Results

3.1 Explanatory parameters

In this study, the following variables were used as indicators of general similarity between microseismic events:

- Cross correlation coefficient of the waveform between the events.
- Relative spatial parameters computed from the relocated hypocenter locations.
- Relative time difference between the events based on the trigger time.

We computed normalized correlation coefficient on waveforms of all pair of the events. Normalized correlation coefficient was computed for the whole 6 sec and 1000

224 Hz waveforms of each three components of station Otterbach2 (OT2) and Riehen2 (RIE2).
225 OT2 and RIE2 recorded relatively clean waveforms, and these stations were used for
226 waveform-similarity evaluation in previous studies (Asanuma et al. 2008; Herrmann et
227 al. 2019). Maximum of normalized correlation coefficient values from three component
228 was averaged. As forementioned, waveform similarity suggest similar source
229 characteristics and ray path. Thus, waveform similarity leads the events originated from
230 same fractures.

231 The Euclidean distances between hypocenter locations of microseismic events
232 were simply computed. The Euclidean distance directly represents the spatial proximity
233 between events and was used to capture the geometric characteristics of local event
234 concentrations and cluster structures in the hypocenter distribution. On the contrary,
235 larger Euclidean distance will not lead the events into the same cluster.

236 To evaluate temporal similarity, we used the time difference between the event
237 occurrences. The occurrence time was defined as the trigger time recorded upon
238 microseismic detection. Microseismic events from the same fractures can occur at
239 different times due to pore pressure migration and asperity distribution in each fracture.

240 Still, the time difference between events within the same cluster can be to be smaller and
241 cannot be very large.

242 Among the explanatory variables used in this study, waveform similarity is
243 measured by the normalized cross correlation coefficient, with values closer to 1
244 indicating greater similarity. In contrast, the spatial and temporal variables are distance-
245 based measures, with smaller values indicating greater proximity between events.
246 Therefore, waveform similarity was converted into a dissimilarity measure. Specifically,
247 the cross correlation coefficient was normalized to the 0 – 1 range, and a flip operation (1
248 – normalized cross correlation coefficient) was applied so that greater similarity
249 corresponds to smaller values. Through this transformation, all explanatory variables can
250 be interpreted consistently, with smaller values indicating higher similarity.

251 These three parameters were organized into matrices, and an integrated matrix
252 was constructed by taking their unweighted simple average avoiding bias on specific
253 variables and to evaluate the influence of each variable on the clustering structure in an
254 unbiased manner (Figure 3 and 4).

255 Furthermore, to treat variables with different scales—waveform, spatial, and

256 temporal parameters—in a more unified manner and to remove the effects of the
257 distributions of the three variables (see Fig. S1), the Aligned Rank Transform (ART) was
258 applied (Wobbrock et al. 2011a). The ART method converts variables based on their
259 relative ranks, thereby reducing the influence of outliers and mitigating differences in
260 scale. Through this transformation, waveform, spatial, and temporal variables can be
261 treated on an equivalent basis in the clustering analysis.

262 For more detailed spatial-similarity evaluation, angular components (elevation and
263 azimuth angles) in a spherical coordinate system centered on the fluid feed point in
264 injection well Basell were also used. Evaluations based solely on distance information
265 may not adequately extract linear or planar cluster structures aligned along fault planes,
266 since one-dimensional distance lacks angular information. The angular components
267 introduced in this study capture the relative directional relationships among events and
268 emphasize structural alignments inherent in the hypocenter distribution. In particular, the
269 angles in the spherical coordinate system represent the spatial directivity of seismic
270 activity observed from the fluid injection point. They may reflect the propagation of pore
271 pressure and the direction of fracture growth. In the dissimilarity matrix, relative angle

between all pairs are stored after normalization (Figure 3). Note that this operation is based on the assumption that the microseismic cloud migrated from the injection point and extended in the direction to the maximum principal stress, which has been confirmed by previous studies (Häring et al. 2008; Mukuhira et al. 2023). Therefore, measuring the angular information from injection point is reasonable implication for this problem. The Euclidean distance and angular components capture the same spatial structure from different perspectives and can therefore be regarded as complementary. Accordingly, in this study, the distance and angular components were treated as a single spatial measure without assigning separate weights. Using the above preprocessing steps, clustering was performed to ensure that waveform, spatial, and temporal variables were represented in a balanced manner (Figure 4).

3.2 Clustering result by K-means

In this section, the K-means clustering results, based solely on the Euclidean distance between microseismic events, are presented. This analysis was conducted as a

288 preliminary investigation before applying the proposed data-driven clustering method,
289 which integrates multiple explanatory variables. The purpose of this preliminary analysis
290 is to evaluate how consistent the results are with previous analyses based solely on
291 hypocenter spatial information after relative relocation.

292 In this analysis, 1442 single events from a previous study (Asanuma et al. 2008;
293 Kumano 2008; Mukuhira et al. 2021) were excluded from the total dataset of 2849 events,
294 remaining 1407 events for clustering. Because the K-means algorithm requires the
295 number of clusters, k , to be specified in advance, $k=100$ was used, following the
296 configuration of the previous study.

297 Figure 5 shows the clustering results visualized as three orthogonal projections
298 and a lower hemisphere stereographic projection. The three-view projections show
299 several clusters exhibiting fracture-like geometries that are visually similar to those
300 reported in the previous study as shown in Figure 2 (Asanuma et al. 2008; Kumano 2008;
301 Mukuhira et al. 2021). In the lower hemisphere projection, a concentration of poles to the
302 clusters is also observed near the EN–WS direction along the circumference, which is
303 qualitatively consistent with the trend reported in the previous analysis (Asanuma et al.

2008; Kumano 2008; Mukuhira et al. 2021). Note that this analysis condition is ideal, and single events were removed in advance.

These results indicate that when relocated locations are used, clustering based solely on Euclidean distance by K-means can reproduce cluster structures broadly consistent with those obtained by conventional multiplet clustering (Asanuma et al. 2008; Kumano 2008; Mukuhira et al. 2021). However, the clustering results for hypocenter locations before relocation were not interpretable, as shown in Figure S3, where hypocenter location determined by SED method (Asanuma et al. 2007) was used. In addition, when the dataset included single events, the K-means clustering results were not interpretable (Figure S3).

3.3 Clustering result by Agglomerative Hierarchical Clustering (AHC) based on arithmetic mean matrix

Agglomerative hierarchical clustering (AHC) was performed on a dissimilarity matrix constructed as the simple average of waveform similarity, inter-event distance, and inter-event time of microseismic events. Target events are all microseismicity that

320 occurred before the 13th of December, 2006, approximately one week after the injection
321 stopped. In this analysis, we limited the target event by its occurrence time because the
322 inter-event time was considered an explainable parameter and occurrence time here after
323 went very long due to prolonged microseismic activity (Mukuhira et al. 2013). In the
324 injection case, microseismicity is primarily driven by pore-pressure migration; therefore,
325 the time constant of pore-pressure migration may control its occurrence within the same
326 fractures in some degree. Note that in AHC method, clusters with fewer than five events
327 were identified as single events.

328 Under these conditions, microseismic events were divided into 49 clusters, with
329 36 events (1.4% of the target events) classified as single events; this proportion is much
330 lower than that observed in the conventional analysis, although comparison is not fair as
331 both approach employed different approach in algorithm. Figure 6 shows the resulting
332 cluster structures visualized as three orthogonal projections and a lower hemisphere
333 stereographic projection. The visualization results showed several clusters exhibiting
334 fracture-like geometries, but many were at a larger scale and had larger populations than
335 those in the previous study. Each cluster contains around 50 events. The lower hemisphere

projection also shows a concentration of cluster poles in the EN–WS direction, consistent with the trend of the previous analysis (Kumano et al., 2008). But the distribution appears somewhat more dispersed, as seen in the middle of the lower hemisphere, suggesting inconsistency in the *in-situ* stress field.

3.4 Clustering result based on aligned rank-transformed matrix

We also employed the AHC, but on a dissimilarity matrix constructed using the Aligned Rank Transform (ART) (Wobbrock et al. 2011b). In 3.3, we used a dissimilarity matrix that was constructed by simply normalizing and averaging waveform similarity, inter-event distance, and inter-event time. However, the resulting cluster dimensions, shapes, and orientations, including strike and dip, exhibited greater variability than those reported in a previous study (Kumano et al., 2008). As a possible cause, the distributions of dissimilarity values differ substantially among the explanatory variables (Figure S2). Therefore, a simple summation of these variables yields biased dissimilarity distributions. At least, the contribution to dissimilarity from some variables can be dominant, undermining the use of multi-explainable variables. To avoid this situation, the Aligned

352 Rank Transform (ART) was applied to evaluate similarity based on the relative ordering
353 of values rather than the absolute magnitudes of the explainable variables (Wobbrock et
354 al. 2011a). The ART was applied to the dissimilarity matrices of each waveform similarity,
355 inter-event distance, and inter-event time. Then, AHC was performed on the combined
356 dissimilarity matrix obtained by averaging the three dissimilarity matrices after ART.
357 Analysis conditions are the same as in 3.3.

358 Figure 7 shows the resulting cluster visualized as three orthogonal projections
359 and a lower hemisphere stereographic projection. Under this condition, the events were
360 divided into 78 clusters, with 46 events (1.8% of target events) classified as single events;
361 this represents a still much smaller proportion than in the previous study (Kumano et al.,
362 2008). Many clusters are on a slightly smaller scale than those in the previous section
363 (3.3) but are still slightly larger than those in the previous study, with correspondingly
364 more populations. Each cluster contains around 32 events. The projections show several
365 clusters exhibiting fracture-like geometries similar to those reported in a previous study
366 (Kumano et al., 2008). The lower hemisphere projection also shows a concentration of
367 points in the EN–WS direction, consistent with a previous study (Kumano et al., 2008)

and the *in-situ* stress state. However, the still-pole distribution appears more dispersed. In particular, many clusters exhibit relatively small dip angles.

3.5 Clustering result based on aligned rank-transformed matrix including angular information

In this section, the results of agglomerative hierarchical clustering based on a dissimilarity matrix incorporating angular information are presented. Angular components (elevation and azimuth) in the spherical coordinate system were adopted as explanatory variables (see 2.1), and the angular difference between microseismic events was defined as a measure of dissimilarity. A smaller angular difference indicates that the spatial directions of microseismic events, as viewed from the fluid injection point, are more similar. As with the other explanatory variables, the Aligned Rank Transform (ART) was applied to the angular variables. Angular information reflects the global spatial structure of the hypocenter distribution with respect to directionality relative to the fluid injection point. Therefore, in this study, a spatial dissimilarity matrix was constructed by taking the simple average of the Euclidean-distance-based dissimilarity matrix and the

angular-difference-based dissimilarity matrices derived from azimuth and elevation. Then, an integrated dissimilarity matrix was obtained by averaging the spatial dissimilarity matrix with those based on waveform similarity and on occurrence-time differences. AHC was then performed on this integrated dissimilarity matrix. Clusters containing fewer than five events were treated as single events.

Figure 8 shows the resulting cluster visualized as three orthogonal projections and a lower hemisphere stereographic projection. Under this condition, the events were divided into 77 clusters, with 58 events (22 % of target events) classified as single events. Again, the clustering result shows a lower proportion of single events than in the previous study (Asanuma et al. 2008; Kumano 2008). However, the cluster dimension and cluster population might be optimized and converge to the results of the previous study. Each cluster contains around 32 events. Many clusters are slightly smaller than other clustering results in this study, but are still slightly larger than those in the previous study.

The three orthogonal projections reveal several clusters exhibiting fracture-like geometries similar to those reported in a previous study. The lower hemisphere projection

also shows a concentration of points in the EN–WS direction, consistent with a previous study and *in-situ* stress state . indicating that fractures with clearly different orientations are relatively limited. In addition, the dispersion of points is smaller than that observed with the conventional method.

However, the still-pole distribution appears more dispersed. In particular, many clusters exhibit relatively small dip angles.

4 Discussion

We have started data driven microseismic clustering with unsupervised machine learning from K-means on solely spatial information of absolute hypocenter location (Figure S3) and relocated hypocenter location (Figure 5 and Figure S4). The biggest challenge with K-means approach is to predefine cluster numbers. As we used known cluster number of 100, K-means clustering performance on relocated hypocenter showed equivalent performance to that by previous study. With the possible cluster number, K-means clustering might lead plausible clustering results, although this approach still increase arbitrariness. In addition, K-means did not identify single events

and sort all events into some clusters. K-means clustering on all hypocenters (single events included) showed the clusters solely depending on the Euclidean distance. That led homogeneously distributed pole plots in Figure S4, suggesting unsuccessful identification on fracture shaped clusters. Thus, K-means methods relying on solely Euclidean distance between the events cannot perform very well unless the seismic catalog was well defined with relocation techniques and reasonable pre-known cluster number is available. Therefore, in our problem, K-means method performs limitedly and we somehow confirmed the importance of our approach of multi physics based clustering.

As next step, we introduced waveform similarity, and inter-event time in addition to the spatial information and employed AHC method for clustering, as multi physics based data driven clustering approach. Three input parameters initially had different distribution (Figure S2), resulting in unsatisfiable clustering results (Figure 6). Introduction of ART significantly improved the clustering results (Figure 7). This result provided more realistic results than the case of K-means algorithm and showed similar performance to the clustering results by previous study. But, still pole distribution in Figure 7 suggested existences of the horizontal fractures, which would is not consistent

432 with *in-situ* stress.

433 We further introduced relative spatial information as relative angular
434 information. Dissimilarity matrices for spatial information (distance, azimuth,
435 inclination) are averaged before merged with other information, so that the weight on
436 spatial information is same. The result of clustering (Figure 8) demonstrated similar
437 clustering results with that of previous work at a glance. Also, pole plots situated more
438 circumstantial region, suggesting the consistency with in-situ stress condition, although
439 there are some horizontal fractures. Apparently, relative angular information made a
440 difference in clustering performance. In principal, azimuth information enhances the
441 detectability of oriented hypocenters, and inclination information provided vertical
442 expansion limit of clusters. Thus, AHC methods works very well in microseismic
443 clustering problem if well located hypocenter catalog is available. Figure 9 shows the
444 time series evolution of identified clusters. Compared to the time series evolution of
445 clusters by previous study (Figure S1), cluster activation time were limited as less than 3
446 days due to the introduction of inter-event time. Microseismic cluster evolution detected
447 in previous research (Asanuma et al. 2008) shown in Figure S1 showed more prolonged

cluster activation, which sometime continued even after the end of stimulation. Shear slip induced by pore pressure did not need to occur simultaneously so that delayed shear slip can be acceptable phenomena, however, revealing delayed effect and interpretation are extremely challenging especially shut-in phase (Mukuhira et al. 2017; Kim et al. 2022; Boyet et al. 2023). Therefore, the clustering activity of our current result can be interpretable result that removing uncertain prolonged clustering activity. It is very hard to rule out the possibility of occurrence of delayed seismicity as real phenomena. Proposed framework might over-segmented some events into different cluster due to the introduction of inter-event time. We have to treat this clustering result delineates the clusters that activated nearly during the stimulation.

In AHC, setting the threshold on the dendrogram highly affects clustering results and also involves human decision making process (see 2.2.2). We checked the stability of clustering results with AHC. In the best case, the threshold was determined based on rapid change in distance and convergence. Figure 10 shows the dendrogram of AHC at the best case. At a glance, very tiny clusters were converged each other and started forming bigger clusters around 6 in distance. To check the stability of our clustering result,

we changed the thresholds around 6 and checked the number of clusters as shown in Figure 11. The number of the cluster decreases with increase of threshold but the gradient changes around 6, which suggests clustering results stabilized. Note that due to the definition of single events (unclustered events), number of the cluster is small for smaller thresholds. Averaged number in clusters increases at the same rate but the range of the the number of the events significantly enlarged after 6, suggesting the cluster population trend changed. Consequently, threshold 6 is reasonable choice although there are some human related aspects.

Throughout our attempt data-driven microseismic clustering with machine learning methods, AHC performed as good as conventional waveform similarity based clustering with expert judgement, if appropriate input data (waveform similarity, spatial information, relative time information) are available as well as well located catalog. Therefore, as data driven clustering method, proposed way can be useful. However, there are still some technical limitations in this approach. In a current framework, single events (events that do not belong to any clusters) were only identified when number of events in cluster was less than 5. This is partially because the distribution of the explainable

480 parameters and ART implementation to manage the integration of different explanatory
481 variables into one dissimilarity matrix. For example, the threshold of similar waveforms
482 based on cross correlation based method is often 0.6-0.7 (Kwiatek et al. 2014; Darisma
483 et al. 2024), however, after ART such similarity information was lost and the events were
484 just ranked into the order of similarity. Similarity or dissimilarity matrix of waveform
485 similarity after ART lost the physical meaning. Therefore, more suitable matrix
486 integration way would be beneficial, such as distribution conversion of explainable
487 parameters. Also, it would be much better to introduce the algorithm to identify single
488 events independently. These technical challenges will be addressed in the study in next
489 phase.

490 Our attempt of multi-physics data driven microseismic clustering undergo
491 successfully and understand the condition of its best performance as on the high quality
492 relocated catalog. This is a bit paradoxical as clustering is usually performed in the
493 process of relocation in standard relocation procedure. One can relocate the events with
494 standard procedure and then conduct clustering with our methods. Our clustering method
495 can be installed into the relocation process and contribute the relocation routine.

496 Relocation in initially defined clusters by our method might provide more precise location

497 efficiently as relocation group can be initially limited.

498

499 Table 1 Summary of the clustering analysis

Analysis num.	Figure	Clustering method	Catalog quality	Explanatory variables	Number of cluster	Ave. num. of events in cluster	Percentage of single events
1	Figure S3	K-means	SED	Space	100 (predefined)	30.7	0.0
2	Figure 5	K-means	DD w/o single events	Space	100 (predefined)	28.5	0.0
3	Figure S4	K-means	DD w/ single events	Space	100 (predefined)	14.1	50.6
4	Figure 6	AHC	DD	Wave, Space, Time	49	50.6	1.4
5	Figure 7	AHC+ART	DD	Wave, Space, Time	78	31.6	1.8
6	Figure 8	AHC+ART	DD	Wave, Space (angle), Time	77	31.9	2.3
7	Figure 12	AHC+ART	SED	Wave, Space (angle), Time	71	37.5	1.3
8	Figure S5	AHC+ART	DD	Wave	37	67.5	0.6
9	Figure S6	AHC+ART	DD	Space (angle)	89	27.1	3.86
10	Figure S7	AHC+ART	DD	Time	98	25.5	0.4
11	Figure S8	AHC+ART	DD	Wave, Space (angle)	101	23.5	5.6
12	Figure S9	AHC+ART	DD	Wave, Time	48	52.1	0.5
13	Figure S10	AHC+ART	DD	Space (angle), Time	89	27.4	2.8

501 **5 Conclusions**

502 In this study, we proposed a multi-parameter data-driven clustering framework
503 based on unsupervised machine learning to characterize subsurface fracture systems from
504 microseismic activity in geothermal reservoirs. In contrast to conventional methods that
505 rely primarily on waveform similarity, our approach integrates waveform cross
506 correlation, spatial proximity, directional angular information, and temporal proximity
507 into a unified dissimilarity matrix using the Aligned Rank Transform (ART) and
508 Agglomerative Hierarchical Clustering (AHC). Applying this methodology to the
509 induced microseismicity dataset from the Basel EGS project in Switzerland yielded the
510 following key conclusions:

511 For the optimized condition—where relocated hypocenters were used and
512 spatial parameters were enhanced by incorporating elevation and azimuth angles relative
513 to the injection feed point—the proposed method identified 77 nearly vertical clusters.
514 These orientations are highly consistent with the in-situ stress field and show structural
515 agreement comparable to conventional multiplet analysis requiring expert judgment.

516 Incorporating inter-event time into the clustering matrix effectively suppressed

the clustering of temporally distant events, constraining cluster activation durations to less than three days and providing a clearer spatiotemporal representation of fracture growth.

This study investigated the potential of unsupervised machine learning into microseismic clustering problems as data-driven methodology. These findings demonstrate that the proposed framework serves as a robust, flexible, and physics-interpretable clustering tool for microseismic analysis. Furthermore, it holds strong potential for integration into hypocenter relocation workflows as an automated pre-clustering routine, establishing consistent event groupings prior to relative location refinement.

Competing interests

The authors declare that they have no competing interest.

Funding

This study is supported by JSPS KAKENHI Grant Number 25K01720.

533 **Authors' contributions**

534 YM designed this study. RT and YM contributed to data analysis. All authors participated
535 in discussion and interpretation. All authors contributed and approved the manuscript.

536

537 **Acknowledgements**

538 We thank to Takaaki Kawahito, Justin Rubinstein, and Takahiro Shina for meaningful
539 discussions and support.

540

541 **Authors' information**

542 RT: Graduate School of Environmental Studies, Tohoku University, Sendai, Japan.

543 YM: Institute of Fluid Science, Tohoku University, Sendai, Japan.

544 KO: Fukushima Renewable Energy Institute, National Institute of Advanced Industrial
545 Science and Technology, Koriyama, Japan.

546 HA: New Energy and Industrial Technology Development Organization (NEDO),

547 Kawasaki, Japan

548

549 **References**

- 550 Aden-Antoniów F, Frank WB, Seydoux L (2022) An Adaptable Random Forest Model
551 for the Declustering of Earthquake Catalogs. JGR Solid Earth
552 127:e2021JB023254. <https://doi.org/10.1029/2021JB023254>
- 553 Asanuma H, Kumano Y, Hotta A, et al (2007) Analysis of Microseismic Events from a
554 Stimulation at Basel, Switzerland. In: GRC Transaction. pp 265–270
- 555 Asanuma H, Kumano Y, Niitsuma H, et al (2008) Interpretation of Reservoir Structure
556 from Super-Resolution Mapping of Microseismic Multiplets from Stimulation at
557 Basel, Switzerland in 2006. In: GRC Transaction. Transactions for GRC, pp 65–
558 70
- 559 Baria R, Jung R, Garnish J (1999) European HDR research programme at Soultz-sous-
560 Forêts (France) 1987±1996
- 561 Bentz S, Martínez-Garzón P, Kwiatek G, et al (2019) Analysis of Microseismicity
562 Framing $M_L > 2.5$ Earthquakes at The Geysers Geothermal Field, California.
563 JGR Solid Earth 124:8823–8843. <https://doi.org/10.1029/2019JB017716>

564 Boyet A, De Simone S, Ge S, Vilarrasa V (2023) Poroelastic stress relaxation, slip stress
 565 transfer and friction weakening controlled post-injection seismicity at the Basel
 566 Enhanced Geothermal System. *Commun Earth Environ* 4:104.
 567 <https://doi.org/10.1038/s43247-023-00764-y>

568 Darisma D, Mukuhira Y, Okamoto K, et al (2024) Building the fracture network model
 569 for the Okuaizu geothermal field based on microseismic data analysis. *Earth
 570 Planets Space* 76:107. <https://doi.org/10.1186/s40623-024-02049-w>

571 Dyer BC, Schanz U, Ladner F, et al (2008) Microseismic imaging of a geothermal
 572 reservoir stimulation. *The Leading Edge* 27:856–869.
 573 <https://doi.org/10.1190/1.2954024>

574 Dyer BC, Schanz U, Spillmann T, et al (2010) Application of microseismic multiplet
 575 analysis to the Basel geothermal reservoir stimulation events. *Geophysical
 576 Prospecting* 58:791–807. [https://doi.org/doi:%2010.1111/j.1365-
 577 2478.2010.00902.x](https://doi.org/doi:%2010.1111/j.1365-2478.2010.00902.x)

578 Evans KF, Moriya H, Niitsuma H, et al (2004) Microseismicity and permeability

579 enhancement of hydrogeologic structures during massive fluid injections into
580 granite at 3 km depth at the Soultz HDR site: Induced seismicity and flow in deep
581 granite. *Geophysical Journal International* 160:389–412.
582 <https://doi.org/10.1111/j.1365-246X.2004.02474.x>

583 Feng P, Wang R, Zhang H, et al (2026) Pre-Existing Structures and Stress Variations
584 Jointly Control the Induced Seismicity in Enhanced Geothermal System of
585 Gonghe Basin, China. *JGR Solid Earth* 131:e2025JB033158.
586 <https://doi.org/10.1029/2025JB033158>

587 Gardner JK, Knopoff L (1974) Is the sequence of earthquakes in southern California, with
588 aftershocks removed, Poissonian? *Bulletin of the Seismological Society of*
589 *America* 64:1363–1367.
590 <https://doi.org/https://doi.org/10.1785/BSSA0640051363>

591 Häring MO, Schanz U, Ladner F, Dyer BC (2008) Characterisation of the Basel 1
592 enhanced geothermal system. *Geothermics* 37:469–495.
593 <https://doi.org/10.1016/j.geothermics.2008.06.002>

594 Herrmann M, Kraft T, Tormann T, et al (2019) A Consistent High-Resolution Catalog of
 595 Induced Seismicity in Basel Based on Matched Filter Detection and Tailored Post-
 596 Processing. JGR Solid Earth 124:8449–8477.
 597 <https://doi.org/10.1029/2019JB017468>

598 Kim K-I, Min K-B, Kim K-Y, et al (2018) Protocol for induced microseismicity in the
 599 first enhanced geothermal systems project in Pohang, Korea. Renewable and
 600 Sustainable Energy Reviews 91:1182–1191.
 601 <https://doi.org/10.1016/j.rser.2018.04.062>

602 Kim K-I, Yoo H, Park S, et al (2022) Induced and triggered seismicity by immediate stress
 603 transfer and delayed fluid migration in a fractured geothermal reservoir at Pohang,
 604 South Korea. International Journal of Rock Mechanics and Mining Sciences
 605 153:105098. <https://doi.org/10.1016/j.ijrmms.2022.105098>

606 Kumano Y (2008) Analysis of Characteristics of AE Multiplets Induced by Hydraulic
 607 Stimulation and Evaluation of Dynamic Behavior in Subsurface Fracture
 608 Networks. Ph.D., Tohoku University

609 Kwiatek G, Bulut F, Bohnhoff M, Dresen G (2014) High-resolution analysis of seismicity
 610 induced at Berlín geothermal field, El Salvador. *Geothermics* 52:98–111.
 611 <https://doi.org/10.1016/j.geothermics.2013.09.008>

612 Kwiatek G, Saarno T, Ader T, et al (2019) Controlling fluid-induced seismicity during a
 613 6.1-km-deep geothermal stimulation in Finland. *Sci Adv* 5:eaav7224.
 614 <https://doi.org/10.1126/sciadv.aav7224>

615 Lloyd S (1982) Least squares quantization in PCM. *IEEE Trans Inform Theory* 28:129–
 616 137. <https://doi.org/10.1109/TIT.1982.1056489>

617 Majer EL, Baria R, Stark M, et al (2007) Induced seismicity associated with Enhanced
 618 Geothermal Systems. *Geothermics* 36:185–222.
 619 <https://doi.org/10.1016/j.geothermics.2007.03.003>

620 McBrearty IW, Beroza GC (2025) Double difference earthquake location with graph
 621 neural networks. *Earth Planets Space* 77:127. [https://doi.org/10.1186/s40623-](https://doi.org/10.1186/s40623-025-02251-4)
 622 [025-02251-4](https://doi.org/10.1186/s40623-025-02251-4)

623 Moriya H, Nakazato K, Niitsuma H, Baria R (2002) Detailed Fracture System of the

624 Soultz-sous-Forêts HDR Field Evaluated Using Microseismic Multiplet
625 Analysis

626 Moriya H, Niituma H, Bria R (2003) Multiplet-Clustering Analysis Reveals Structural
627 Details within the Seismic Cloud at the Soultz Geothermal Field, France. Bulletin
628 of the Seismological Society of America 93:1606–1620.
629 <https://doi.org/10.1785/0120020072>

630 Mukuhira Y, Asanuma H, Niituma H, Häring MO (2013) Characteristics of large-
631 magnitude microseismic events recorded during and after stimulation of a
632 geothermal reservoir at Basel, Switzerland. Geothermics 45:1–17.
633 <https://doi.org/10.1016/j.geothermics.2012.07.005>

634 Mukuhira Y, Dinske C, Asanuma H, et al (2017) Pore pressure behavior at the shut-in
635 phase and causality of large induced seismicity at Basel, Switzerland. JGR Solid
636 Earth 122:411–435. <https://doi.org/10.1002/2016JB013338>

637 Mukuhira Y, Kumano Y, Asanuma H, et al (2021) Catalog of locations of microseismic
638 events accompanying the 2006-2007 fluid injection into the Basel Enhanced

639 Geothermal System

640 Mukuhira Y, Yang M, Ishibashi T, et al (2023) Scaling Microseismic Cloud Shape During

641 Hydraulic Stimulation Using In Situ Stress and Permeability. JGR Solid Earth

642 128:e2023JB026839. <https://doi.org/10.1029/2023JB026839>

643 Phillips WS (2000) Precise Microearthquake Locations and Fluid Flow in the Geothermal

644 Reservoir at Soultz-sous-Forêts, France. Bulletin of the Seismological Society of

645 America 90:212–228. [https://doi.org/https://doi.org/10.1785/0119990047](https://doi.org/10.1785/0119990047)

646 Pine RJ, Batchelor AS (1984) Downward Migration of Shearing in Jointed Rock During

647 Hydraulic Injections. International Journal of Rock Mechanics and Mining

648 Sciences 21:249–263. [https://doi.org/https://doi.org/10.1016/0148-](https://doi.org/10.1016/0148-9062(84)92681-0)

649 9062(84)92681-0

650 Reasenber P (1985) Second-order moment of central California seismicity, 1969–1982.

651 J Geophys Res 90:5479–5495. <https://doi.org/10.1029/JB090iB07p05479>

652 Ross ZE, Cochran ES, Trugman DT, Smith JD (2020) 3D fault architecture controls the

653 dynamism of earthquake swarms. Science 368:357–1361

654 Sawaki Y, Shiina T, Sagae K, et al (2025a) Fault Geometries of the 2024 Mw 7.5 Noto
 655 Peninsula Earthquake from Hypocenter-Based Hierarchical Clustering of Point-
 656 Cloud Normal Vectors

657 Sawaki Y, Shiina T, Uchide T (2025b) Steep intraplate reverse faulting adjacent to the
 658 hypocenter of the 1923 Kanto earthquake: the Mw 5.0 western Kanagawa
 659 earthquake in eastern Japan on 9 August 2024. *Earth Planets Space* 77:186.
 660 <https://doi.org/10.1186/s40623-025-02311-9>

661 Tester, J. W. (2006) The future of geothermal energy: Impact of enhanced geothermal
 662 systems (EGS) on the United States in the 21st century. Idaho National Laboratory,
 663 Idaho Falls

664 Trugman DT, Shearer PM (2017) GrowClust: A Hierarchical Clustering Algorithm for
 665 Relative Earthquake Relocation, with Application to the Spanish Springs and
 666 Sheldon, Nevada, Earthquake Sequences. *Seismological Research Letters*
 667 88:379–391. <https://doi.org/10.1785/0220160188>

668 Valley B, Evans KF (2019) Stress magnitudes in the Basel enhanced geothermal system.

669 International Journal of Rock Mechanics and Mining Sciences 118:1–20.
 670 <https://doi.org/10.1016/j.ijrmms.2019.03.008>

671 Waldhauser F, Ellsworth W (2000) A Double-Difference Earthquake Location Algorithm:
 672 Method and Application to the Northern Hayward Fault, California. Bulletin of
 673 the Seismological Society of America 90:1353–1368.
 674 <https://doi.org/10.1785/0120000006>

675 Waldhauser F, Ellsworth WL (2002) Fault structure and mechanics of the Hayward Fault,
 676 California, from double-difference earthquake locations. J Geophys Res 107:..
 677 <https://doi.org/10.1029/2000JB000084>

678 Ward JH (1963) Hierarchical Grouping to Optimize an Objective Function. Journal of the
 679 American Statistical Association 58:236–244.
 680 <https://doi.org/https://doi.org/10.2307/2282967>

681 Wilding JD, Zhu W, Ross ZE, Jackson JM (2023) The magmatic web beneath Hawai‘i.
 682 Science 379:462–468. <https://doi.org/10.1126/science.ade5755>

683 Wobbrock JO, Findlater L, Gergle D, Higgins JJ (2011a) The aligned rank transform for

684 nonparametric factorial analyses using only anova procedures. In: Proceedings of
685 the SIGCHI Conference on Human Factors in Computing Systems. ACM,
686 Vancouver BC Canada, pp 143–146

687 Wobbrock JO, Findlater L, Gergle D, Higgins JJ (2011b) The aligned rank transform for
688 nonparametric factorial analyses using only anova procedures. In: Proceedings of
689 the SIGCHI Conference on Human Factors in Computing Systems. ACM,
690 Vancouver BC Canada, pp 143–146

691 Zaliapin I, Ben-Zion Y (2020) Earthquake Declustering Using the Nearest-Neighbor
692 Approach in Space-Time-Magnitude Domain. JGR Solid Earth
693 125:e2018JB017120. <https://doi.org/10.1029/2018JB017120>

694 Ziegler M, Evans KF (2020) Comparative study of Basel EGS reservoir faults inferred
695 from analysis of microseismic cluster datasets with fracture zones obtained from
696 well log analysis. Journal of Structural Geology 130:103923.
697 <https://doi.org/10.1016/j.jsg.2019.103923>

698

699 **Figure legends**

700 Figure 1. Hypocenter distribution of microseismic events that occurred in the Basel EGS
701 field (Asanuma et al. 2007). Hypocenter location was determined with single event
702 determination method. The upper right panel also shows the microseismic monitoring
703 network and its depth.

704

705 Figure 2. Results of microseismic clustering by the previous study (Asanuma et al. 2008;
706 Kumano, 2008; Mukuhira et al. 2021). Clustering results are color-coded by hypocenter
707 location using DD methods. A different color represents each cluster. The top-right panel
708 shows the pole plots in the lower hemisphere. Poles to the fractures were estimated using
709 principal component analysis of the hypocenter locations of events in each cluster,
710 assuming that the clustering results detected the existing fractures.

711

712 Figure 3. Heatmap of dissimilarity matrices for all explanatory variables (Cross
713 correlation, distance, inter-event time, azimuth and elevation angle). Both axes represent
714 event ID. Diagonal parts are identically zero. Each value are diagonal symmetry.

715

716 Figure 4. Heatmap of integrated dissimilarity matrices. AHC is performed on these
717 matrices.

718

719 Figure 5. Clustering results by K-means in three orthogonal projections of map view (top
720 left), EW (bottom left), and NS (bottom right) cross sections. The top-right panel shows
721 the lower hemisphere projection of the poles of identified microseismic clusters.

722

723 Figure 6. Clustering results by AHC on the arithmetic mean-based dissimilarity matrix in
724 three orthogonal projections of map view (top left), EW (bottom left), and NS (bottom
725 right) cross sections. The top-right panel shows the lower hemisphere projection of the
726 poles of identified microseismic clusters. The color of the dots corresponds to the clusters.

727

728 Figure 7. Clustering results by AHC on the dissimilarity matrix estimated as the
729 arithmetic mean of ART applied to each explainable variable. Clustering results are
730 projected into map view (top left), and EW (bottom left) and NS (bottom right) cross

sections. The top-right panel shows the lower hemisphere projection of the poles of identified microseismic clusters. The color of the dots corresponds to the clusters.

Figure 8. Clustering results by AHC on the dissimilarity matrix estimated as the arithmetic mean of ART applied to each explainable variable. The clustering result is projected onto the map view (top left) and the EW (bottom left) and NS (bottom right) cross sections. The top-right panel shows the lower hemisphere projection of the poles of identified microseismic clusters. The color of the dots corresponds to the clusters.

Figure 9. Clustering results by AHC on the dissimilarity matrix estimated as the arithmetic mean of ART applied to each explainable variable. The clustering result is projected onto the map view (top left) and the EW (bottom left) and NS (bottom right) cross sections. The top-right panel shows the lower hemisphere projection of the poles of identified microseismic clusters. The color of the dots corresponds to the clusters.

Figure 10. Dendrogram plot as a result of AHC for the best case. The threshold for

747 clustering was visually determined as 6.

748

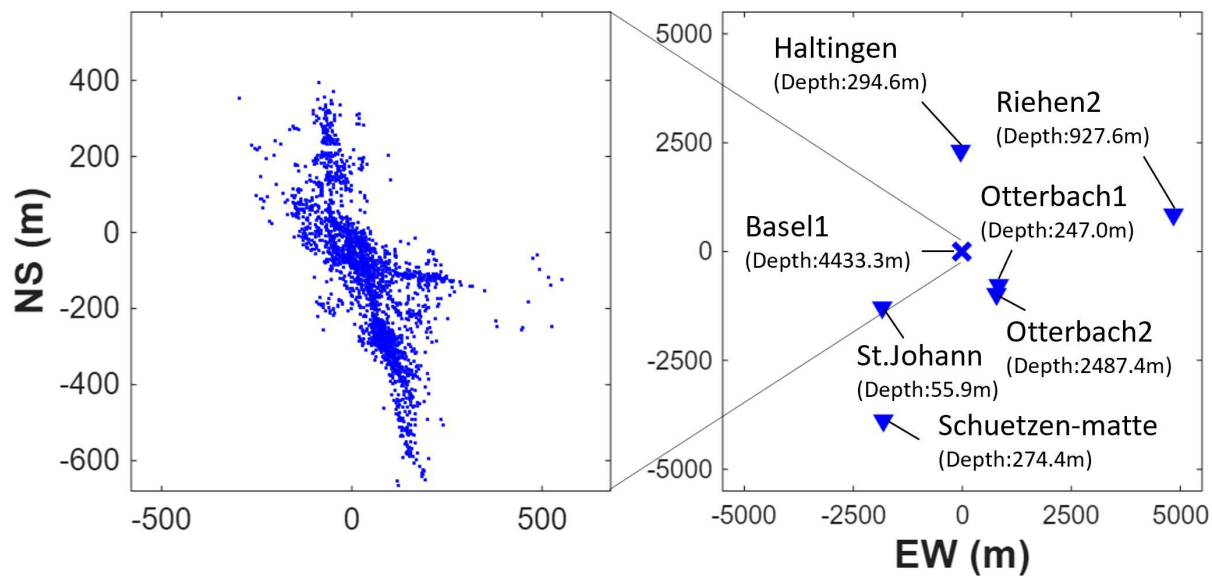
749 Figure 11. Stability of clustering results. Variation in number of clusters (top), average

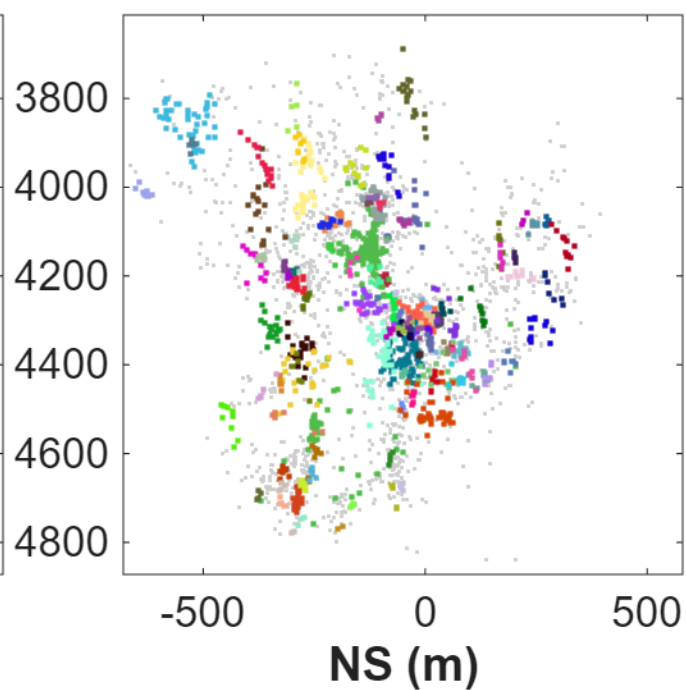
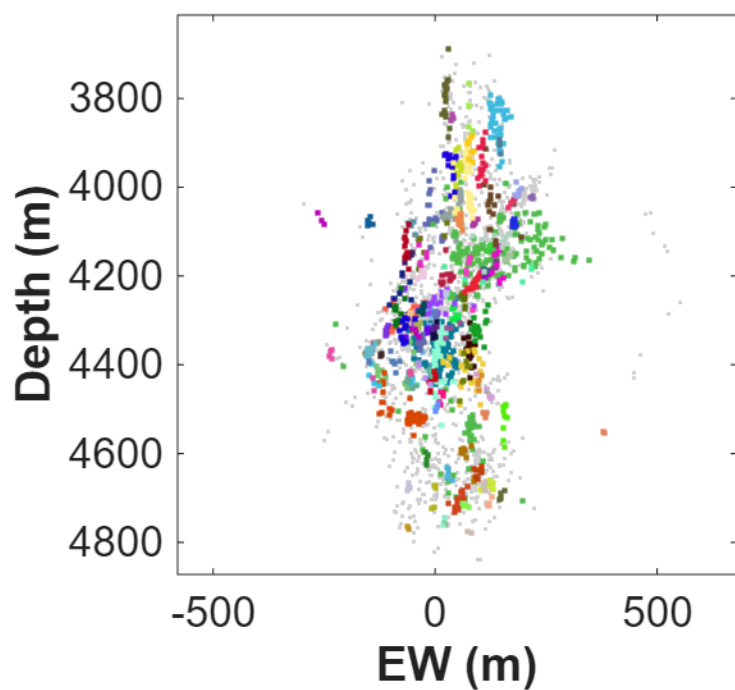
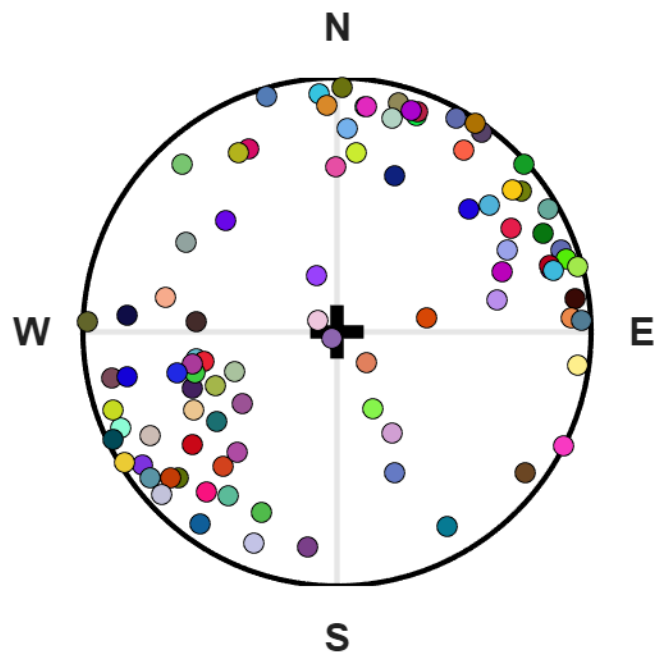
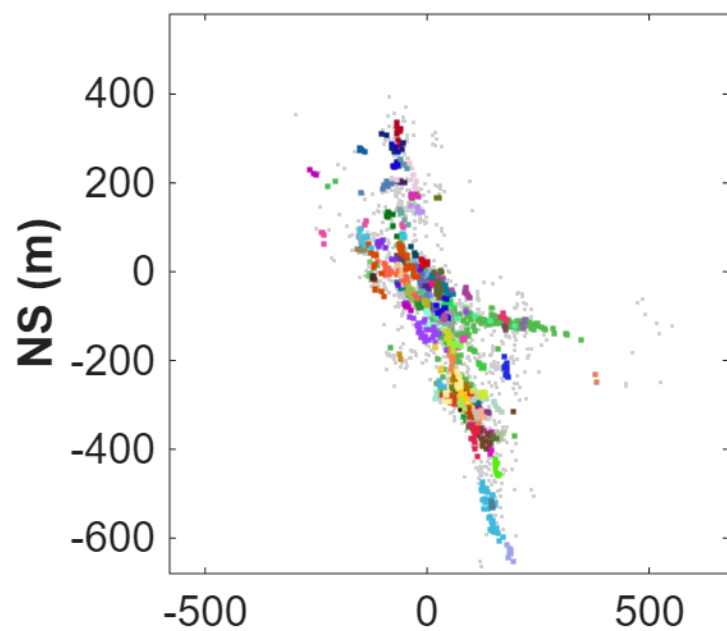
750 events number in a cluster (middle), and average inter invent distance in respect to various

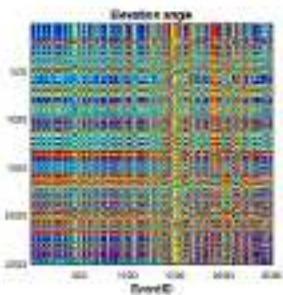
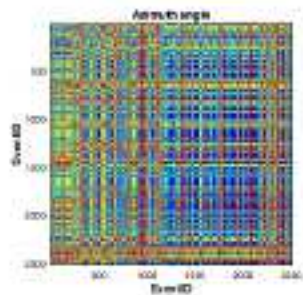
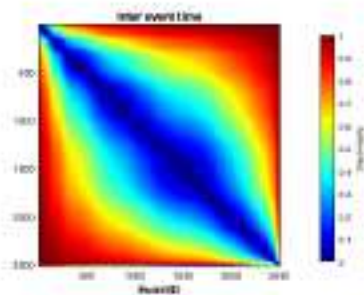
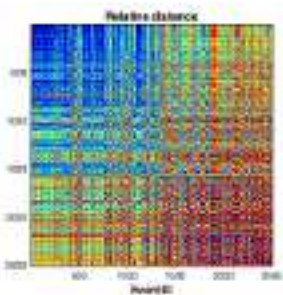
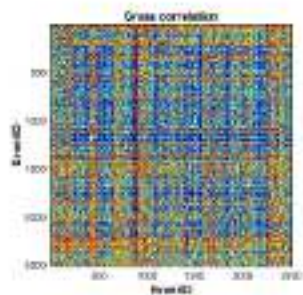
751 threshold.

752

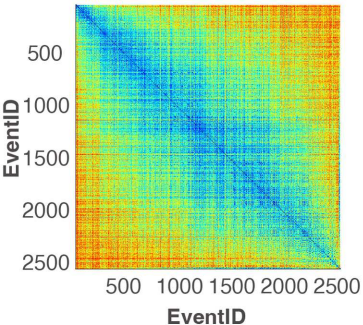
753 Table 1 Summary of the clustering analysis



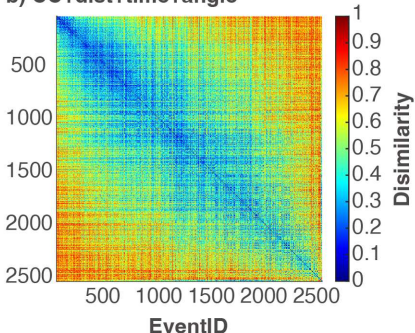


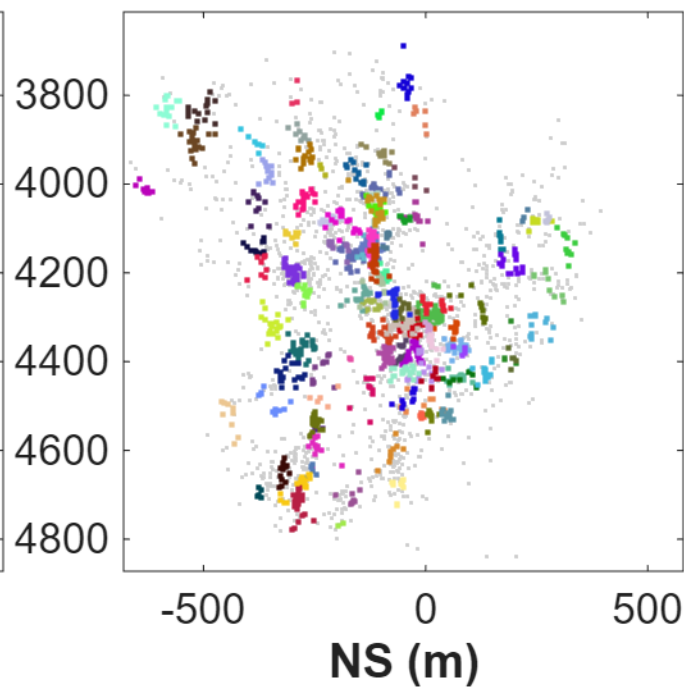
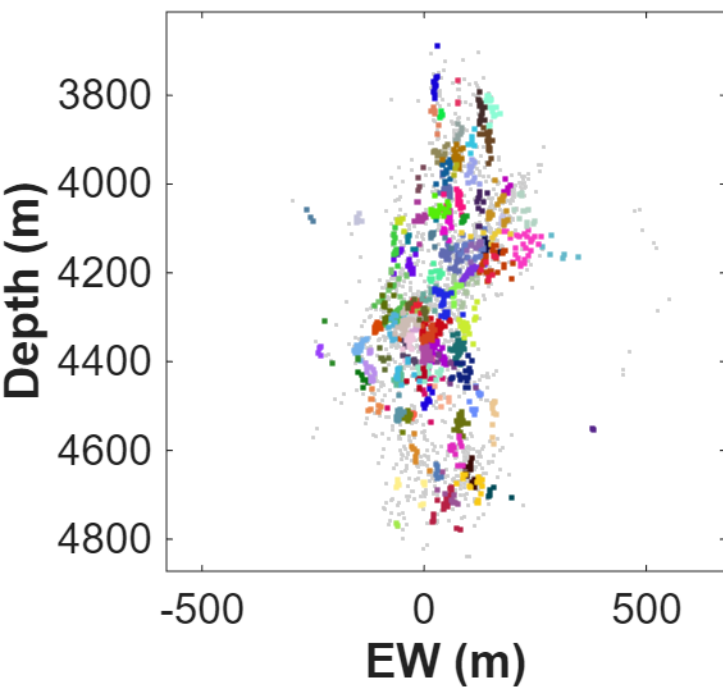
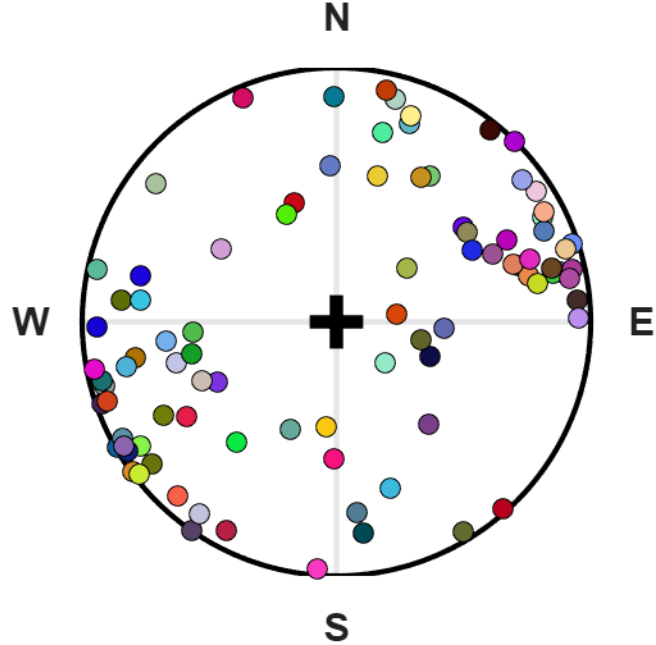
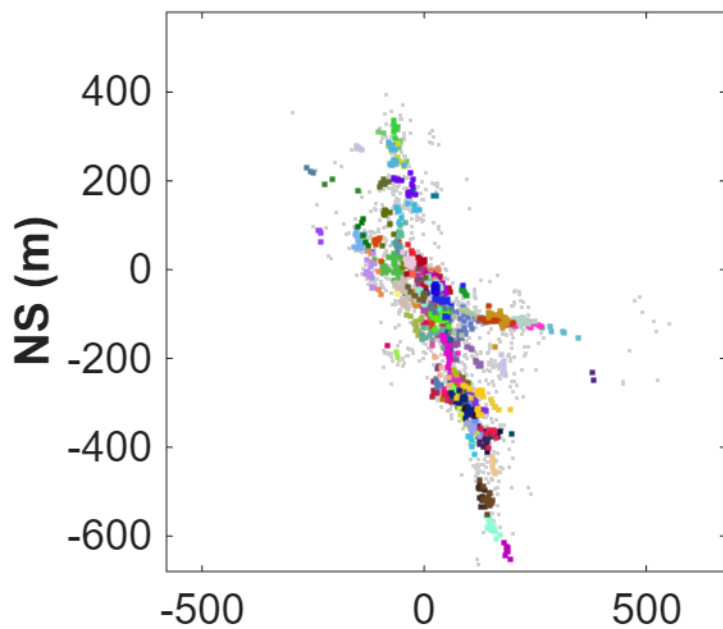


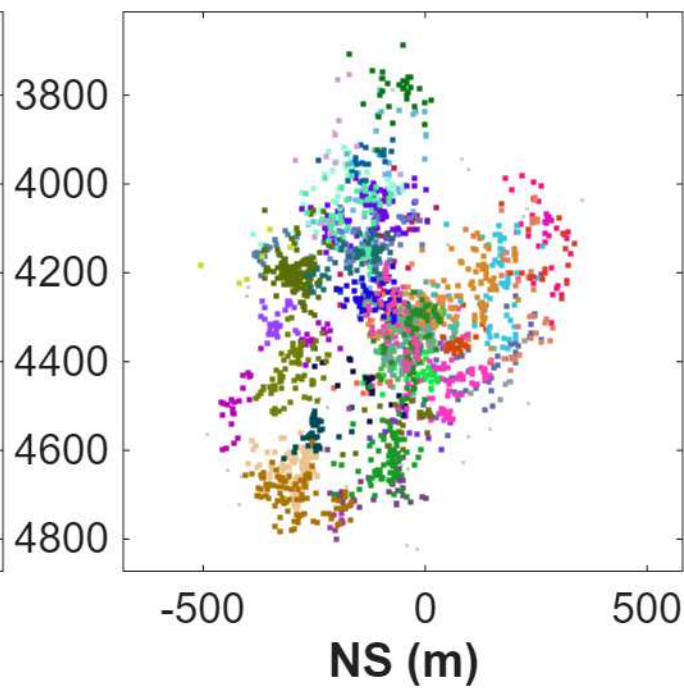
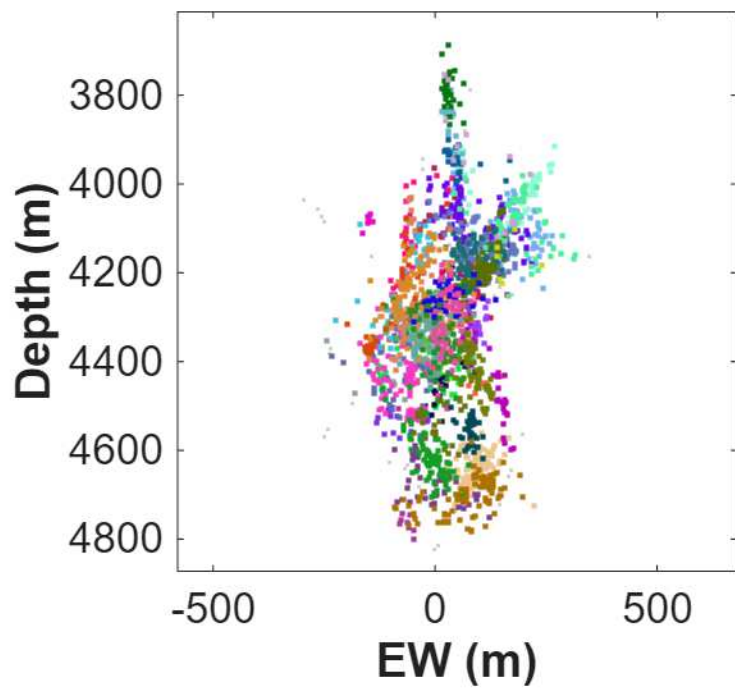
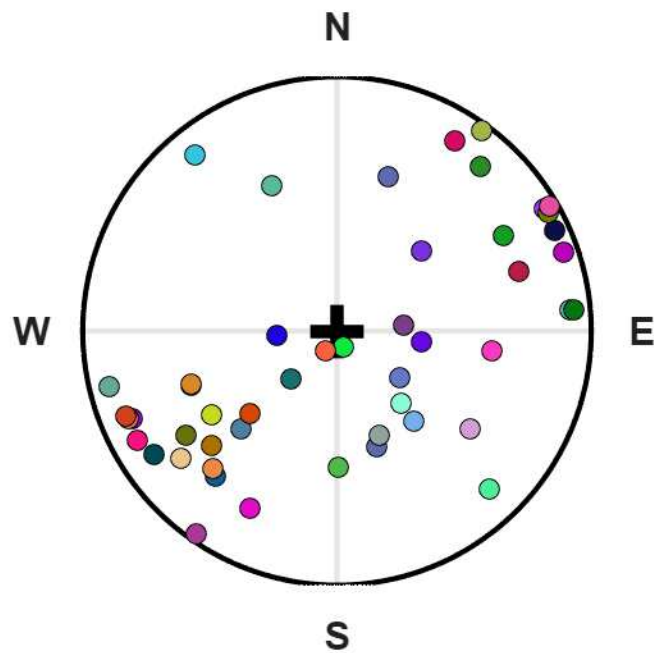
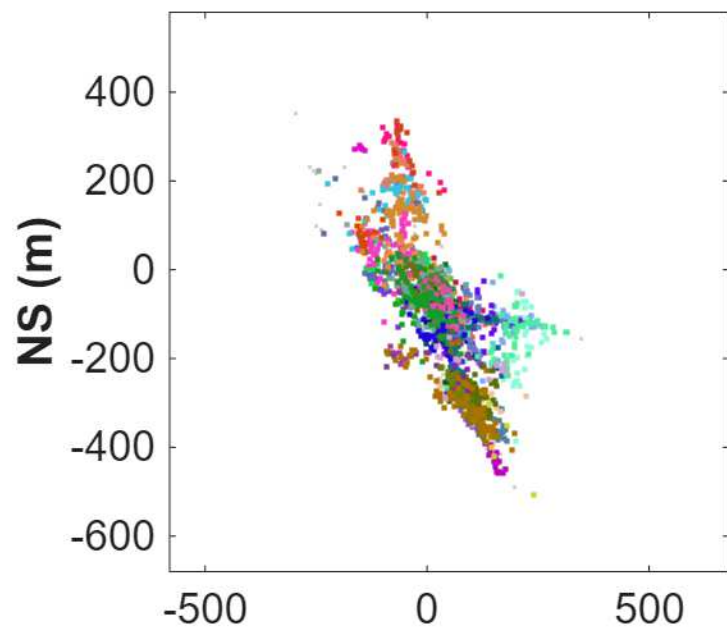
a) CC+dist+time

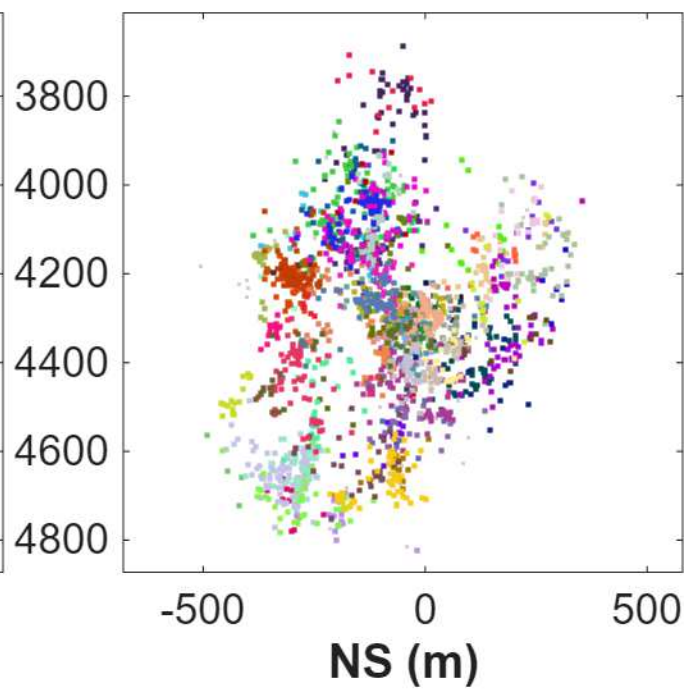
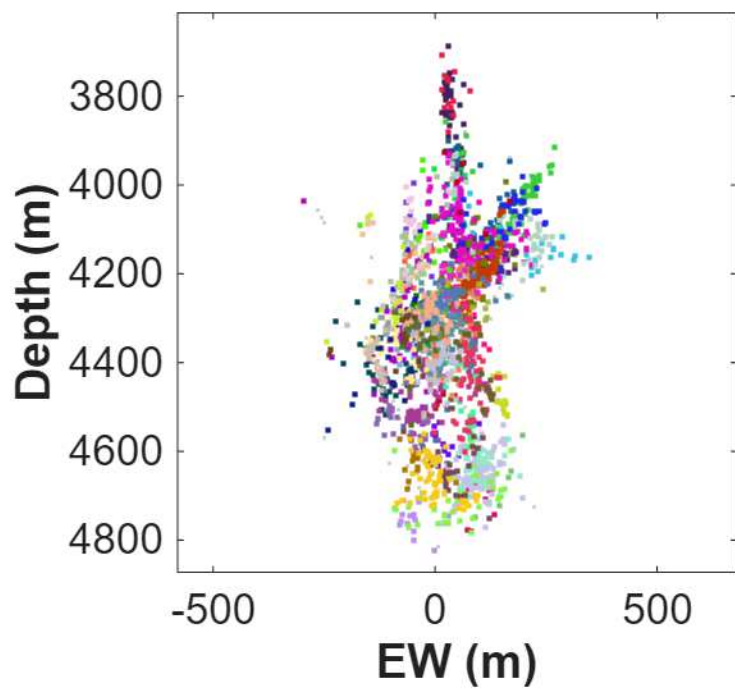
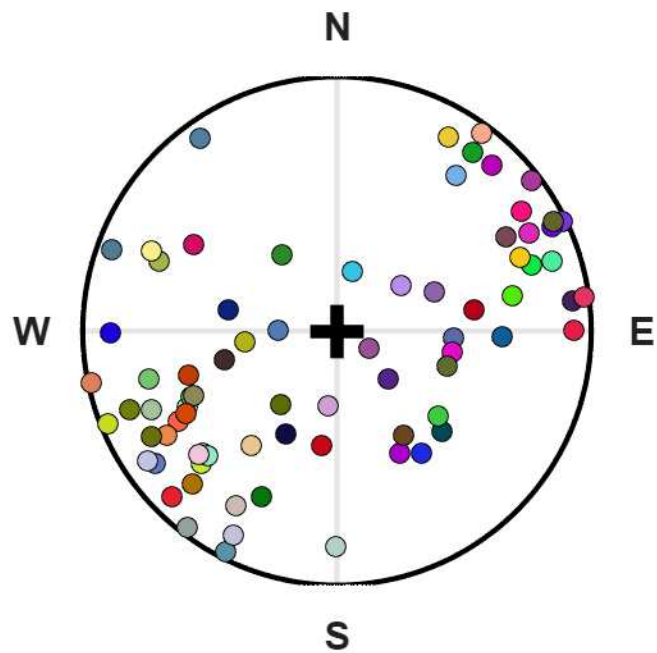
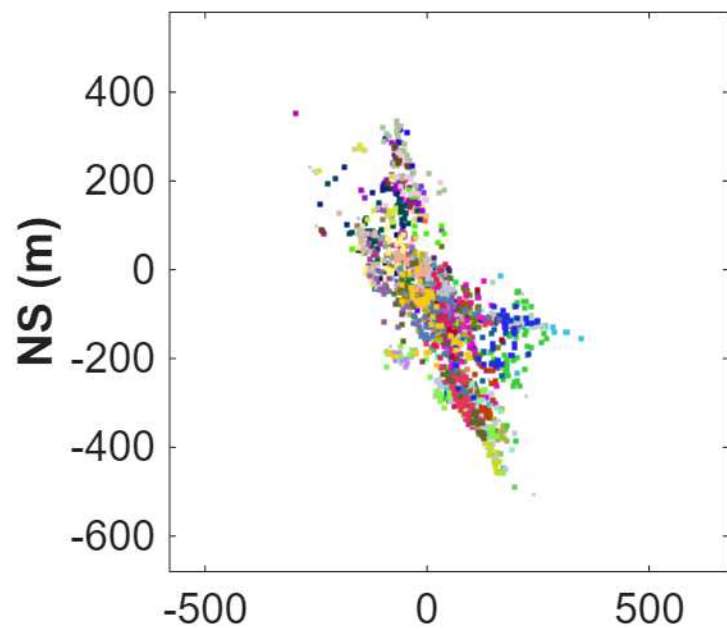


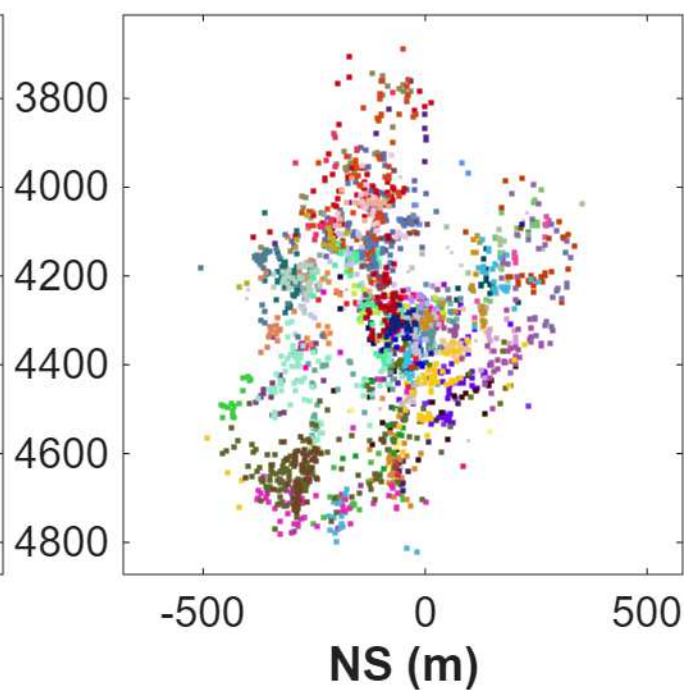
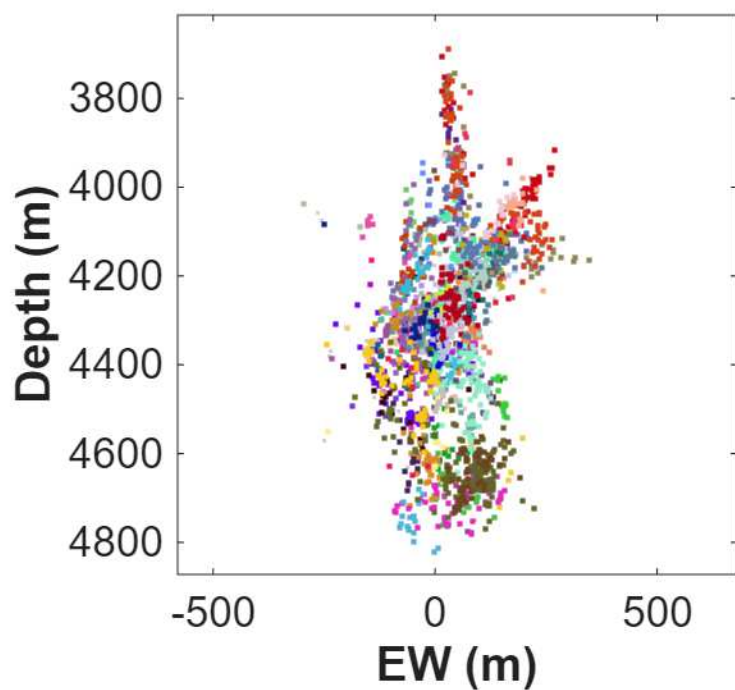
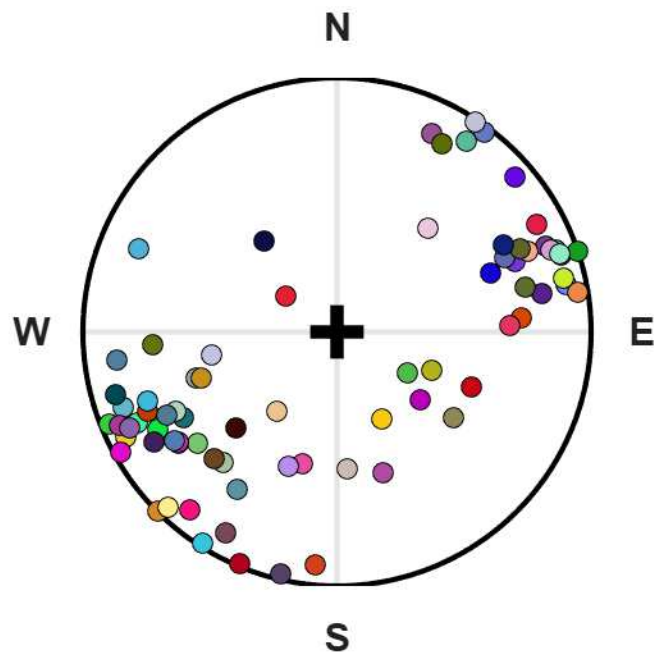
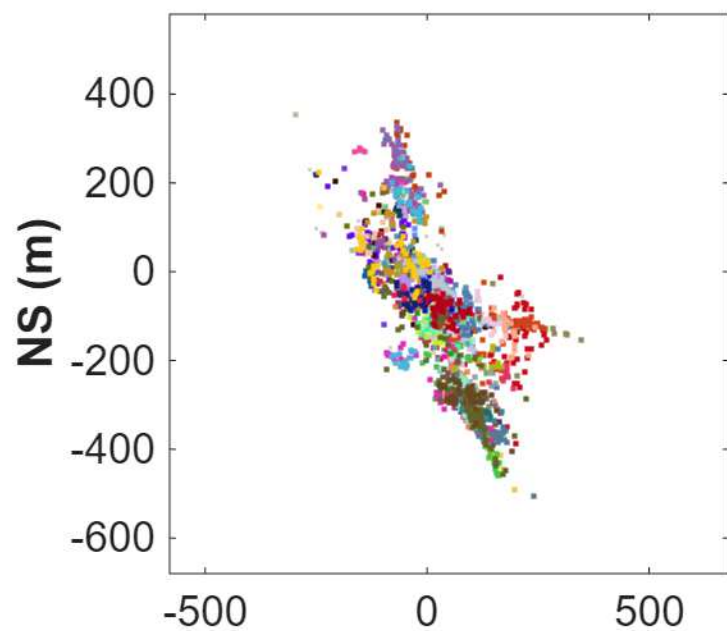
b) CC+dist+time+angle

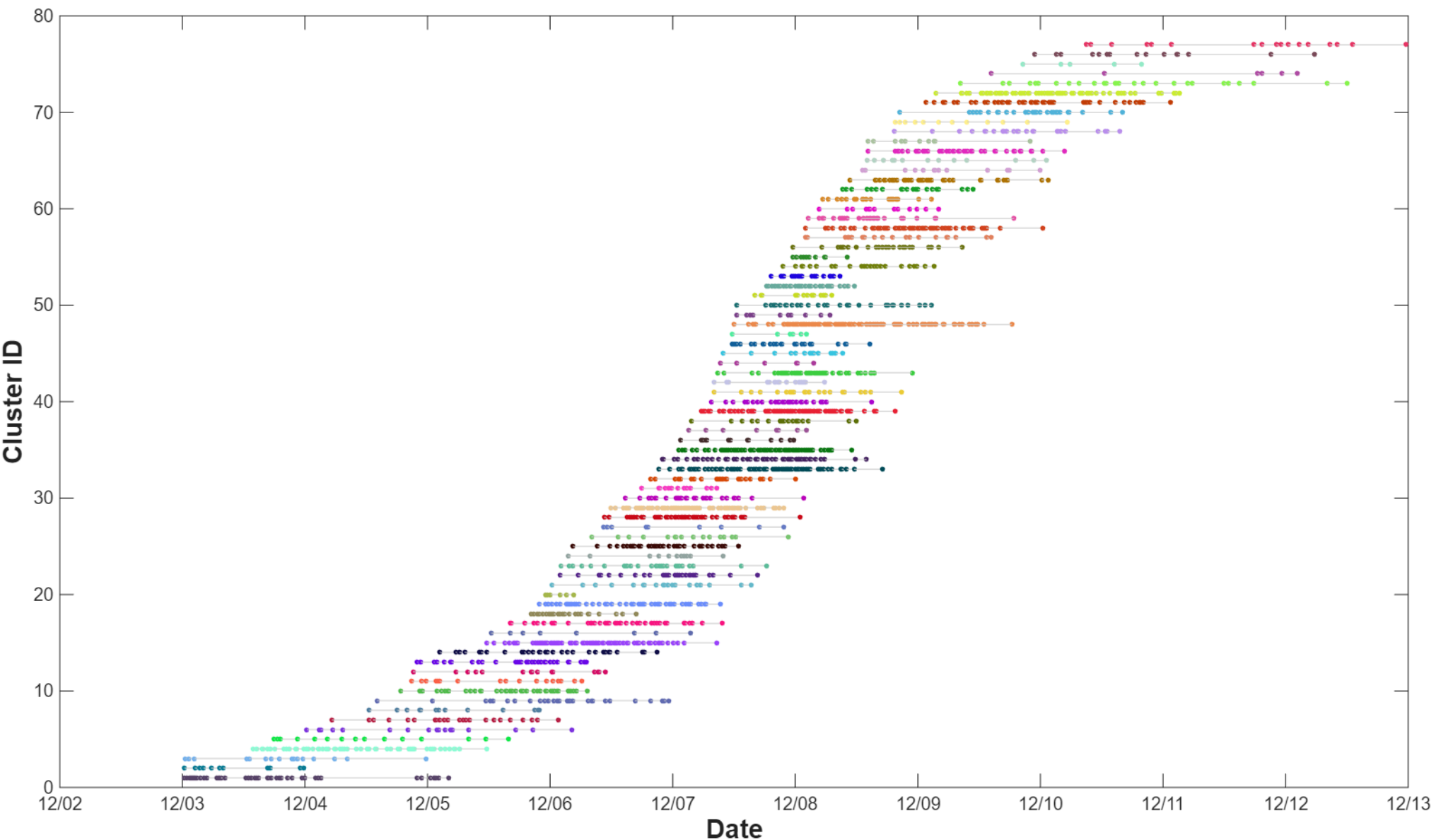
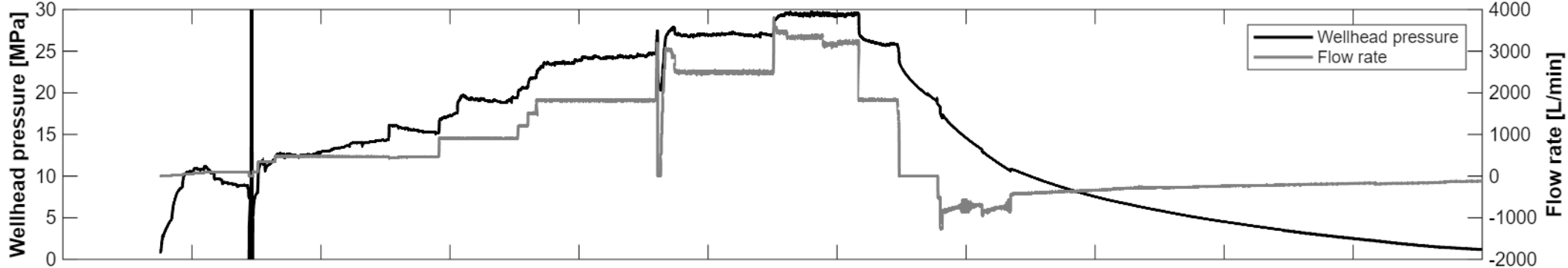




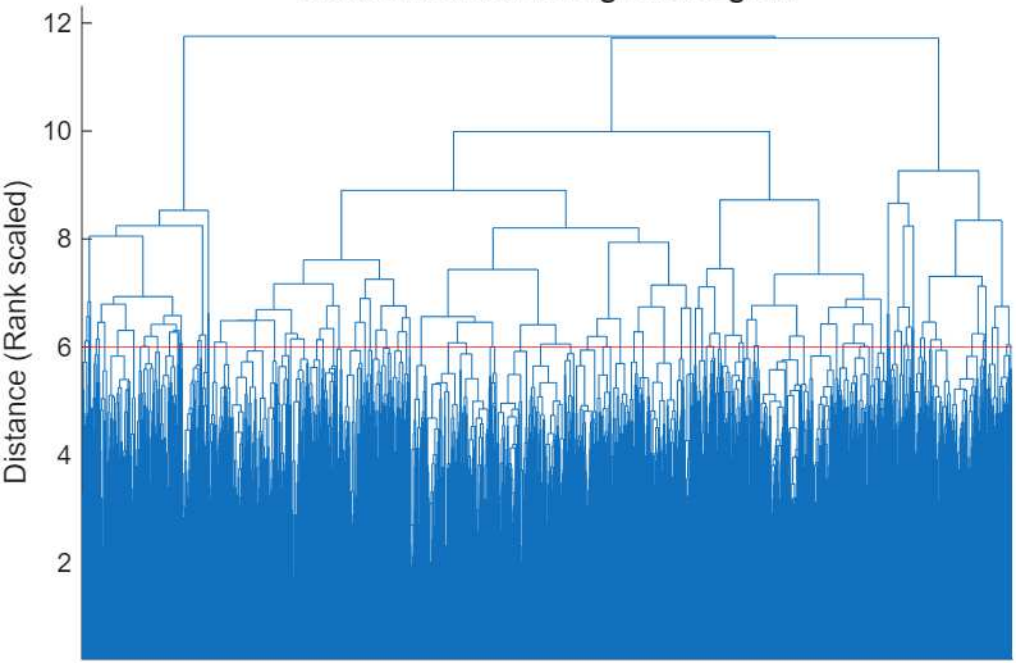


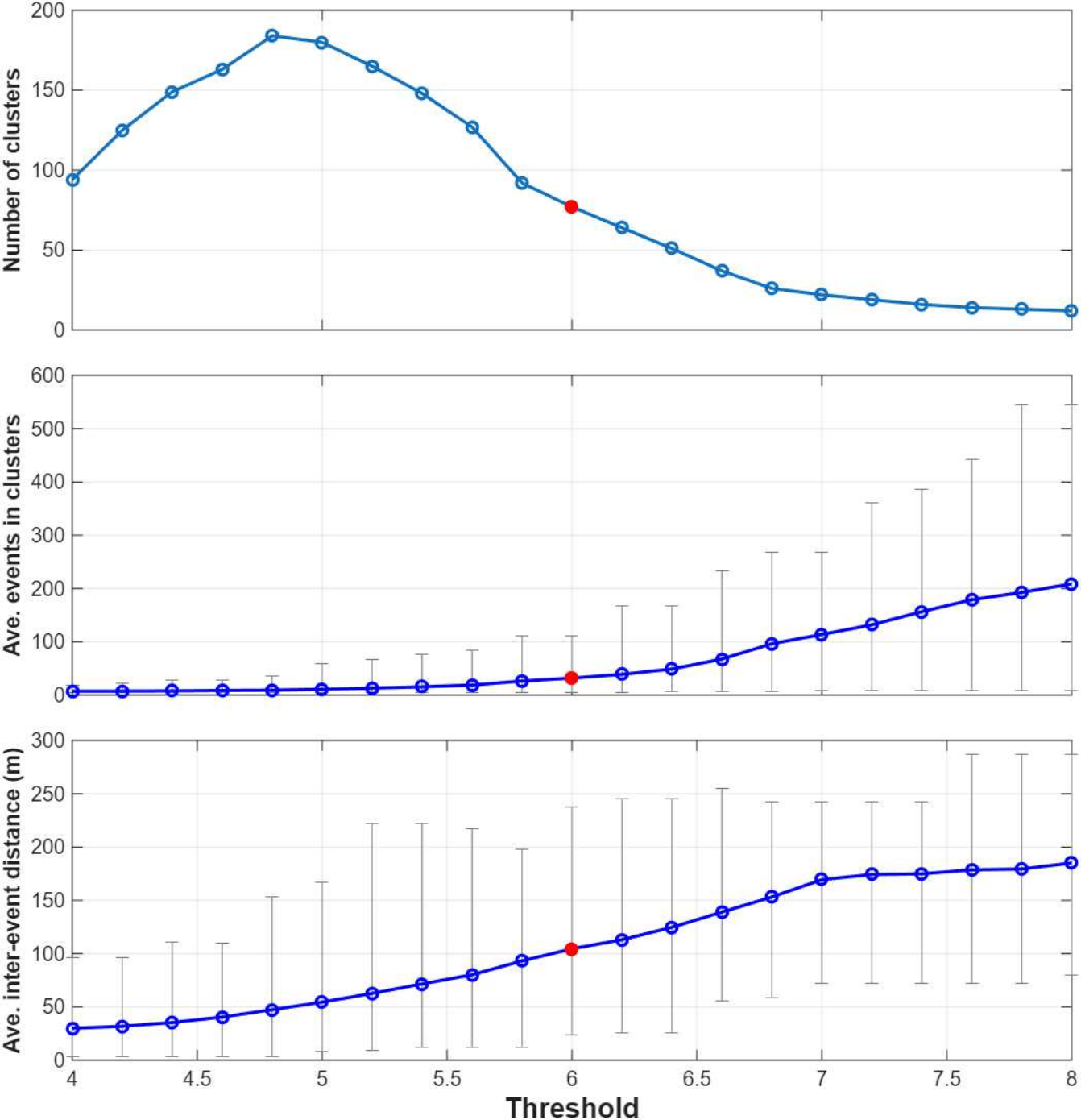


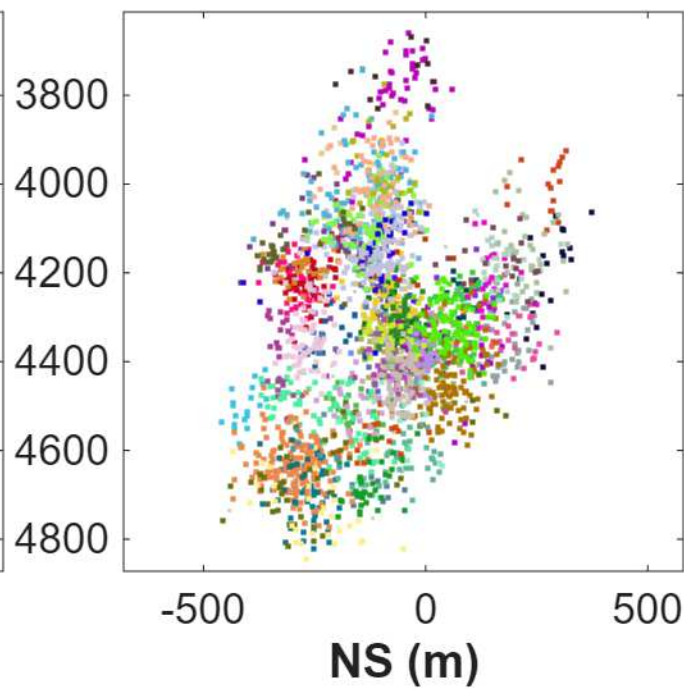
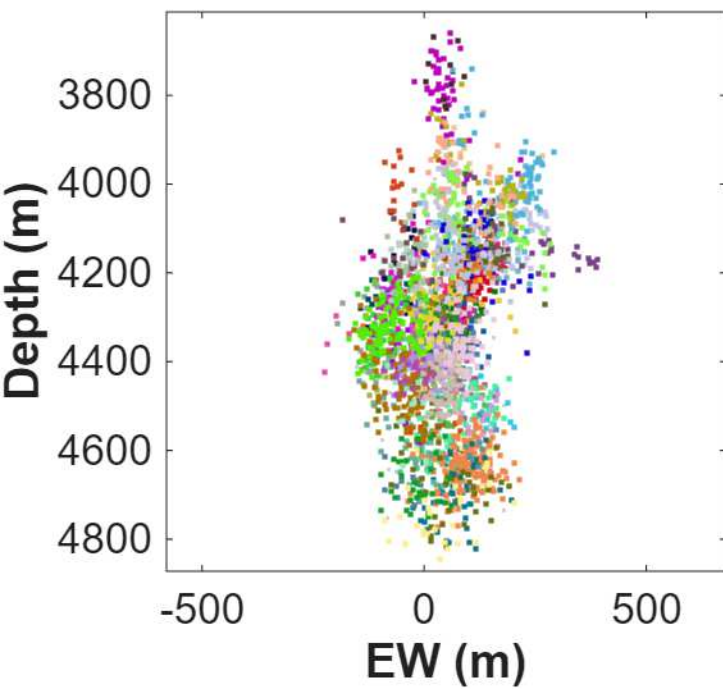
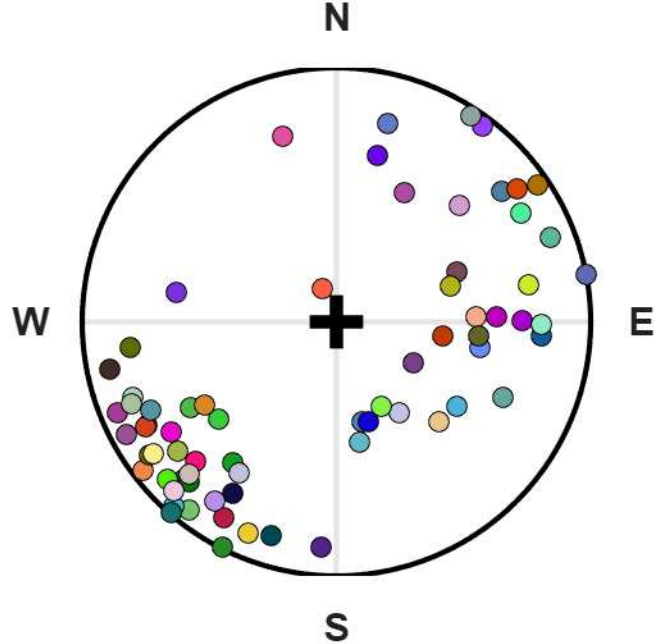
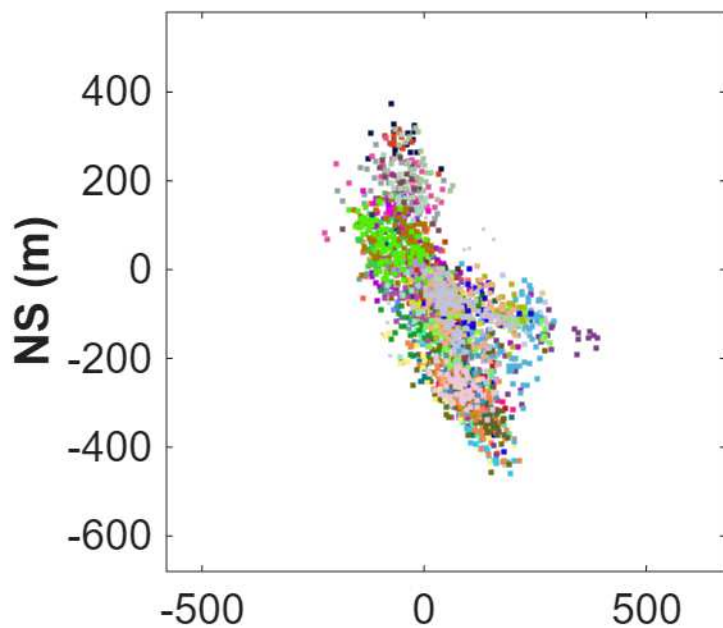




Hierarchical Clustering Dendrogram







Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additiononalfile1.1.docx](#)
- [GA.png](#)