

Fault reactivation analysis using microearthquake clustering based on signal-to-noise weighted waveform similarity

*Manuscript accepted by **Pure and Applied Geophysics***

*The final publication is available at **link.springer.com***

*via **<http://link.springer.com/article/10.1007/s00024-016-1281-4>***

Michael Grund^{1*}, Jörn C. Groos^{1,2}, Joachim R. R. Ritter¹

¹ Karlsruhe Institute of Technology, Geophysical Institute

Hertzstr. 16, 76187 Karlsruhe, Germany

² now at German Aerospace Center (DLR), Institute of Transportation Systems,
Lilienthalplatz 7, 38108 Braunschweig, Germany

* corresponding author: michael.grund@kit.edu, Tel: +49 (0)721 608 44491

Abbreviated title: Microearthquake clustering

Abstract

The cluster formation of about 2000 induced microearthquakes (mostly $M_L < 2$) is studied using a waveform similarity technique based on cross-correlation and a subsequent equivalence class approach. All events were detected within two separated but neighbouring seismic volumes close to the geothermal powerplants near Landau and Insheim in the Upper Rhine Graben, SW Germany between 2006 and 2013. Besides different sensors, sampling rates and individual data gaps, mainly low signal-to-noise ratios (SNR) of the recordings at most station sites provide a complication for the determination of a precise waveform similarity analysis of the microseismic events in this area. To include a large number of events for such an analysis, a newly developed weighting approach was implemented in the waveform similarity analysis which directly considers the individual SNRs across the whole seismic network. The application to both seismic volumes leads to event clusters with high waveform similarities within short (seconds to hours) and long (months to years) time periods covering two magnitude ranges. The estimated relative hypocenter locations are spatially concentrated for each single cluster and mirror the orientations of mapped faults as well as interpreted rupture planes determined from fault plane solutions. Depending on the waveform cross-correlation coefficient threshold, clusters can be resolved in space to as little as one dominant wavelength. The interpretation of these observations implies recurring fault reactivations by fluid injection with very similar faulting mechanisms during different time periods between 2006 and 2013.

Keywords: waveform similarity, cross-correlation, microearthquakes, induced seismicity, fractures

1. Introduction

Since 2006 induced microearthquakes have occurred close to the sites of geothermal powerplants near Landau ($-1 \leq M_L \leq 2.7$, depth 3.5-5 km) and Insheim ($-0.5 \leq M_L \leq 2.4$, depth 4.4-6.5 km) in Southern Palatinate in the Upper Rhine Graben in SW Germany (Fig. 1). There fluids are injected into a deep fault network related to the rifting processes (Schad, 1962; Meier and Eisbacher, 1991; Frietsch et al., 2015) at approximately 3.3-3.7 km depth. Between March 2006 and September 2013, 1982 microseismic events (Landau seismic volume: 1363, Insheim: 619) were detected using a cross-correlation algorithm (Plenkers et al., 2013; Groos et al., 2013a). To group potential events with almost identical hypocenters, often a cluster analysis based on waveform similarities is applied on such large datasets (e.g. Aster and Scott, 1993; Maurer and Deichmann, 1995). Due to high anthropogenic noise in the investigated area, a waveform similarity analysis in most cases is directly biased by low signal-to-noise ratios (SNR). Furthermore, data of different seismic networks and experiments were merged in the analyses to cover as long as possible observation time periods. Taking into account these difficulties, this study is focussed on three main questions: Is it possible to realize a precise waveform similarity analysis including as many possible recording stations and waveform records from a heterogeneous recording network? Can we detect clusters of high inter-event waveform similarity? If so, can we also resolve these clusters spatially within estimated relative hypocenter locations as well as find coinciding mapped faults and fault plane solutions of selected events?

High waveform similarities have been observed in several fields of seismological research with tectonic (Geller and Mueller, 1980; Ferretti et al., 2005; Gibbons et al.,

2007) volcanic (Frémont and Malone, 1987) and induced (Moriya et al., 2003; Asanuma et al., 2007; Wehling-Benatelli et al., 2013; Deichmann et al. 2014; Kraft and Deichmann, 2014) settings. In this context, waveform similarity mostly implies recurring fault plane reactivations or repeated slip on the same fault plane with very similar source mechanisms and travel paths over varying time periods (Geller and Mueller, 1980; Deichmann and Garcia-Fernandez, 1992; Got and Fréchet, 1993; Schaff et al., 2004). Two events that have nearly identical time series records at a single seismic station are called doublets and a series of events is called a multiplet (Poupinet et al., 1984) or cluster. The similarity of waveforms is often assessed by means of the maximum cross-correlation coefficient (hereinafter termed CC) for an event pair (e.g. Augliera et al. 1995; Schaff et al. 2004; Wehling-Benatelli et al. 2013). To achieve a reliable validation of a cluster, the event pairs should be characterized separately by waveform similarity in terms of high CCs on all available components (vertical, N-S, E-W) of each single seismic station or across a whole seismic network (Maurer and Deichmann, 1995; Igarashi et al., 2003). For each component of a recording station the corresponding CCs are placed in symmetrical component similarity matrices (e.g. Aster and Scott, 1993; Maurer and Deichmann, 1995). Mostly, not all events of a dataset are recorded at each station, which causes data gaps in the corresponding similarity matrices due to missing CCs. Creation of a network similarity matrix allows for the inclusion of all available data of all one- and three-component stations of a seismic network, whereby occurring data gaps mainly could be closed (Maurer and Deichmann, 1995). Due to a specific high noise influence in our waveform similarity analysis, we developed a signal-to-noise weighted network matrix which directly considers the individual site-specific SNRs of an event pair across the whole recording network. In the following sections, the creation of the SNR-weighted network matrix is presented as well as the application

to the two datasets of the closely neighbouring areas of Landau and Insheim (Fig. 1). Separately for both seismic volumes an equivalence class approach (e.g. Aster and Scott, 1993; Ferretti et al., 2005) is applied on the computed weighted network similarity matrices, which performs event clustering based on a predefined CC threshold (short: threshold). Iterative application of three different and increasing thresholds allows us to study the clustering evolution and additionally identify waveforms of very high similarity. For both areas, also a comparison with results based on the corresponding standard network matrix is done, which reveals the benefits of the developed weighting approach for our datasets. Finally, a cross-check with results from other studies on the microseismic events in the area provides obvious indications of repeated failure of pre-existing fault planes.

2. Area and seismological network

The study region (Fig. 1) is located in the central segment of the Upper Rhine Graben (URG), SW Germany. Rift formation in the URG began in Late Eocene to Early Oligocene and the main extensional tectonic activity ended in the Miocene (Meier and Eisbacher, 1991). The graben is composed of mainly downward faulted blocks and the faults are preferentially NNW-SSE oriented in the nearby study region (Schad, 1962; GeORG-Projektteam, 2013). Only minor to moderate seismicity has been observed in the central URG at the present time (Plenefisch and Bonjer, 1997; Ritter et al., 2009; Ritter, 2011; Barth et al., 2015). No historical events ($M_L > 3$) are known around Landau within a radius of 10 km (Leydecker, 2011) and a recent increase in the microseismicity rate is related to induced earthquakes due to deep fluid injection at geothermal installations (Cuenot et al., 2008; Plenkers et al., 2013; Groos et al., 2013a).

The earthquake activity at Landau started in 2006 with the first stimulation tests in a geothermal reservoir (Plenkers et al., 2013). Systematic seismic monitoring was initiated in 2008 and due to an increase in seismic activity, additional recording stations were added over time (Fig. 2). For some events, detected after the powerplant started operation on 21/11/2007 (see Fig. 2), hypocenters were located in a depth range between 3.5 km and 5 km (Zeiß, 2013). The determined magnitudes for the Landau events are in the range of $-1 \leq M_L \leq 2.7$.

In Insheim, just 4-5 km south of Landau, microseismicity was also induced during stimulation tests in the geothermal reservoir in 2009 and 2010 ($-0.5 \leq M_L \leq 2.4$, depth 5.1-5.6 km). Since the production start in autumn 2012 the activity continued ($-0.5 \leq M_L \leq 2$, depth 4.4-6.5 km) and it is closely monitored. The used seismic stations were operated by KIT - Karlsruhe BroadBand Array (KABBA, station code TMO) from 2006-2013 and the BESTEC GmbH (station code SOS) from 2008-2013. All used KABBA sensors are three-component instruments of type Streckeisen STS2, Geotech KS2000, Guralp CMG-40T, Lennartz LE-3Dlite or LE-3D/5s connected to 24-bit Earthdata loggers. The sensors installed at the SOS stations are all of type Guralp CMG-6T. The early recording period during the TIMO project (Ritter et al., 2008) allowed us to explore whether there was any correlation between the first events detected during a stimulation experiment in Landau 2006 (Plenkers et al., 2013) and later observed seismicity. For this purpose, station TMO20 is used (Figs. 1 and 2).

3. Waveform similarity and cluster analysis

3.1 Data selection and preprocessing

Since the microearthquake dataset contains events of two different and independent seismic volumes, already within the detection process all events were assigned either

to the Landau or Insheim seismic volume (Groos et al., 2012; Groos et al., 2013a). Due to the difficult (low SNR) absolute localization of the very weak seismic events ($M_L < 0.5$), a comparison of waveform similarities with the absolute located events of both seismic volumes was used. Detected events were compared with all absolutely located events by cross-correlating time windows, containing P - as well as S -wave onsets, on all three components (Z, N-S, E-W). The located event that delivered the highest cumulated CC (sum of the Z, N-S, and E-W component CCs) was used to determine the seismic source volume of the detected event. The automated source volume determination based on the waveform similarities was checked by a manual analysis of seismogram sections (seismograms over distance plots) of all available station recordings of an event for the assumed source volume. Hereinafter only for one processing step (see section 3.2, minimum travel times) a distinction between the two seismic volumes must be made because most processing and investigation steps are equal. Results based on the following approaches for both seismic volumes are presented and discussed separately in section 4.

Sequential events i and $i + 1$ with origin time differences less than 1 s ($t_{0,i+1} - t_{0,i} < 1$ s) were excluded from the further analysis in order to avoid complications due to overlapping waveforms. Following this constraint, in total 19 events were removed, all assigned to the Landau seismic volume. Componentwise all waveforms were preprocessed by removing the mean, a possible linear trend and the instrument response function to reconstitute the true ground motion velocity. Restitution is necessary because different sensors were operated during the recording periods and comparable recordings are required. Frequency spectra were calculated and analysed to estimate the frequency range of interest. The highest amplitudes were found in the frequency range 2-20 Hz. Hence, a second order Butterworth bandpass filter was applied on the data forward and reverse with corner frequencies at 2 Hz

and 20 Hz. As not all instruments were recording during the same time periods (Fig. 2), data gaps exist at some recording sites. For the following data analysis, the number/dimension of events at all stations must be equal. This requirement was realized by replacing missing events at a recording station by a zero trace on all three components.

3.2 Cross-correlation and time window selection

The detection of possible event clusters often is handled by applying a waveform similarity analysis based on the normalized cross-correlation function (e.g. Augliera et al. 1995; Schaff et al. 2004; Wehling-Benatelli et al. 2013):

$$C_{xy}(\tau) = \frac{\sum_{i=1}^n x(t_i) y(t_i+\tau)}{\sqrt{\sum_{i=1}^n x(t_i)^2} \cdot \sqrt{\sum_{i=1}^n y(t_i)^2}} , \quad (1)$$

where $x(t_i)$ and $y(t_i)$ denote two discrete recorded event time series and τ is the time lag. For each possible event pair, the maximum CC of the corresponding normalized cross-correlation function C_{xy} is used as a measure for waveform similarity separately for all three components of a recording station. Comparing two events with identical waveforms ($x(t_i) = y(t_i)$, $\tau = 0$) results in a CC of one (autocorrelation). Due to the large number of around 2000 microseismic events, an automatic procedure was used to cut the time windows for the cross-correlation. For this purpose, we defined station-specific minimum P -wave travel times derived from absolutely located events (Zeiß, 2013) to mark automatically the beginning of the waveforms used for cross-correlation relative to the estimated origin times (Fig. 3). Due to different hypocenter-station distances the estimated minimum travel times of the P -waves vary between 0.9 s and 2.25 s for the Landau area and 0.85 s and 2.3 s for the Insheim area. The set time windows for all events of a single recording station contain the first four

seconds after the corresponding time mark including the onsets of the *P*- and *S*-waves plus the *P*-wave coda and the most significant parts of the *S*-wave coda (see e.g. Baisch et al. 2008). Therefore, *S* - *P* differential travel times are directly considered in the similarity analysis. Using time windows without the period between the origin time and the *P*-wave onset minimizes possible decreasing CCs caused by the varying noise influences on otherwise equal or similar propagation paths (see e.g. Arrowsmith and Eisner, 2006). Although all events are located within two small seismic volumes, each of around 50 km³, the hypocenters differ in their absolute location due to the complicated 3-D tectonic structure (Figs. 1b and c). Therefore some of the automatically estimated time windows contain a small part of pre-event noise before the corresponding *P*-wave onsets. Due to significant time savings by setting the time marks automatically instead by hand, this small noise part is accepted.

For *N* events of a dataset, the maximum CC values of all cross-correlated event pairs were sorted chronologically in symmetric matrices of dimension *N* × *N* separately for each component of a recording station (component similarity matrices). Due to partly zero traces, resulting NaN values (Not a Number) may occur which imply no data availability for an event and finally no observation of the corresponding event pair at this recording station and components. The zero traces guarantee a consistent dimension of the different component similarity matrices of all stations.

At some recording stations, difficulties due to varying sampling rates were handled individually by applying a resampling routine before cross-correlating the waveforms of an event pair. If necessary, the lower sampling rate of the analyzed event pair was used to resample the more frequently sampled time series. The sampling rates vary for KABBA stations (50-200 Hz) as well as SOS stations (100-400 Hz). A comparison

of the amplitudes of the frequency spectra before and after the resampling revealed no significant change in the frequency content for the relevant signal frequency band of 2-20 Hz. Hence, no influence on the CC due to the resampling process is expected.

3.3 Network similarity matrix and noise influences

Due to time-dependent heterogeneous station configurations and data availabilities during the entire investigation period (Fig. 2), we applied a network matrix to combine the cross-correlation results of all recording stations into one matrix. In the most simplified case, the mean of all non-zero or rather non-NaN elements of the different component similarity matrices (uniform dimensions) is calculated (Maurer and Deichmann, 1995; Wehling-Benatelli et al., 2013). However, first tests indicated a strong impact on our recordings due to anthropogenic noise such as traffic, industry and infrastructure, unfavorable site conditions with unconsolidated rift sediments or weather conditions (Groos and Ritter, 2010). The influence of anthropogenic noise was also documented by Plenkers et al. (2013) for the detection of microseismic events in the study region. Unfortunately, a large number of transient noise signals are characterized by a similar frequency content and similar amplitude values as the detected microearthquakes. To obtain an overview on the influence of the noise level, the SNR of 416 selected Landau events (recorded between March and August 2012) was computed for a shallow borehole station (TMO64, depth about 70 m, see Fig. 1) and for several surface stations (TMO53, SOS2, SOS3) using the following definition:

$$\text{SNR} = \frac{\max(|x_{sw}(t_i)|)}{\sqrt{\frac{1}{n} \sum_{j=1}^n x_{nw}(t_j)^2}}, \quad (2)$$

where $x_{sw}(t_j)$ and $x_{nw}(t_j)$ indicate the corresponding time series of the signal and noise time windows. The required time windows were determined using also the minimum *P*-wave travel times of the corresponding recording station (Fig. 3). All noise windows include the time series part 0.75 s before the set time mark (see section 3.2). Signal windows contain the same part as the windows used for cross-correlation (first 4 s after the time mark). The denominator on the right-hand side of equation (2) corresponds to the definition of the root-mean-square (rms) of the time series $x_{nw}(t_j)$.

After cross-correlating the 416 events componentwise for each station, in general we observe higher CCs on all three components at the borehole station TMO64 as well as patches of concentrated high CCs compared to the outcome of the exemplary surface station TMO53 (Fig. 4). This variation occurs though both stations have approximately the same epicentral distances to the center of gravity of all located Landau events (Fig. 1). However, as shown by Schaff et al. (2004), even for nearly identical event-receiver distances the CCs of known clustered events can vary intensely within a seismological station network. Considering the corresponding SNR distributions of the 416 events (histograms in Fig. 4), as expected, a strong variation for the SNRs becomes obvious. For the vertical component of surface station TMO53, approximately 90 % (red step functions in histograms) of the 416 events have SNRs smaller than 15 while at station TMO64 (borehole station) only 60 % of the same events reveal SNRs below 15. Further, at TMO64 more than 10 % of the recordings have SNRs higher than 50, compared to 1-2 % at TMO53. The other *P* components (N-S and E-W) show similar trends (Fig. 4). Since the borehole station TMO64 was installed in March 2012 (Fig. 2), using only low-noise borehole data was no option to cover as long as possible time periods of observation. Hence the application of a standard network matrix for our waveform similarity analysis is

significantly influenced by the strongly varying noise levels across the recording network.

3.4 Signal-to-noise weighted network similarity matrix

To overcome problems with site-specific noise, we developed a modified network similarity matrix which takes into account the varying SNRs of each event pair across the whole network. By means of Fig. 5, in the following section the work flow to create a signal-to-noise weighted network similarity matrix for 15 exemplary seismic events is outlined:

- **Step 1:** Applying equation (2), each event pair delivers two SNRs per component (Z, N-S, E-W). Componentwise the lower of the two SNRs is stored in a SNR-matrix of the same dimension $N \times N$ as the corresponding similarity component matrices (Fig. 5). Matrix entries with NaN values (gray color in Fig. 5) indicate event pairs not observed at the corresponding component or station.
- **Step 2:** Weighting the single CCs of all event pairs in terms of the corresponding SNR is handled by applying a sigmoid function (e.g. Priddy and Keller, 2005; von Seggern, 2006) that is defined as:

$$sig(SNR) = \frac{1}{1 + e^{-\frac{SNR-a}{b}}}, \quad (3)$$

where the parameters a and b determine the characteristics of the weighting function. Depending on the parameter selection, the sigmoid function assigns a weighting factor $wf = sig(SNR)$ between 0 and 1 for each SNR. Parameter a shifts the linearly weighted range of the function along the abscissa (SNR) and b controls the portion of the curve that assigns the weighting factors to the

corresponding SNRs linearly (e.g. Horstmann et al. 2013). The parameter configuration for our dataset ($a = 7$ and $b = 0.8$, see Fig. 5) was determined by studying the effect of different noise levels on the CCs. For this purpose, quasi-autocorrelations were performed for recorded reference waveforms (each with a fixed SNR) with several versions of themselves (varying SNRs). The SNRs were adjusted by adding randomly generated synthetic noise of different amplitude orders on the recorded waveform. As a result, the correlation of a reference waveform (SNR=15) with waveforms whose SNRs are approximately equal to or greater than 15 offers a small variation of the corresponding CCs ranging from 0.9 to 0.95. On the other hand, for a more noisy reference waveform (SNR=5) the correlation with waveforms of similar noise level, delivers CCs varying between 0.5 and 0.6 which only in unusual cases indicates a plausible waveform similarity. Since for each event pair the lower SNR of both is relevant for weighting (step 1), we decided to down weight the CCs with SNRs < 5 towards 0. Furthermore, the tests showed that for waveforms with SNR > 15 the noise has little influence on the CC in our case. Hence CCs with corresponding SNRs above 15 (Fig. 5) are weighted with a factor of 1. For the SNR range between approximately 5 and 15 the corresponding CCs are linearly weighted by the function displayed in Fig. 5. For other datasets the parameters a and b will differ and need to be adjusted depending on the local noise conditions.

- **Step 3a:** For each non-NaN entry of the SNR-matrices, the weighting factors $[0 \ 1]$ determined by $wf = sig(SNR)$, componentwise are stored in a weighting matrix. **b:** Summation of all non-NaN elements of the weighting matrices results in a matrix of normalization factors which is used at a later stage.

- **Step 4:** Weighting matrices are multiplied elementwise with the corresponding component similarity matrices including the CCs [0 1] of each event pair. This step leads to the weighted component similarity matrices which contain the signal-to-noise weighted CCs [0 1] elementwise for each component of a station.
- **Step 5:** Separately each of the weighted component similarity matrices (non-NaN entries) is elementwise divided by the normalization factor matrix of step 3b which results in normalized weighted component similarity matrices.
- **Step 6:** Across all used stations, these matrices contain individual signal-to-noise weighted and normalized CCs [0 1] for each possible event pair on all three components of a single station.
- **Step 7:** Summation of the normalized weighted component similarity matrices delivers a signal-to-noise weighted network similarity matrix. Applying the normalization of step 5 guarantees that finally all diagonal elements of the network matrix are equal to one (Fig. 5). Matrix elements with NaN values indicate event pairs not observed across the entire seismic station network.

3.5 Clustering algorithm

To group events with similar characteristics into clusters, several techniques were developed including approaches using waveform similarities based on cross-correlation methods (e.g. Maurer and Deichmann, 1995) or moment tensor clustering (e.g. Cesca et al., 2014). For our analysis, we use an often applied bridging technique (e.g. Aster and Scott, 1993; Cattaneo et al., 1997; Igarashi et al. 2003; Ferretti et al., 2005; Massa et al., 2006) which is based on an equivalence class algorithm (Press et al., 1988). If two event couples (A, B) and (B, C) separately

exceed a defined CC threshold, indicating waveform similarity, then events A, B and C are sorted to the same event cluster. Thus, this algorithm connects events A and C, although their corresponding CC may not exceed the given threshold. In this case, event B represents the bridge-event between A and C (Cattaneo et al., 1997). Using the bridging technique combined with a network matrix instead of several component matrices allows us to search for clusters coherently in one single matrix which contains information about all available recordings of a heterogeneous seismic network in terms of CCs for all possible event pairs.

3.6 Selection of thresholds to cluster events

The selection of a suitable CC threshold for clustering is not trivial and there is no universal way to set a perfect value (Ferretti et al., 2005). Applying the bridging technique, a too low threshold often leads to a large group of events without strong similarity. On the other hand a too high value, very close to 1, leads to very small groups with events of very high similarity caused by favorable SNRs (Aster and Scott, 1993; Cattaneo et al., 1999). Thus, noisy records of events with potentially high waveform similarities mostly are not considered by high thresholds due to their decreasing CCs. To investigate both, the development of cluster allocation in time and space as well as the occurrence of nearly identical waveforms, a stepwise clustering with three predefined thresholds is studied separately for both seismic volumes. Firstly, a threshold of 0.7 is applied on the events of each dataset. Combined with available relative locations, we expect to resolve the spatial distribution of different large scale structures, each with individual rupturing characteristics. In a second step, subsequently the events of each resulting 0.7-cluster were separately clustered with a threshold of 0.8. For each 0.8-cluster the final clustering with a threshold of 0.9 follows the same procedure. Applying the

thresholds of 0.8 and 0.9 may help to detect recurring rupturing of finer small scale structures within the larger ones. Together with the temporal evolution of the clusters this allows one to speculate on potential migration paths of the injected fluids and thus on (re)activated faults in the investigation area which are disclosed by the directions of the detected event clusters (see below).

3.7 Cluster naming

The cluster names used in the following section consist of a maximum of five characters which directly indicate the cluster assignment of an event depending on the clustering steps of the three sequentially used thresholds. The first capital letter indicates the affiliation to one of the two seismic volumes, Landau (L) or Insheim (I). The second capital letter denotes a cluster assignment with a threshold of 0.7 (Landau: A-P; Insheim: A-N). Numbers of 01-15 (Landau) and 01-28 (Insheim) at positions three and four imply that the corresponding event was clustered with a threshold of 0.8. Finally, a possible cluster classification with a threshold of 0.9 is indicated by a small letter in the fifth position (Landau: a-f; Insheim: a-b). For example the cluster name LB04a implies that the corresponding Landau event (L) was clustered with thresholds of 0.7 (B), 0.8 (04) and 0.9 (a).

4. Application to the seismic volumes of Landau and Insheim

4.1 Evaluating the performance of SNR-weighting

In order to evaluate the performance of our approach, separately for both seismic volumes, we compare the results based on the SNR-weighted network matrix with those of the standard mean network matrix.

4.1.1 Landau

Including all available preprocessed waveforms for the Landau area, a mean and a weighted network matrix was computed (Figs. 6a and b) with events during the period between March 2006 and September 2013 (Fig. 6c). In general, the weighted network matrix, compared to the mean network matrix, reveals similarity patterns with higher CCs which are displayed as reddish encoded patches in Figs. 6a and b. High CCs close to matrix diagonal elements imply events with very similar waveforms which occurred within narrow time windows, while off-diagonal high values indicate events not tightly clustered in time. This is due to the chronological ordering/numbering of the events which increases with origin time (see Fig. 6c). Detailed visual inspection implies that most events that occurred during the injection experiment in 2006 (event no. 1-132) have high waveform similarity to each other. This inter-event similarity was also observed by Plenkers et al. (2013) during event detection. Further, a high correlation between most of the stimulation events in 2006 and a part of swarm-like events in June 2012 (event no. 708-1034) is likely regarding the matrix entries around event pairs labelled (I) in Figs. 6a and b. Compared to the corresponding lower CCs of the mean network matrix (Fig. 6a) the weighting procedure leads to maximum values of around 0.7 (Fig. 6b). These mainly low magnitude swarm events ($M_L < 0$, see Fig. 6d) all occurred within a short time window of around five hours after a shutdown of the geothermal powerplant (Fig. 6c). The corresponding CCs also indicate a high inter-event waveform similarity. Improved quantification of this similarity is not possible, since the corresponding event pairs were observable only on three components of one recording station (TMO20) which was operated 2006 and reinstalled at the same site in 2010 (Fig. 2). Nevertheless, the SNR-weighting provides a slightly higher correlation between the events in 2006 and 2012 by reducing the noise influence on the 2012 low magnitude swarm events. Further events, possibly correlating to the 2006 events or 2012 events, occurred

during the years 2007 (event no. 133-192) and 2008 (event no. 193-260). This relation becomes obvious only by regarding the weighted matrix elements around (II) in Fig. 6b with CCs of approximately 0.75. Thus, the relation to the 2006 events can be demonstrated only indirectly (compare Fig. 5, step 7) because the operation of the SOS stations did not start before 2007. SNR-weighted matrix entries around (I) and (II) in Fig. 6b indirectly imply a linkage between the corresponding events that occurred during 2006, 2007/08 and the 2012 swarm events though no direct observation for event pairs consisting of earthquakes of 2006 and 2007/08 exists (gray shaded areas). Besides the mentioned relationships a few other single events between 2009 and 2013 are characterized with a similarity to those earthquakes (e.g. event no. 1116-1120). This recurring similarity indicates a recurrence of microearthquakes which had already occurred in 2006 on the same fault. Lower CCs imply partially different source mechanisms or a different location in the years 2007-2013 than for the events in 2006. Besides several larger patches with high CCs also a few doublets (e.g. event no. 671-672) and sparse-event clusters (e.g. event no. 549-568) are visible which mostly indicate waveform similarity between each other during short time periods.

Applying our bridging algorithm on the weighted network matrix, in total around 60 % of the entering 1344 earthquakes are clustered with a threshold of 0.7 (Fig. 7). For the mean network matrix just 30 % of the events got a cluster assignment with threshold 0.7. For increasing thresholds of 0.8 and 0.9, the number of clustered events decreases whereby the percentual amount of events for the mean network matrix is consistently lower. Besides the total number of clustered events, Fig. 7 displays the distribution of clustered events in four different magnitude ranges (N/A means that no magnitude was estimated). For a threshold of 0.7, no significant

differences are found for the mean and weighted network matrices in the magnitude ranges $0.5 \leq M_L < 1$ and $M_L \geq 1$. For $M_L < 0.5$ a strong discrepancy is revealed, because the clustering based on the weighted approach is superior to detect similar events with a low SNR. Using the weighted network matrix, for thresholds 0.7 and 0.8 around three times more weak events are clustered in that magnitude range than for the mean matrix. For a threshold of 0.9, the frequency distributions of both network matrices are dominated by events with larger magnitudes ($M_L \geq 0.5$) which result in recordings with a high SNR in our case.

4.1.2 Insheim

For the Insheim area, a mean and a weighted network matrix (Figs. 8a and b) was computed for events between September 2008 and September 2013 (Fig. 8c). The weighted network matrix (Fig. 8b) is characterized by distinct similarity patterns with near- and off-diagonal CCs of 0.8-0.95 which occur especially for event pairs in the years 2012 and 2013 (event no. 389-619). Only a few of these show a likely correlation in the corresponding areas of the mean network matrix (Fig. 8a). Besides this, by visual inspection the weighted matrix (Fig. 8b) reveals two distinctly separated and widespread patches of event pairs with CCs mostly between 0.6 and 0.7 (matrix entries around (A) and (B) in Fig. 9b). Those events, which occurred during several injection and circulation tests between September 2008 and July 2009 (event no. 1-83), are characterized by a high inter-event similarity regarding the matrix entries around event pairs labelled (A) in Fig. 8b. These events also have a high waveform similarity to a few events later in time (between event no. 120 to 260 in Fig. 8b) but they are not corresponding to event pairs of the other large patch (B) between 2009 and 2010. This distinction is not visible in the mean matrix (Fig. 8a). The identification is mainly caused by the continuous high CCs of the E-W-

component matrix of station TMO54 (Fig. 9). The recording station TMO54 is located relatively close to the seismic volume of Insheim (see Fig. 1) and thus in general high SNR-values for most of the event pairs are observed, which lead to a high impact during the weighting process although most of the corresponding event magnitudes are quite low (Fig. 8d). A detailed view of ten exemplary earthquake recordings from station TMO54 is presented in Fig. 10 which displays the waveforms and the CCs of the corresponding event pairs. Although the *S-P* arrival time differences for all displayed events are almost equal, the waveforms of the vertical and N-S components partly differ clearly compared to the very high similarities of the E-W component waveforms. Furthermore, the amplitude peak values of the E-W component waveforms are all significantly larger than the vertical and N-S components which directly influences the SNR and thus the weighting procedure. A possible reason for higher waveform similarity observed only at one component of a recording station compared to the other components are spatially tightly located hypocenters but slightly different focal mechanisms and hence varying radiation patterns (Deichmann and Garcia-Fernandez, 1992; Baisch et al., 2008). Furthermore, we assume that the amplitude differences between the recording components most likely are caused by complicated underground structures, radiation patterns as well as presumed amplification effects in this area due to a distinct variation of the sediment thickness (Figs. 1b and c; Groos et al., 2013b).

Evaluating the results of event clustering for the 619 earthquakes of the Insheim area (Fig. 11) compared to the clustering of the Landau dataset (Fig. 7), a similar pattern of the magnitude frequency distribution for small events with $M_L < 0.5$ is found. Thus, for both datasets the application of SNR-weighting is more sensitive to perform clustering in a magnitude range up to $M_L < 0.5$ which in our case helps to associate

much more events to clusters than performing clustering with the mean network matrix. Thus, the following sections only deal with results based on the weighted approach.

4.2 Detailed cluster formation based on SNR-weighted waveform similarity

4.2.1 Landau

Detailed clustering results for the Landau area are illustrated in Fig. 12. Fixing the threshold to 0.7, we find two large clusters containing 404 events (called LA, blue) and 284 (LB, red) events (Fig 12a). Further smaller clusters contain another 48 events (LD, green) and 10 events (LH, gray). The remaining clusters contain six or fewer events whereby seven clusters are composed of only two events defined as doublets. Colors indicate clusters which are discussed below in detail in terms of relative locations and fault plane solutions. Note, not all events of a colored cluster are located, thus the epicenters displayed only represent a part of the overall events of a cluster. Black color in Fig. 12 means that no events of the corresponding cluster are plotted in the spatial distributions due to missing location results.

Using a higher threshold of 0.8 (Fig. 12 b), most clusters, especially LA and LB, split up into smaller groups (hereinafter called subclusters). Considering the temporal and magnitude distributions for the eight largest subclusters (Fig. 12d and e) reveals that events associated to LA01 and LA02 only occurred in March and April 2006 during the stimulation test (Plenkers et al., 2013) with magnitudes ranging from $0 \leq M_L \leq 1.1$. With the exception of one event in May 2013, seismic activity of LA04 (23 events) is limited to a short time period of two days (23 and 24/11/2007) after the initial operation of the powerplant started on 21/11/2007 (Figs. 12d and e). Up to subcluster

LA07, no locations are available for corresponding events due to the poor station coverage (Zeiß, 2013). Events assigned to subcluster LA08 ($-0.4 \leq M_L \leq 1.7$) occurred since 2008 during the whole study period until 2013, except for 2011 when no activity belonging to this group was recorded (Fig. 12d). Furthermore, a temporal variation of seismic activity within cluster LA08 is observable with the highest rate of events in 2009 and 2010. For subcluster LA09, after one single event in October 2010, three events followed within a short time period in May 2011. However, the major part of in total 83 events ($-0.5 \leq M_L \leq 0.7$) in subcluster LA09 is associated to the above outlined earthquake swarm (Fig. 6c) with an activity of around five hours on 25/06/2012 (Figs. 12e and e). A further couple of events belonging to LA09 recurred in December 2012. The three events assigned to subcluster LA10 (Fig. 12b) also were recorded on 25/06/2012 and in December 2012, respectively. For cluster LB a threshold of 0.8 leads to a split into 15 subclusters (Fig. 12b). Overall most of them consist of events that occurred at irregular time intervals between 2007 and 2013 (see LB01,04 and 07 in Figs. 12d and e). Besides several subclusters resulting from the two large clusters LA and LB, additionally for clusters LC, LD, LF, LH and LI subclustering is observable using a threshold of 0.8 (Fig. 12b).

A final cluster classification with a threshold of 0.9 leads to further subclustering of the differently estimated subclusters whereby a few of them consist of more than 2-4 events (Fig. 12c). The success of applying the bridging technique on our weighted network matrix is evaluated in Fig. 13 for six events of cluster LB03. Consistently all events have nearly identical waveforms on all components of three different recording stations.

4.2.2 Insheim

The cluster formation of the Insheim region using a threshold of 0.7 results in 14 clusters (Fig. 14a). Besides several smaller clusters with a maximum of up to 23 events (IF, IG, IJ and IK), the large group IA contains most of the in total 492 clustered events. A rise of the threshold to 0.8 splits up cluster IA into 28 subclusters with IA05 containing a maximum of 82 events (Fig. 14b). Some of these events occurred between 2008 and 2010 ($-0.1 \leq M_L \leq 2.2$) close to time periods when several stimulation and circulation tests were accomplished (Fig. 14d). After an aseismic period, when no fluids were injected, until the initial operation of the powerplant in November 2012, several further events occurred of which a part is assigned to subcluster IA05 (Fig. 14d and e). Out of subclusters IA01-IA18, besides IA05 only subcluster IA02 (3 events) shows a connection to events after the aseismic phase. All other subclusters contain events which only occurred between 2008 and 2010, especially subcluster IA03 with in total 30 events of magnitudes $0.2 \leq M_L \leq 1.7$ (Fig. 14e and e). In contrast, solely subclusters IA19-IA28 include events between 2012 and 2013 which are demonstrated in Figs. 14d and 14e for the two subclusters IA19 and IA26. From the other subclusters only IF01 also contains events before and after the aseismic period. Subclusters IB01, IC01, ID01, IE01 and IF01 only include events between 2009 and 2010 while several events occurred in 2012 and 2013 are associated to subclusters IG01 and IJ01 (Fig. 14d and e) as well as IG02, IK0, IL01 and IM01. Final clustering with a threshold of 0.9 (Fig. 14c) leads to several individual doublets and a few subclusters with a maximum up to eight events (IA19a).

4.3 Comparison with relative event locations and fault plane solutions

The outcome of the cluster analysis depends on the properties of the seismic waveforms which are influenced by the source position, the rupture mechanism and the elastic medium properties along the propagation path. The latter are assumed to

be constant and thus we expect that the hypocenter locations and fault plane solutions (fps) should be very similar for individual events belonging to a common (sub-) cluster. Therefore, we cross-check the clustering results with known relative locations (Zeiß, 2013) and fps (Gaßner et al., 2014) for the Landau and Insheim areas. The relative hypocenters were determined with a double-difference method (Waldhauser and Ellsworth, 2000) and the code hypoDD (Waldhauser, 2001). The uncertainties of the relative locations are about ± 50 m, absolute location errors are up to 500-1000 m mainly due to the uncertainties from the local complex 3-D seismic velocity structure (Zeiß, 2013; Staudt et al., 2016). The fps were calculated from polarities of *P*-, *SV*- and *SH*-waves with the code FOCMEC (Snoke, 2003) with average uncertainties of mainly less than 10° but partly reaching up to 25° for the fault plane orientations (Gaßner et al., 2014). Full waveform modelling was executed with a simplified 3-D velocity model to check the fps with synthetic seismograms (Gaßner et al., 2014). In the following, we use the results from the weighted similarity matrices and thresholds of 0.7 and 0.8. It is noted that not all events located by Zeiß (2013) are assigned to a cluster defined in this study. Furthermore, not all clustered events are located with hypoDD, mainly because no reliable onset time pickings could be determined due to the low SNR. Reliable hypocenter parameters are available from 2009 onwards when the station coverage was sufficient. Unfortunately for earlier events (2006-2008) no reliable hypocenter information exists due to the unbalanced station configuration and an inadequate amount of recordings.

4.3.1 Landau

Two dominant structures with different strike orientations dominate for events with a threshold of 0.7 for clustering in the Landau area (Fig. 15, left column). The cluster LA is aligned in NE-SW direction (ca. 40° from north to east, blue). The cluster LB

with NNW-SSE strike direction (ca. 150° from north, red) includes most events and overlaps with clusters LD and LL. Clusters LE and LI are located at the edge of cluster LB and can be separated with a CC of 0.7 from LB. The cluster LH is located directly below the injection hole and has a N-S orientation. The epicenter distribution in Fig. 15a reveals that the events with a similar waveform, and hence the ones located within a common cluster, are neighbouring events. The vertical cross-sections in Fig. 15c (N-S) and Fig. 15e (W-E) reveal additional information and the 3-D complexity. The cluster LH is located about 1.5 km underneath the injection depth (~ 3 km) at about 4.5-5 km depth (absolute uncertainty about ± 0.5 km). Due to its distance to the main seismic cloud (clusters LA and LB), the clustering algorithm can identify and separate this event group very clearly. The seismic cloud stretches upwards from cluster LH in northern direction and the shallowest events are located in the northern part of clusters LB and LI. Although, event locations from cluster LD are situated within cluster LB, their waveforms are slightly different which leads to a different cluster assignment. Possibly this difference is due to a different rupture mechanism, unfortunately there is no fps available for events of cluster LD due to unclear polarities of first arrival phases.

An increase of the threshold from 0.7 to 0.8 leads to a finer resolution of the clusters (Fig. 15, right column). Cluster LA is subdivided into three subclusters LA08, LA09 and LA10 (blue symbols in Fig. 15) at different depth levels. The 1-1.5 km difference in hypocentral depth between LA08 events and LA09 and LA10 events is enough to separate these subclusters by a clustering with a threshold of 0.8 (Fig. 15d). This may be due to the increase in travel time between the *P*-wave and *S*-wave by about 0.15 s. At the injection area the deepest sited cluster LH is subdivided into subclusters LH01 and LH02 (gray in Figs. 15d and f) which are about 300 m apart.

The large cluster LB (red) is subdivided into 10 subclusters (LB01-09 and LB12) which are spatially grouped. Their typical dimension is about 200-400 m in horizontal as well as in vertical dimensions (about 4-6 times larger than the uncertainty of the relative location between the events). The dominant wavelength of the input seismic waveforms with 8-10 Hz is about 350-430 m inside the seismic cloud with a background *S*-wave velocity of 3.5 km/s. Thus, the clustering with a CC of 0.8 seems to be able to differentiate between clusters that are only one wavelength apart. This could be due to the 3-D tectonic structure at depth which can influence the waveforms. The rupture mechanisms inside cluster LB do not vary strongly (Fig. 15b), because all determined fps have a dominant NNW-SSE oriented normal faulting mechanism. Thus, the division into the subclusters LB is not influenced by the rupture process. Since the events of subclusters LB01-09 and LB12 form a uniform association for a threshold of 0.7, the dataset contains several bridge-events with waveform characteristics of different single events leading to a linkage between the different subclusters.

The events ($M_L < 1$) which occurred during the stimulation with high injection pressure (up to 16-18 MPa) in spring 2006 (see Figs. 12d and e) could not be located, because there were not enough recordings available (Plenkens et al., 2013). However, due to the high correlation of the 2006 waveforms (LA01) with waveforms of the swarm events that occurred on 25/06/2012 and which are sorted to LA09 (Fig. 16), we speculate that the events in 2006 belong to the subclusters LA08-LA10. This means that the stimulation did not activate microevents close to the injection hole but some 600-700 m to the west.

Visual inspection implies that the N-S component waveforms of the 2006 swarm and the 2012 swarm have high waveform similarity, while the other two components only

show weaker similarity (details below). Using the CC of selected events as proxy leads to the component matrix results presented in Fig. 16. On all three components the inter-event-similarity of the 2006 swarm (event no. 1-5 and corresponding matrix entries around A) and 2012 swarm (event no. 6-10 and B, except the E-W results) is clearly visible with CCs often higher than 0.9. While the similarity between 2006 and 2012 is low for the vertical component (matrix entries around C, maximum CC 0.68 as marked in Fig. 16) and the E-W component (maximum CC 0.66), on the N-S component most event pairs have CCs higher than 0.7 and two event pairs have CCs of 0.83. This similarity confirms the visual inspection of the waveforms and allows us to speculate that the 2006 and 2012 swarms occurred in the same area but potentially had different focal mechanisms, due to the lower CCs on the vertical and E-W components. Unfortunately, during the analysis a technical defect on the E-W component of station TMO20 was identified for the data recorded after 2006 which partly can explain the lower inter-event-similarity of the 2012 events and further the low similarity to the 2006 events due to higher instrument noise on this component. On the vertical component additionally a difference in the amplitude ratios between P and S waves is visible and for the 2012 events also a further phase onset is observed between the P -phase and S -phase at 3 s arrival time. Despite the high CCs on the N-S component after creating a network matrix the resulting mean or weighted CCs do not exceed 0.8 and thus there is no link possible between the 2006 and 2012 events for a threshold higher than 0.7 (see Fig. 12). However, it is possible to assign the 2006 events to a limited area of the seismic volume with the similarity analysis that was not possible with the relative location method due to an insufficient number of recordings.

4.3.2 Insheim

For the Insheim area, the spatial distribution of the clusters is shown in Fig. 17 for thresholds of 0.7 (left) and 0.8 (right). Cluster IA (threshold of 0.7) includes most of the clustered events (Figs. 17a, c and e, blue) with available relative hypocenters from Zei (2013). Their epicenter distribution displays a structure with NNW-SSE strike direction (ca. 30° from north to west). In the two vertical cross-sections (Figs. 17c and e) the events of cluster IA are distributed at 4.5-6 km depth. With the exception of six relatively located events of cluster IG, the other clusters contain only one (IC, IK) or two (IE, IJ, IL, IM) events for which relative hypocenters are available. However, these events (IC-IM) occur horizontally and vertically at the edge of the main cluster IA and this difference may explain the waveform deviation relative to IA waveforms.

Using a threshold of 0.8 (Fig. 17, right), cluster IA splits up in 13 smaller, spatially separated subclusters with known hypoDD locations. The spatial dimension of these subclusters is about 200-300 m or one dominant wavelength, leading to waveform differences which are recognized by the clustering algorithm. The waveform difference is attributed to the 3-D structure at depth and not to the rupture mechanisms, because the latter are modelled as quite similar normal faulting mechanisms for the subclusters (Fig. 17b). Subclusters IA03-IA17 overlap in depth as well as northing and easting. The split up of IA with regard to the spatially separated locations implies that several bridge-events lead to a linkage of all corresponding events for a common threshold of 0.7. Thereby the bridge-events contain waveform characteristics of different single events (see e.g. Nishigami, 1987).

Due to the completely different and separated locations of the single subclusters in Figs. 17b, 17d and 17f, this example points out the typical limits of most clustering

algorithms, the bridging technique included. As observed in other regions (e.g. Aster and Scott, 1993; Cattaneo et al., 1999), by applying a too lower threshold (here 0.7), events which are clearly separated in space for a higher threshold (0.8) will be merged together in a single large cluster due to several existing bridge-events in our dataset. The hypocenter depth varies between around 4.4 km and 6.5 km and there is a variation of cluster activity with time. The plotted events of subclusters IA03-IA17 as well as IC01 and IE01 occurred only until 2010. With the exception of one event of subcluster IA05, all other plotted events occurred in completely different areas after an aseismic phase when the geothermal power plant started operation in November 2012.

4.4 Interpretation of clustered events

The knowledge of the time occurrence of the clusters, the location of the clusters and the fps together with the known location and strike of the main faults in the two study areas (see Fig. 1) allows us to speculate on the nature of the rupture processes. Around the Landau seismic cluster the uppermost crust is composed of about 2000 m syn- and post-rifting sediments and 800 m Mesozoic and Palaeozoic sediments. The microseismic events are located in the crystalline basement which is found at about 2800 m depth and which is dissected by rifting-related faults in NNW-SSE direction (Fig. 1) with a dominant normal faulting character (Schad, 1962; GeORG-Projektteam, 2013). The relative locations of the events mimic this strike direction (with the exception of clusters LA and LH) as well as the strike of the normal faulting mechanisms. In addition, the nodal planes determined from the fps of clusters LB01, LB02, LB07, LH01 and LI01 also strike in NNW-SSE direction. This preferred orientation of the seismic activity, which coincides with the tectonic structures as well as observations of seismic anisotropy (Frietsch et al., 2015), is a strong argument for

reactivation of pre-existing faults. As a physical model, we propose the reduction of the normal stress on the faults due to an increase in fluid pressure (Simpson, 1986) which is related to the injection of water (Expertengruppe, 2010; Plenkers et al., 2013). Furthermore, considering the recurring temporal activity of several clusters (Fig. 12d and e) supports the theory of microearthquakes with nearly identical locations and faulting mechanisms induced due to injected fluids. However, some detected clusters are active only within short time periods (LA01-02, LA09) after operating interventions of the geothermal powerplant like the stimulation 2006 or the shut down on 25/06/2012 (Figs. 12d and e).

The flow path of the injected fluid is not well known. As argued by Cornet and Scotti (1993), induced seismicity in general maps not areas of high-fluid flow but rather indicates areas of high-fluid pressure. After first-time injection, the main fluid flow will take place along pathways with the lowest hydraulic impedance. Thus, seismic events contemporary induced after the injection may help to detect highly permeable fractures that are well connected to the injection well (Cornet and Scotti, 1993). Furthermore, aseismic behaviour of several hydraulically significant fractures was reported by Cornet and Scotti (1993), particularly for pathways with direct connections to the production wells but without connections to the injection well. Baisch et al. (2002) demonstrated that high fluid flow over large distances (~1.5 km) is not mandatorily mapped by induced seismicity.

In the following, we consequently assume that the microseismicity mirrors the fluid migration along existing structures that are connected to the injection well and which are close to failure. Flow rate parameters are not publicly available, except for the stimulation experiment in the Landau area 2006 (see Plenkers et al. 2013). Thus, compared with similar analysis of induced seismic events in Soultz-sous-Forêts

(Cornet et al., 1997; Cuenot et al., 2008) or Basel (Häring et al., 2008; Deichmann et al. 2014), we cannot validate if any correlation with the seismicity between 2007 and 2013 occurs. Below the Landau injection hole, the events of cluster LH are aligned N-S as well as their fault planes (Fig. 15b), indicating fluid migration in this direction. The hypocenters represent the deepest located events for that area and a leaching of the injected fluids down to deeper sited structures driven by gravity may be assumed during the ongoing operation of the powerplant. As argued by Baisch et al. (2002), due to the absence of seismicity between the landing point of the injection well and the hypocenters of cluster LH, we speculate that the fluid migration down to around 4.7 km depth takes place along highly permeable pathways. The main seismic clusters deviate into the NNW-SSE direction (clusters LB, LD, LE, LI and LL) with decreasing hypocenter depth towards NNW. Comparing the temporal evolution of clusters LB01, 04 and 07 (Figs. 12d and e) with the hypocentral (Figs. 15) and temporal (Fig. 18) distribution for some relatively located events reveals obviously that events associated to these clusters occurred during the whole studied time period 2007-2013. However, for cluster LB01 first located seismic activity occurred around 1.1 km northwest of the reinjection site (Figs. 15b and 18). The other clusters LB04 and 07 are initially activated later with a shift towards the reinjection hole. Most likely, these clusters either represent the main fluid flow or complicated fault systems which were close to failure and thus easy to trigger seismicity. A second pathway for fluids is assumed in NE-SW direction (cluster LA) which may be due to a connecting fault between the NNW-SSE running normal faults. Most locations representing this direction are part of cluster LA08 with activity during the whole studied time period 2008-2013 (Figs. 12d and e, Fig. 18). Moreover, four events of cluster LA09, containing most of the events of the earthquake swarm on 25/06/2012, are located in that area (Fig. 15b). However, the cross-sections in Figs. 15d and 15f indicate that

both clusters are separated in depth. The similarity between the 2006 and 2012 swarms (Fig. 16) allows us to speculate that after operating interventions with strong injection pressure changes within short time periods (e.g. stimulation or shutdown of the powerplant) mainly the shallower part (LA09) is activated by mostly a large number of events ($-0.5 \leq M_L \leq 1.1$). Events associated to the deeper sited cluster LA08 seem to occur during the normal powerplant operation in irregular time periods with magnitudes M_L varying from -0.4 to 1.7 (Figs. 12d and e). Since the observed seismicity is not migrating directly towards the production well, most likely either a structure impermeable to the fluid circulation and/or pore pressure transfer is present or the structure is highly permeable which consequently implies easy fluid flow without any occurring induced seismicity (Cuenot et al. 2008). Certainly several other flow paths of high permeability exist within the investigated Landau seismic volume, which remain undetected due to the absence of microseismic activity.

The events near Insheim are clustered in groups which are characterized by a NNW-SSE orientation. Mainly normal faulting mechanisms along this direction have been found (Fig. 17b, Gaßner et al., 2014). The sediment thickness is about 3500 m, thus the microseismicity at 4.5-6 km depth occurs in the crystalline basement (Figs. 1 and 17). Fluid flow or fluid migration, if represented by the microseismicity, is outwards from the injection region into NNW and SSE directions with a downward migration component towards west (Figs. 17e and f). The clusters at the western side (IA19, IA20, IA28, IJ01, IC01 and IG01) are located along a NNW-SSE striking line. Joint interpretation of the spatial distribution and the temporal evolution of the clusters reveals that during the stimulations and circulation tests (Figs. 14d and e) seismicity (cluster IA03 and IA05) is spatially limited to a volume below the two injection holes (Figs. 17 and 19). After the initial operation of the powerplant started in November

2012, besides a reactivation of cluster IA05 (black arrow in Fig. 19), several further clusters far away from the injection holes were activated for the first time (e.g. IA19 or IG01). This pattern may indicate that different small patches of a fault were (re)activated with increasing time of operation step by step by increasing pore pressure due to the arriving fluid flow.

Conclusions

In order to detect and cluster microseismic events in an environment with a high noise level and observational gaps, we introduce a network similarity matrix which is weighted by the signal to noise ratio (SNR) of the single recordings (Fig. 5). As result, we achieve a relatively homogeneous evaluation of the waveform coherency within a heterogeneous dataset. Our application to fluid-induced microseismicity successfully identifies event clusters which reoccur during a six year observation period. The event grouping based on waveform similarity spatially correlates well with hypocenter determinations and allows us to present a preliminary interpretation concerning the origin of the fracture mechanisms.

Problems with event clustering due to data gaps in the time series and mostly quite low SNRs are handled by implementing a newly developed weighting approach. This weighting approach is based on the individual SNRs of the seismic recordings including their different components (Z, E-W and N-S). It works well for the two analysed datasets after defining a dataset-specific weighting function that is based on a sigmoid function. The weighted network matrix approach is successful in the tested magnitude (M_L) range between about -0.5 and 2.7. Indirectly, also the source-

receiver distances are weighted, because in general the signal amplitude decreases for longer distances and reduces the SNR. Particularly for weak events ($M_L < 0.5$), the weighted network matrix performs much better in our examples than the standard version without weighting (Figs. 7 and 11). This is because the few well-recorded waveforms dominate the final correlation results (Figs. 9 and 10). For stronger events ($M_L > 0.5$ in our case) the standard network matrix without weighting performs also well, however, this depends on the SNR of an individual dataset. For our analysis no *a priori* waveform selection must be conducted before the waveform similarity analysis, as it was applied in other studies, to limit the dataset or to analyse just a group of very nearby located events (Maurer and Deichmann, 1995). Only 19 events were removed from our dataset due to their small differences in origin time to avoid overlapping waveforms.

The applied combination of SNR-weighted CCs and the three step clustering procedure allows us to detect distinct event clusters which are not resolved without the SNR-weighting, because the CCs for most waveform pairs are very low. The weighting procedure increases the influence of well-recorded waveforms, e.g. on a specific component, which may have a high SNR due to a favourable source radiation pattern. In this way also borehole recordings with a high SNR are treated preferentially. For instance single recording components, as the E-W component of station TMO54 for events in the Insheim area in our case, may dominate the weighted network matrix (Figs. 9 and 10). A low threshold (e.g. 0.7 for the Insheim area) leads to a cluster that contains a large number of events of the entire dataset. Applying a higher threshold (e.g. 0.8 or 0.9) such a large cluster splits up into several smaller ones which can be clearly separated in space.

Recordings from borehole seismometers are a clear benefit for an area with low SNR at the surface. For event detection, we find an average reduction of the detection threshold from around $M_L = 0$ with surface recordings to $M_L < -0.5$ with borehole recordings (station TMO64). Thus, we recommend the installation of borehole stations in a setting with high anthropogenic noise and unconsolidated sediment sites as in our setting in the Upper Rhine Graben to ensure a precise cluster analysis of induced microearthquakes with mainly $M_L < 0$.

For the two analysed seismic volumes, we find reoccurrences of highly similar waveforms within short time periods (minutes to hours) as well as after several months to years. The comparison of the clustered microearthquakes for Landau and Insheim with their relative hypocenters from double-difference locations show a high spatial coincidence. The clusters with thresholds of at least 0.8 (Figs. 15 and 17, right columns) have an extension of about 200-400 m that is in order to one wavelength of the shear wave. Together with the fault plane solutions such a detailed event characterization allows the reconstruction of rupture processes: the spatial orientation of clusters containing events with highly similar waveforms coincide with the NNW-SSE trend which is found in tectonic studies as well as the interpreted orientations of the nodal planes determined from fault plane solutions. This consistency is a strong argument for the reactivation of the existing faults close to the fluid injection in varying time intervals. Such an identification of recurring event clusters can be applied to tectonic as well as induced rupture processes for hazard analyses or reservoir characterization.

Acknowledgements

We thank P. Knopf, R. Plokarz and W. Scherer for help with data acquisition and processing, Prof. G. Eisbacher for fruitful discussions and Dr. J. Daniell for language editing. Jens Zeiß provided hypocentre information. Seismic waveforms for this study were kindly provided by BESTEC GmbH, geo x GmbH and the KIT KABBA datacentre. Data of tectonic elements were provided by the EU-project GeORG (www.geopotenziale.eu, GeORG Projektteam 2013) and we thank J. Tesch and B. Schmidt (LGB Mainz) for information on the local geology. Further, we thank Dr. A. Barth for his helpful comments to an early version of the manuscript. Two anonymous reviewers helped to clarify several points in the manuscript. This study is part of the research project MAGS (Microseismic Activity of Geothermal Systems) which is funded by the Federal Ministry for the Environment, Nature Conservation and Nuclear Safety of the Federal Republic of Germany (FKZ 0325191A-F) and supervised by Projektträger Jülich (PT-J).

References

- Arrowsmith, S.J., and Eisner, L. (2006), A technique for identifying microseismic multiplets and application to the Valhall field, North Sea, *Geophysics* 71(2), 31–40.
- Asanuma, H., Kumano, Y., Hotta, A., Schanz, N.H., and Häring, M. (2007), Analysis of microseismic events from a stimulation at Basel, Switzerland, *GRC Transactions* 31, 265–270.
- Aster, R.C., and Scott, J. (1993), Comprehensive characterization of waveform similarity in microearthquake data sets, *Bull. Seism. Soc. Am.* 83(4), 1307–1314.
- Augliera, P., Cattaneo, M., and Eva, C. (1995), Seismic multiplets analysis and its implication in seismotectonics, *Tectonophysics* 248, 219-234.
- Baisch, S., Bohnhoff, M., Ceranna, L., Tu, Y., and Harjes, H.-P. (2002), Probing the crust to 9-km depth: fluid-injection experiments and induced seismicity at the KTB superdeep drilling hole, Germany, *Bull. Seism. Soc. Am.* 92(6) 2369–2380.
- Baisch, S., Ceranna, L., and Harjes, H.P. (2008), Earthquake cluster: what can we learn from waveform similarity?, *Bull. Seism. Soc. Am.* 98(6), 2806–2814.
- Barth, A., Ritter, J., and Wenzel, F. (2015), Spatial variations of earthquake occurrence and coseismic deformation in the Upper Rhine Graben, Central Europe, *Tectonophysics* 651-652, 172-185.

- Cattaneo, M., Augliera, P., Spallarossa, D., and Eva, C. (1997), Reconstruction of seismogenic structures by multiplet analysis: an example of Western Liguria, Italy, *Bull. Seism. Soc. Am.* 87(4), 971–986.
- Cattaneo, M., Augliera, P., Spallarossa, D., and Lanza, V. (1999), A waveform similarity approach to investigate seismicity patterns, *Natural Hazards* 19(2–3), 123–138.
- Cesca, S., Şen, A.T., and Dahm, T. (2014), Seismicity monitoring by cluster analysis of moment tensors, *Geophys. J. Int.* 196, 1813–1826.
- Cornet, F.H. and Scotti, O. (1993), Analysis of induced seismicity for fault zone identification, *Int. J. Rock Mech. Min. Sci. & Geomech. Abstr.* 30(7), 789-795.
- Cornet, F.H., Helm, J., Poitrenaud, H., and Etchecopar, A. (1997), Seismic and aseismic slips induced by large-scale fluid injections, *Pure appl. Geophys.* 150, 563-583.
- Cuenot, N., Dorbath, C., and Dorbath, L. (2008), Analysis of the microseismicity induced by fluid injections at the EGS site of Soultz-sous-Forêts (Alsace, France): implications for the characterization of the geothermal reservoir properties, *Pure appl. geophys.* 165, 797–828.
- Deichmann, N., and Garcia-Fernandez, M. (1992), Rupture geometry from high-precision relative hypocentre locations of microearthquake clusters, *Geophys. J. Int.* 110, 501–517.
- Deichmann, N., Kraft, T., and Evans, K.F. (2014), Identification of faults activated during the stimulation of the Basel geothermal project from cluster analysis and focal mechanisms of the larger magnitude events, *Geothermics* 52, 84-97.
- Expertengruppe (2010), Das seismische Ereignis bei Landau vom 15. August 2009, Abschlussbericht, http://www.lgb-rlp.de/fileadmin/cd2009/images/content/endbericht_landau/Landau_Endbericht_101103_corr.pdf (last access 5 March 2015).
- Ferretti, G., Massa, M., and Solarino, S. (2005), An improved method for the recognition of seismic families: application to the Garfagnana–Lunigiana area, Italy, *Bull. Seism. Soc. Am.* 95(5), 1903–1915.
- Frémont, M.J., and Malone, S.D. (1987), High precision relative locations of earthquakes at Mount St. Helens, Washington, *J. Geophys. Res.* 92(B10), 10,223–10,236.
- Frietsch, M., Groos, J., and Ritter, J. (2015), Detection and delineation of a fracture zone with observation of seismic shear wave anisotropy in the upper rhine graben, sw germany, *Pure appl. geophys.* 172, 267–282.
- Gaßner, L., Groos, J., and Ritter, J. (2014), Herdflächenanalyse induzierter Erdbeben in der Südpfalz: Reaktivierung präexistenter Bruchflächen und Spannungszustand, *Mainzer geowiss. Mitt.* 42, 195–213.
- Geller, J., and Mueller, C.S. (1980), Four similar earthquakes in central California, *Geophys. Res. Lett.* 7(10), 821–824.

- GeORG-Projektteam (2013), Geopotenziale des tieferen Untergrundes im Oberrheingraben. Fachlich–Technischer Abschlussbericht des Interreg–Projekts GeORG, Teile 1–4. – Internet (PDF-Dokument: <http://www.geopotenziale.eu>, last access 5 March 2015).
- Gibbons, S., Böttger Sørensen, M., Harris, D., and Ringdal, F. (2007), The detection and location of low magnitude earthquakes in northern Norway using multi-channel waveform correlation at regional distances, *Phys. Earth Planet. Inter.* 160, 285–309.
- Got, J.L., and Fréchet, J. (1993), Origins of amplitude variations in seismic doublets: source or attenuation process?, *Geophys. J. Int.* 114, 325-340.
- Groos, J., and Ritter, J. (2010), Seismic noise: A challenge and opportunity for seismological monitoring in densely populated areas, in Ritter, J., and Oth, A. (eds.) *Proceedings of the Workshop: Induced Seismicity, Cahiers du Centre Européen de Géodynamique et de Séismologie* 30 87-97.
- Groos, J., Grund, M., and Ritter, J. (2012), Automated detection of microseismic events in the Upper Rhine valley near the city of Landau/South Palatinate, *Geophys. Res. Abstr.* 14, EGU2012-10482.
- Groos, J., Zeiß, J., Grund, M., and Ritter, J. (2013a), Microseismicity at two geothermal power plants at Landau and Insheim in the Upper Rhine Graben, Germany, *Geophys. Res. Abstr.* 15, EGU2013-2742.
- Groos, J.C., Fritschen, R., and Ritter, J.R.R. (2013b), Untersuchung induzierter Erdbeben hinsichtlich ihrer Spürbarkeit und eventueller Schadenswirkung anhand DIN 4150, *Bauingenieur* 88, 374–384.
- Häring, M.O., Schanz, U., Ladner, F., and Dyer, B.C. (2008), Characterisation of the Basel 1 enhanced geothermal system, *Geothermics* 37, 469-495.
- Horstmann, T., Harrington, R.M., and Cochran, E.S. (2013), Semiautomated tremor detection using a combined cross-correlation and neural network approach, *J. geophys. Res.* 118, 4827-4846.
- Igarashi, T., Matsuzawa, T., and Hasegawa, A. (2003), Repeating earthquakes and interplate aseismic slip in the northeastern Japan subduction zone, *J. geophys. Res.* 108(B5), 2156-2202.
- Kraft, T., and Deichmann, N. (2014), High-precision relocation and focal mechanism of the injection-induced seismicity at the Basel EGS, *Geothermics* 52, 59-73
- Leydecker, G. (2011), Erdbebenkatalog für Deutschland mit Randgebieten für die Jahre 800 bis 2008, *Geologisches Jahrbuch, Reihe E* 59, E. Schweizerbart'sche Verlagsbuchhandlung, Stuttgart.
- Massa, M., Eva, E., Spallarossa, D., and Eva, C. (2006), Detection of earthquake clusters on the basis of waveform similarity: an application in the monferrato region (Piedmont, Italy), *J. Seismol.* 10(1), 1–22.

- Maurer, H., and Deichmann, N. (1995), Microearthquake cluster detection based on waveform similarities, with an application to the western Swiss Alps, *Geophys. J. Int.* 123, 588–600.
- Meier, L., and Eisbacher, G. (1991), Crustal kinematics and deep structure of the northern Rhine graben, Germany, *Tectonics* 10, 621–630.
- Moriya, H., Niitsuma, H., and Baria, R. (2003), Multiplet–clustering analysis reveals structural details within the seismic cloud at the Soultz Geothermal Field, France, *Bull. Seism. Soc. Am.* 93(4), 1606–1620.
- Nishigami, K. (1987), Clustering structure and fracture process of microearthquake sequences, *J. Phys. Earth* 35, 425–448.
- Plenefisch, T., and Bonjer, K.P. (1997), The stress field in the Rhine Graben area inferred from earthquake focal mechanisms and estimation of frictional parameters, *Tectonophysics* 275, 71–91.
- Plenkers, K., Ritter, J.R.R., and Schindler, M. (2013), Low signal–to–noise event detection based on waveform stacking and cross–correlation: application to a stimulation experiment, *J. Seismol.* 17(1), 27–49.
- Poupinet, G., Ellsworth, W.L., and Frechet, J. (1984), Monitoring velocity variations in the crust using earthquake doublets: an application to the Calaveras fault, California, *J. Geophys. Res.* 89(B7), 5719–5731.
- Press, W.H., Flannery, P.B., Teukolsky, S.A., and Vetterling, W.T. (1988), *Numerical Recipes in C, The Art of Scientific Computing*, 1st edn (Cambridge University Press, Cambridge, UK).
- Priddy, K.L., and Keller, P.E. (2005) *Artificial Neural Networks: An Introduction* (SPIE Publications, Bellingham, WA, USA).
- Ritter, J., Wagner, M., Bonjer, K.P., and Schmidt, B. (2009), The Heidelberg and Speyer earthquakes and their relationships to active tectonics in the central Upper Rhine Graben, *Int. J. Earth Sci.* 98, 697–705.
- Ritter, J.R.R. (2011), Konzeptionelle Ansätze zur Überwachung induzierter Seismizität im Oberrheingraben in Rheinland–Pfalz, *Mainzer geowiss. Mitt.* 39, 157–176.
- Ritter, J.R.R., Wagner, M., Wawerzinek, B., and Wenzel, F. (2008), Aims and first results of the TIMO project – Tiefenstruktur des mittleren Oberrheingrabens, *Geotect. Res.* 95(1), 151–154.
- Schad, A. (1962), Das Erdölfeld Landau, *Abh. geol. Landesamt Baden–Württemberg* 4, 81–101.
- Schaff, D.P., Bokelmann, G.H.R., Ellsworth, W.L., Zankerka, E., Waldhauser, F., and Beronza, G.C. (2004), Optimizing correlation techniques for improved earthquake location, *Bull. Seism. Soc. Am.* 94(2), 705–721.
- Simpson, D. (1986), Triggered earthquakes, *Ann. Rev. Earth Planet. Sci.* 14, 21–42.

- Snoke, J. (2003), FOCMEC: FOCal MEChanism determinations (Inter. Handbook of Earthquake and Engineering Seismology, 81B, 29–30, Academic Press, London).
- Staudt, S., Brüstle, A., and Ritter, J.R.R. (2016), Bestimmung von 1-D seismischen Geschwindigkeitsmodellen in der Südpfalz, Mainzer geowiss. Mitt., submitted.
- von Seggern, D.H. (2006), CRC Standard Curves and Surfaces with Mathematica, 2nd edn (Chapman and Hall/CRC, Boca Raton, FL, USA).
- Waldhauser, F. (2001), HypoDD – a computer program to compute double – difference earthquake locations. U.S. Geological Survey Open–File Report 01–113, <http://pubs.usgs.gov/of/2001/01113/> (last access 5 March 2015).
- Waldhauser, F., and Ellsworth, W. (2000), A double–difference earthquake location algorithm: Method and application to the northern Hayward fault, Bull. Seism. Soc. Am. 90, 1353–1368.
- Wehling-Benatelli, S., Becker, D., Bischoff, M., Friederich, W., and Meier, T. (2013), Indications for different types of brittle failure due to active coal mining using waveform similarities of induced seismic events, Solid Earth 4(2), 405–422.
- Zeiß, J. (2013), Präzise Lokalisierung mikroseismischer Ereignisse im Umfeld geothermischer Kraftwerke in der Südpfalz, diploma thesis, Karlsruhe Institute of Technology, Geophysical Institute, in German

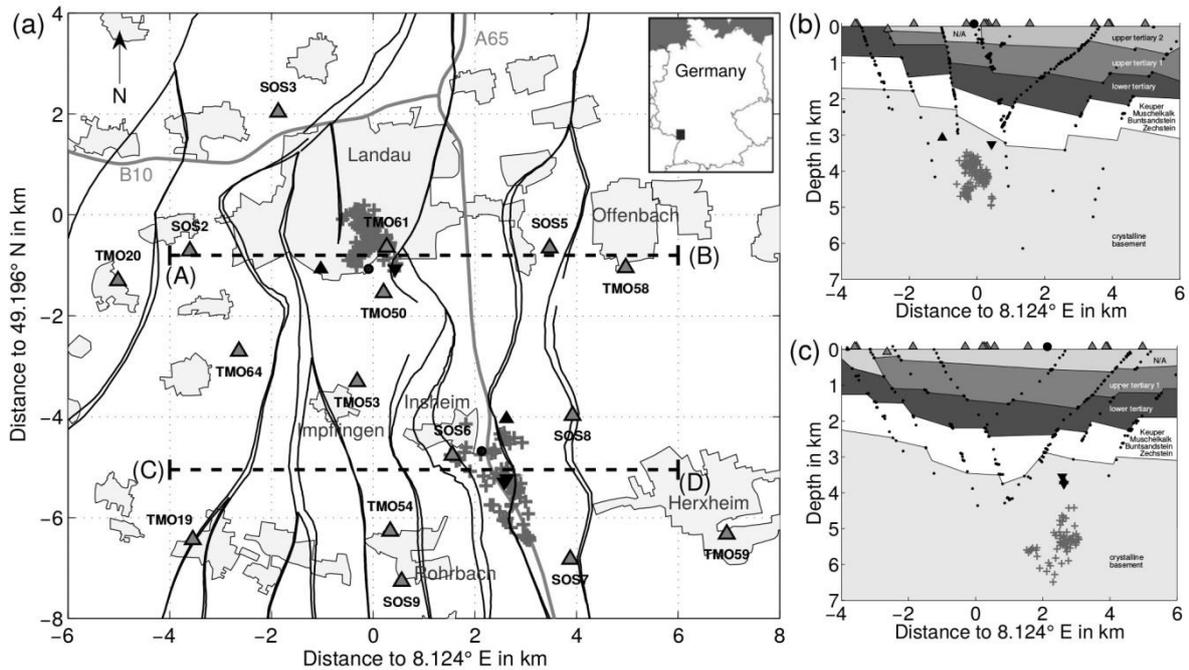


Figure 1: a) Seismic recording stations used for the clustering analyses and epicentral distribution of all relatively located earthquakes. Note, there are also other stations operated in that area which were used for hypocenter and fault plane solution studies. TMO64 was the only borehole station used at this time. There are two seismic volumes, one underneath the city of Landau and one next to the village of Insheim. Black circles mark positions of the two geothermal powerplants and black triangles the landing points of the production (upward) and reinjection (downward) wells. Settlements are indicated in light gray, lines show two major roads (A65 and B10). Known faults at basement level are taken from the GeORG project (www.geopotenzielle.eu, GeORG Projektteam 2013) and marked in black. b) Depth section for the dashed line (A) to (B) in a) with simplified geological units and faults (black dotted lines) modified after GeORG. The east-west distribution of the surface/borehole recording stations and the powerplant Landau is indicated for reference. c) Corresponding depth section for the dashed line (C) to (D) through the seismic volume of Insheim. Note that the geological units and faults in the GeORG model are based only on single known grid points (e.g. from core samples or reflection seismic data) which were used for interpolation. For clarity, vertical exaggeration is about 1.3.

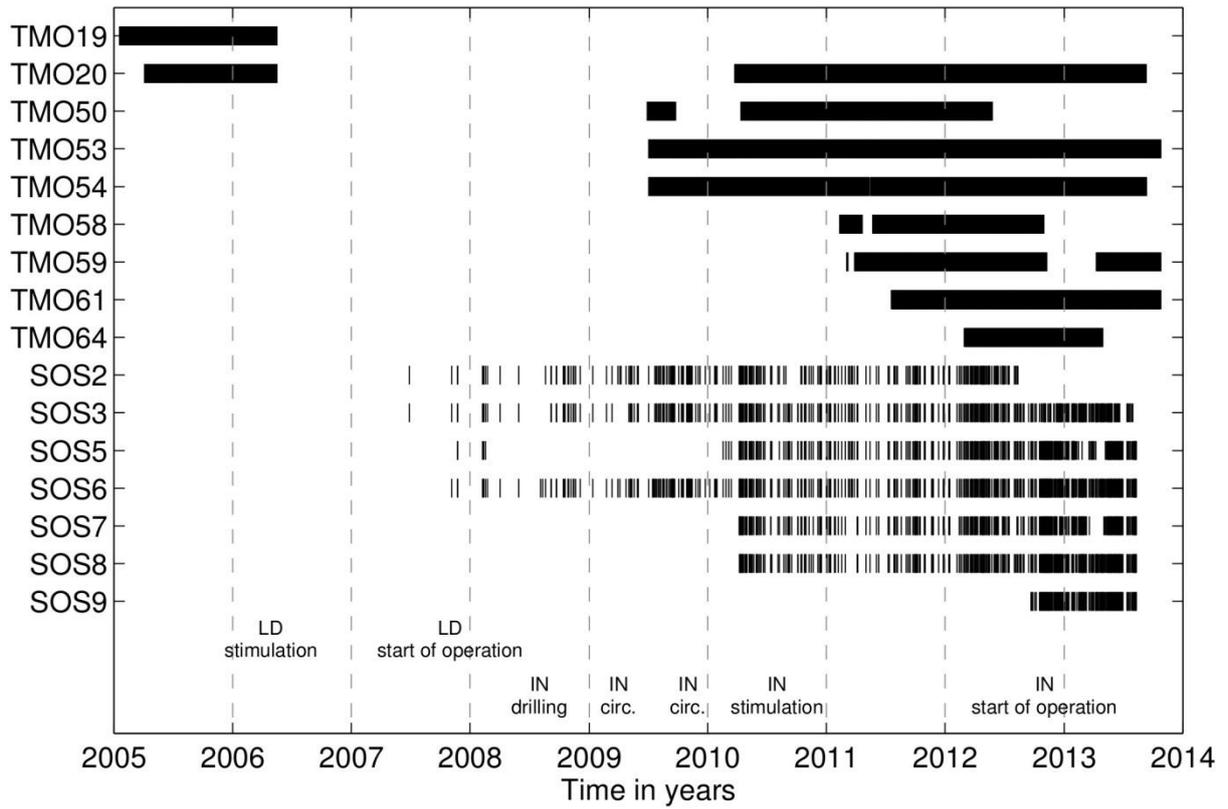


Figure 2: Daily data availability of the recording stations used during the time period 2005-2014 (Fig. 1). Note the data gaps, especially for the SOS stations. Important time stamps (drilling, stimulation, circulation, start of operation) for both seismic areas are marked by LD (Landau) and IN (Insheim).

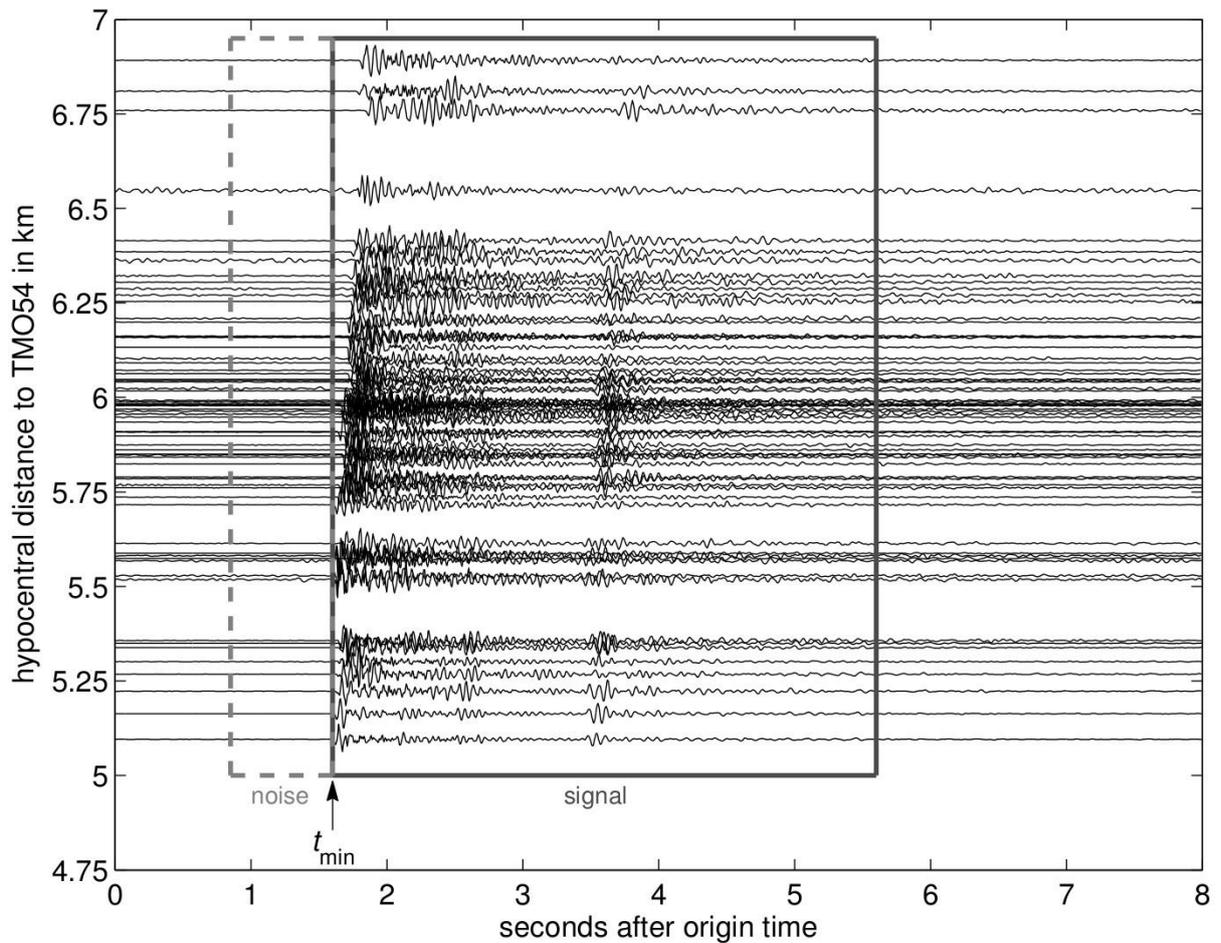


Figure 3: Seismogram section to set the minimum travelttime t_{\min} at station TMO54 for the area Insheim using the time series of all absolutely located events on the vertical component. The black time window setting was used for all three components. Normalized cross-correlation functions were computed using the time series part within the marked signal window. SNR-values are computed using the time series parts of both, the signal and noise windows (dashed).

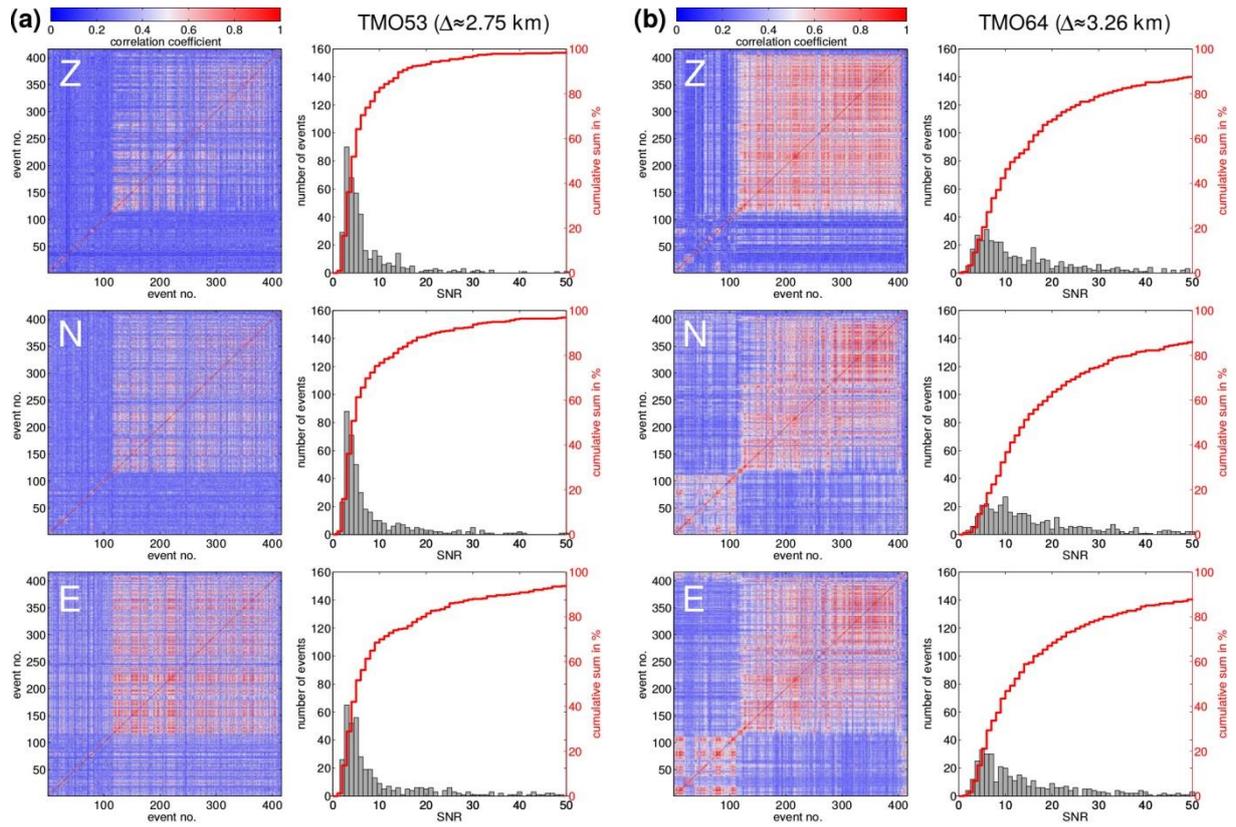


Figure 4: Component similarity matrices (left panels) and SNR distributions for each component (right panels) of 416 Landau events recorded between March and August 2012 for a) surface station TMO53 and b) borehole station TMO64, respectively. Cumulative sums in the SNR-histograms (red step functions) indicate, how many of the total 416 events are included up to a specific SNR at the corresponding component (expressed as a percentage). For clarity, only SNRs of up to 50 are included.

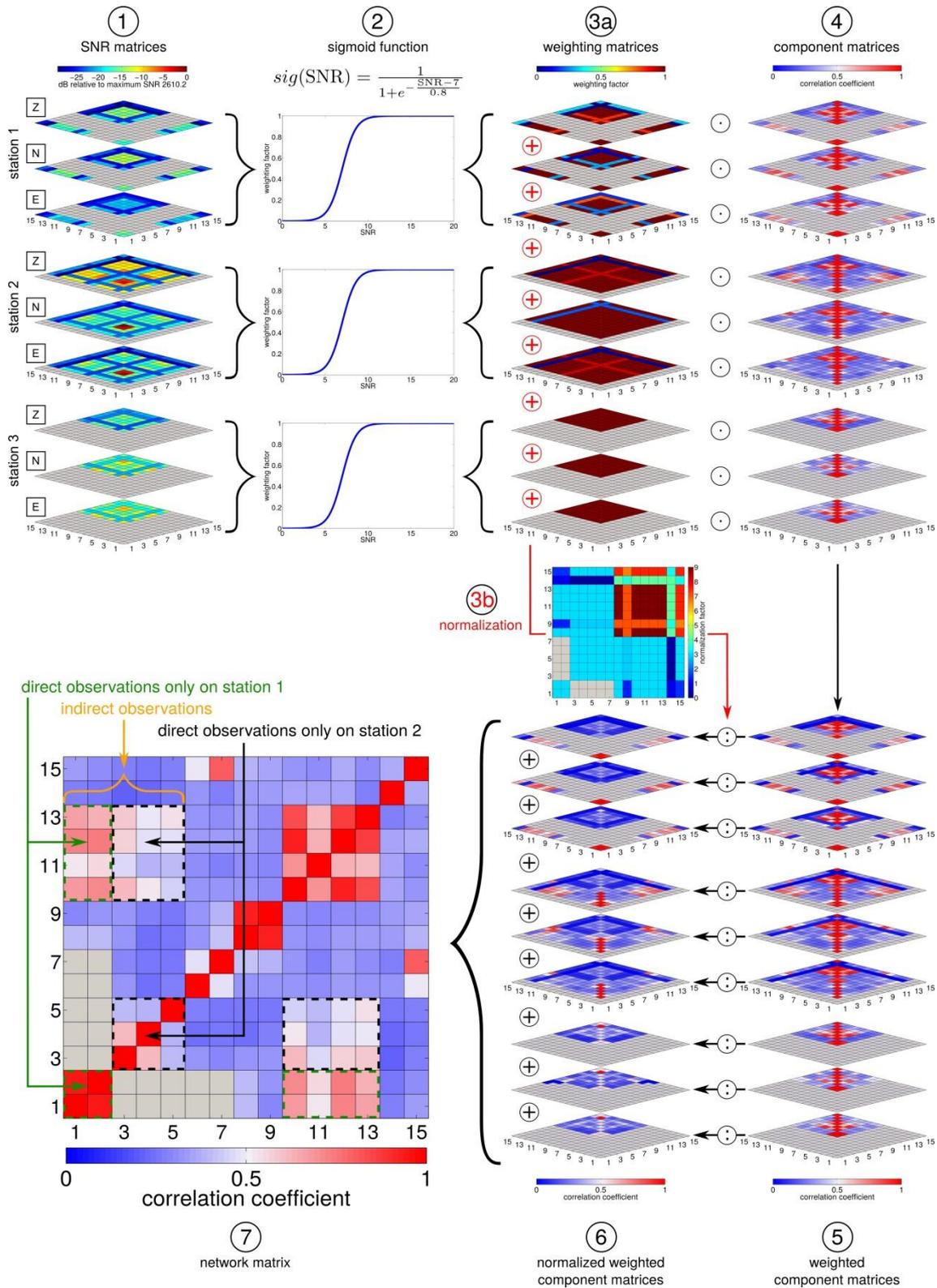


Figure 5: Work flow to construct a weighted network matrix considering the individual SNRs of all possible event pairs for three exemplary seismic stations. Gray shaded areas indicate event pairs for which no observations (NaN values) are available at the corresponding station/components or rather the entire seismic station network. The benefits of using a network matrix (mean or weighted) instead of several single station matrices for a station

network of heterogeneous recording periods is indicated in step 7. CC matrix entries enclosed by green dashed lines are directly observed only at station 1 while CC entries within the black dashed boxes only result from direct observations on station 2. However, a comparison between the two blocks covered by the orange brace allows to speculate about similar characteristics indirectly due to the displayed CCs. For detailed description of steps 1 to 7, see text.

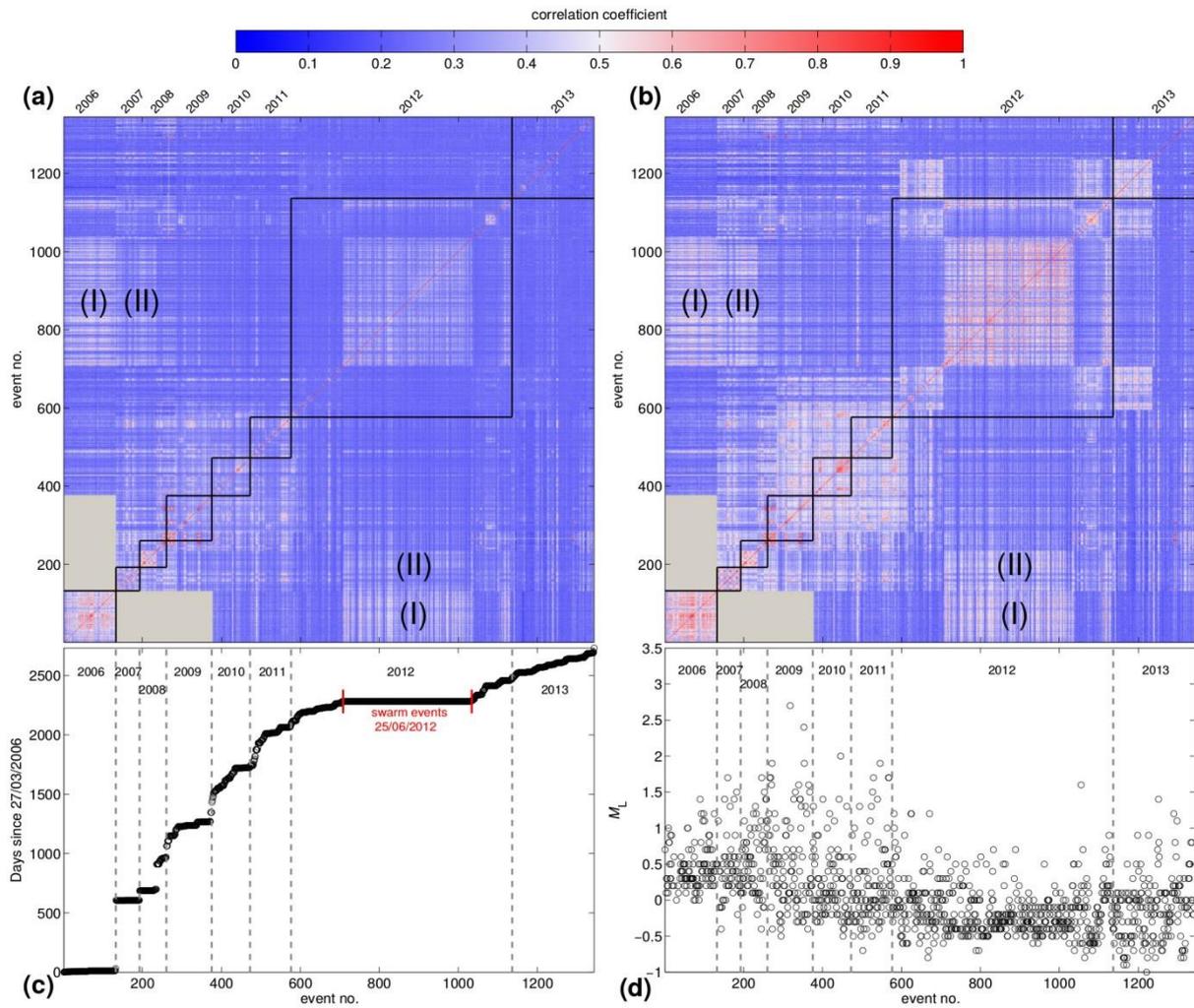


Figure 6: Mean (a) and weighted (b) network matrices for all possible event pairs of the Landau area. Gray shaded areas indicate event pairs for which no observation is available over the entire seismic station network (event no. 133-377). Areas (I) and (II) indicate event pairs which are discussed in detail in the text. Individual years of the investigation period 2006-2013 are shown as black squares. Temporal and magnitude distributions over the investigation period for all events are illustrated in c) and d).

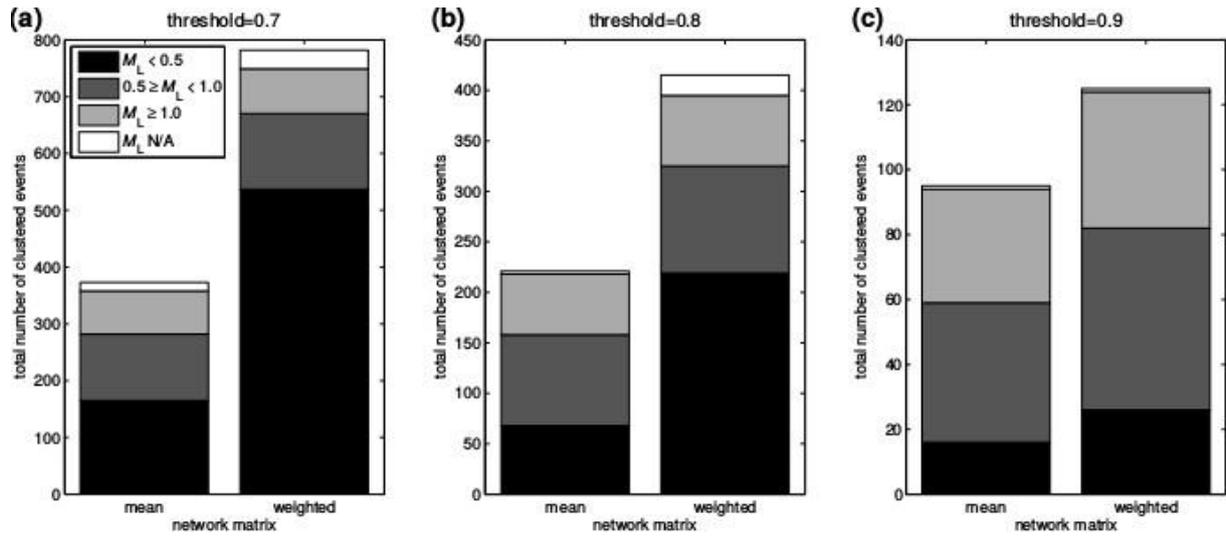


Figure 7: Comparison of clustering results for the Landau area with the mean and weighted network matrices. Note the different axis scales of the ordinates for the three used thresholds: a) 0.7, b) 0.8, c) 0.9.

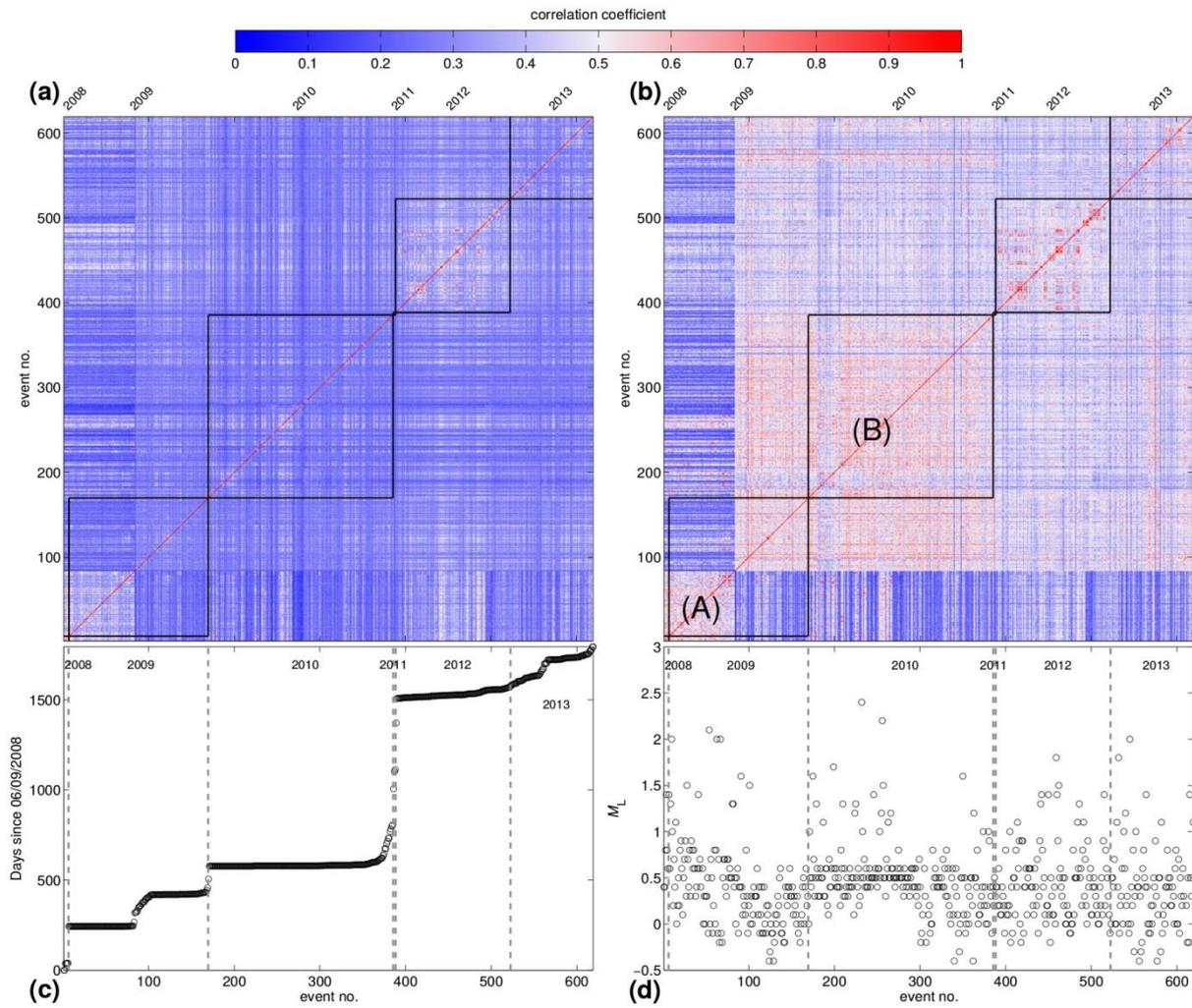


Figure 8: Mean (a) and weighted (b) network matrices for all possible event pairs of the Insheim area. Areas (A) and (B) indicate event pairs which are discussed in detail in the text. Individual years of the investigation period 2008-2013 are shown as black squares. Temporal and magnitude distributions over the investigation period for all events are illustrated in c) and d).

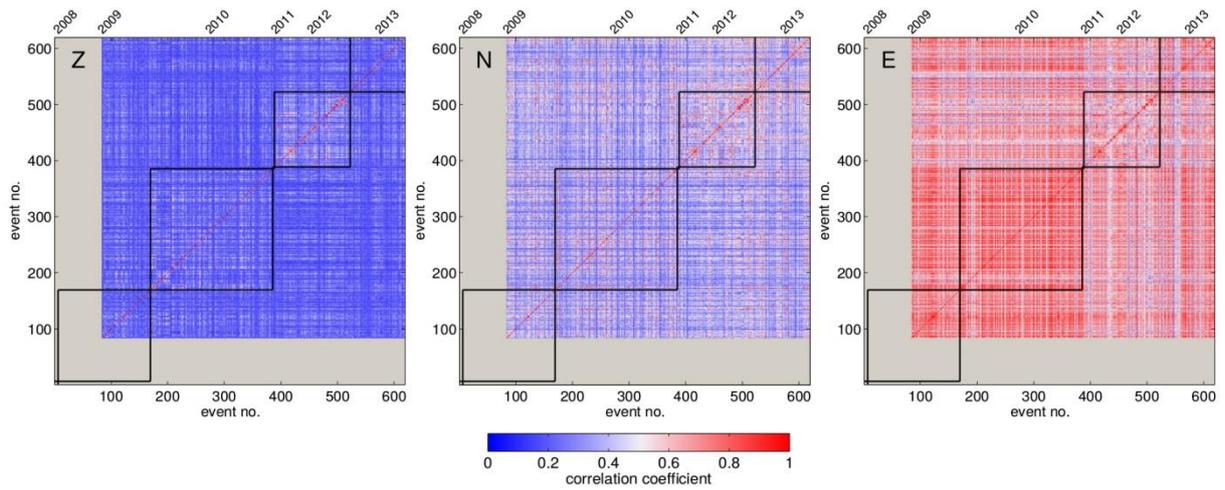


Figure 9: Component matrices (Z, N-S, E-W) of the Insheim area events for recording station TMO54. Gray shaded areas indicate event pairs for which no observation is available at this station. Individual years of the investigation period 2008-2013 are shown as black squares.

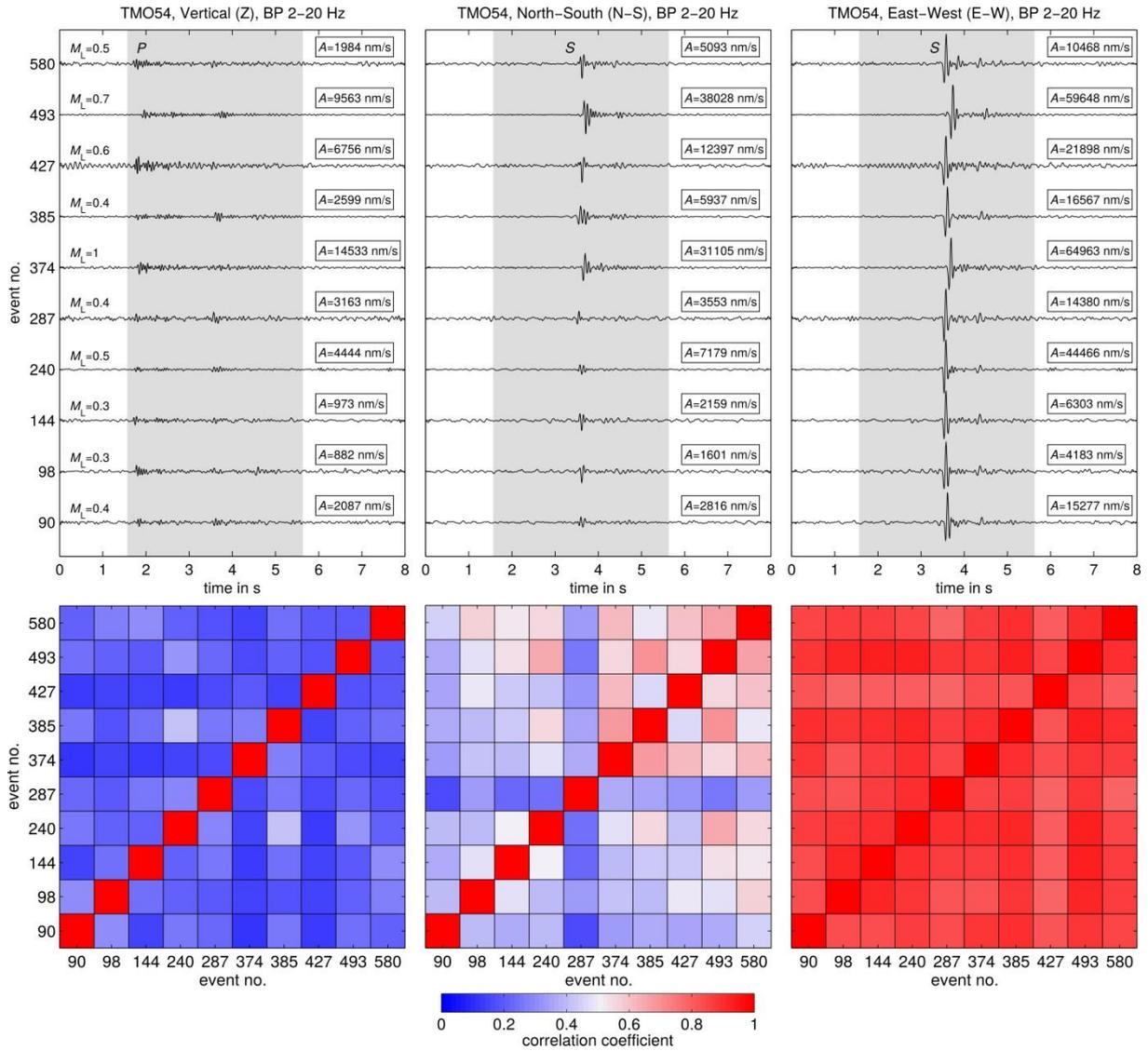


Figure 10: Waveforms and corresponding component matrices (Z, N-S, E-W) of selected Insheim events at recording station TMO54 during the investigation period 2008-2013 (event no. corresponds to Fig. 9). To display amplitude variations between the recording components for each event, the corresponding waveforms are normalized to the maximum peak of the three components. The peak amplitude A is given above each trace. The gray shaded areas mark the 4 s waveform windows used for cross-correlation.

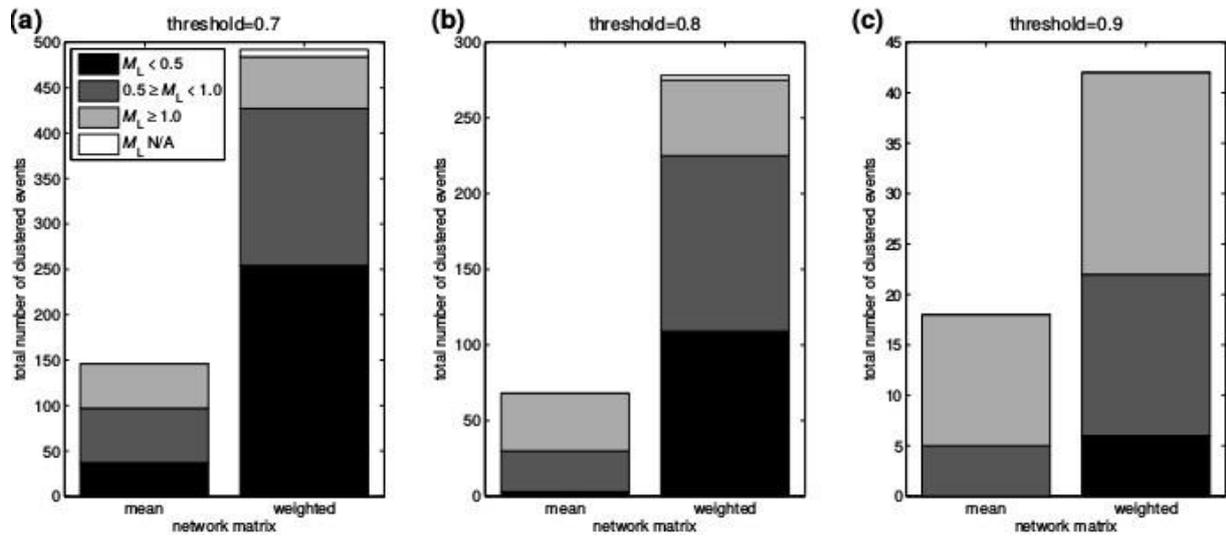


Figure 11: Comparison of clustering results for the Insheim area with the mean and weighted network matrices. Note the different axis scales of the ordinates for the three used thresholds: a) 0.7, b) 0.8, c) 0.9.

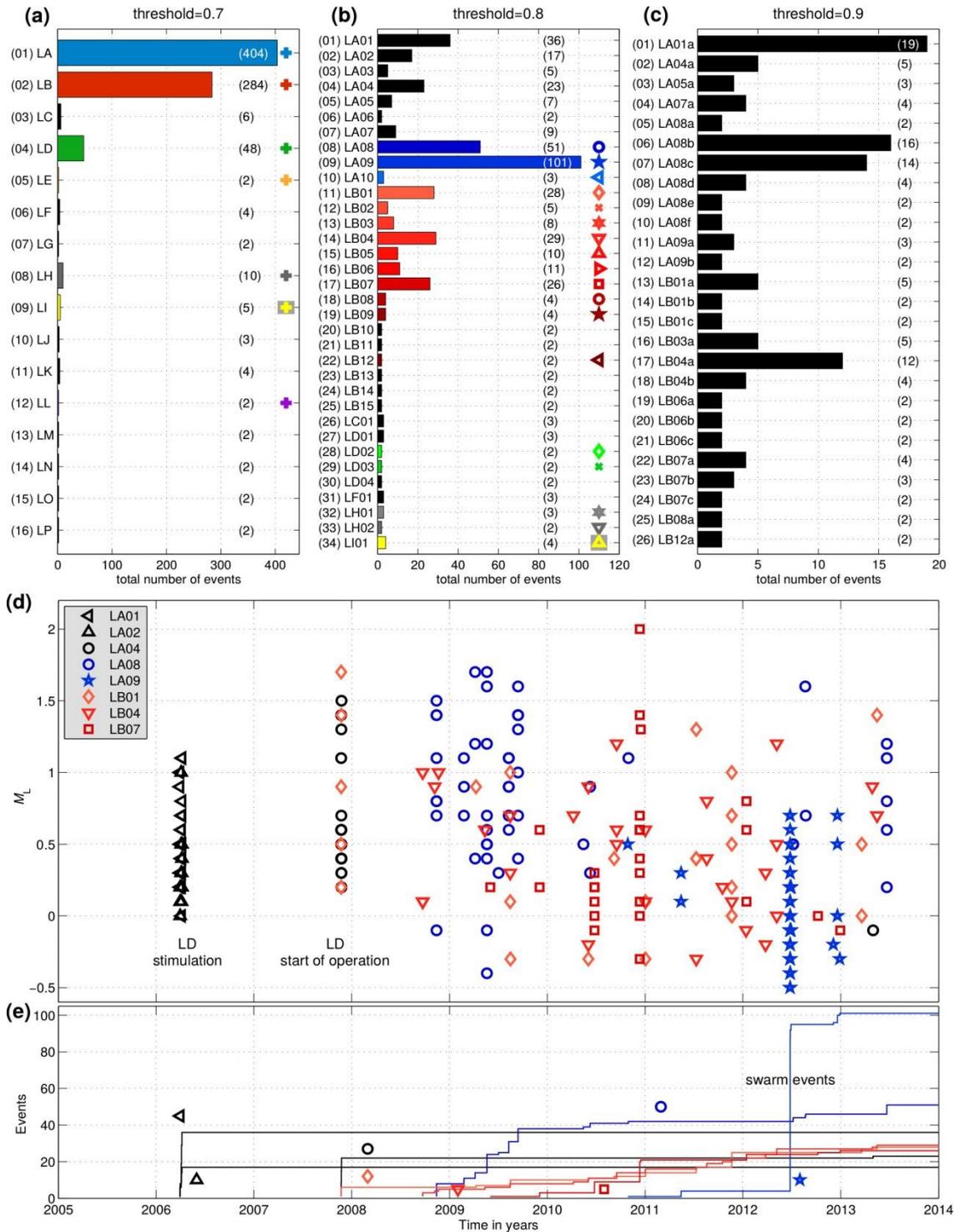


Figure 12: Detailed SNR-weighted clustering results for the Landau area: Separately for each threshold a) 0.7, b) 0.8 and c) 0.9, the number of the clusters is plotted with respect to the number of events in each cluster. Colors indicate different clusters for thresholds of 0.7 and symbols (except LA01, 02 and 04) indicate subclusters for thresholds of 0.8 for which relative locations of several events are available (see Fig. 15). Note the different axis scales in a)-c). Magnitude distributions and event size evolution of the eight largest clusters are illustrated in d) and e) for a threshold of 0.8.

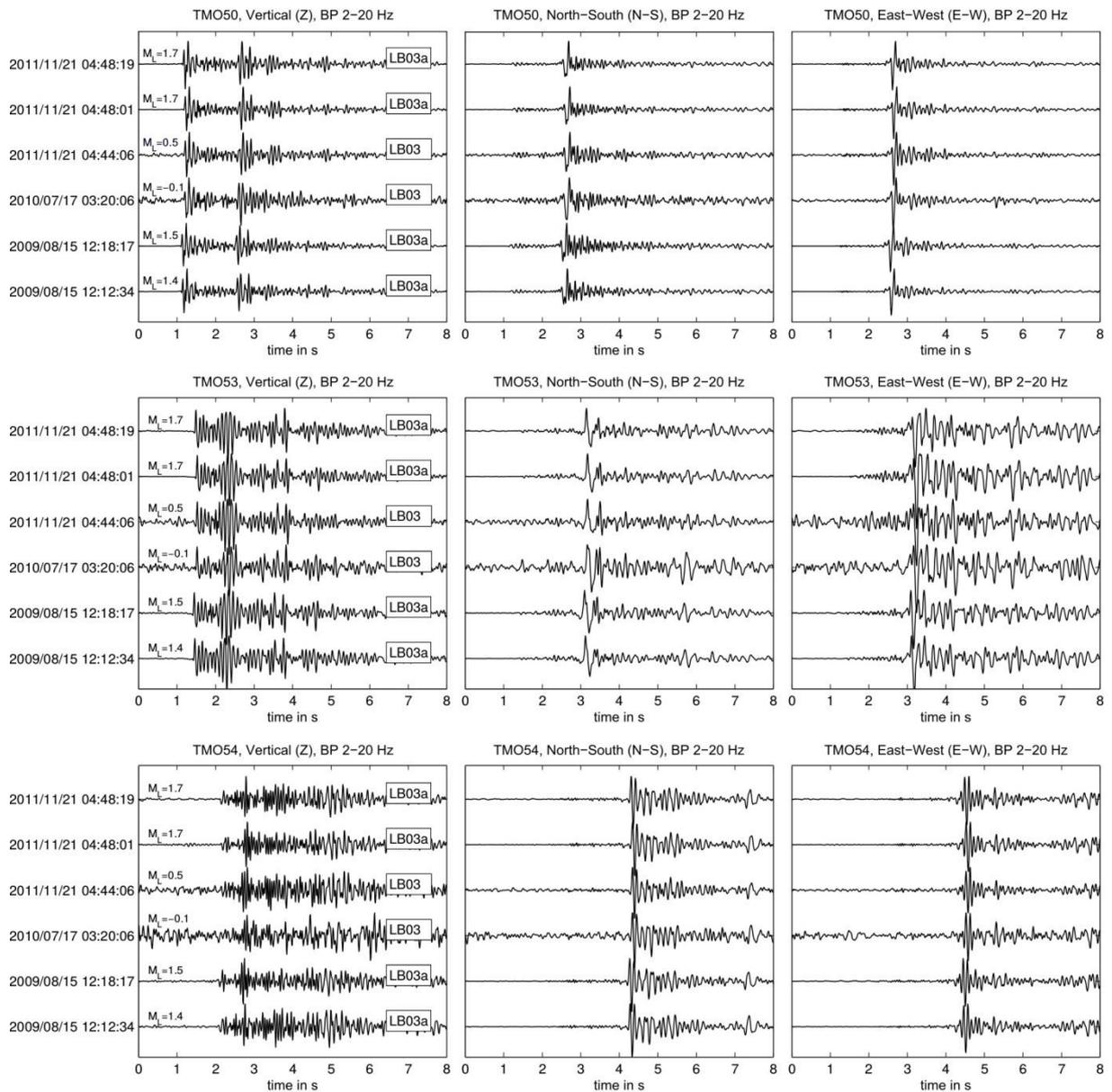


Figure 13: Amplitude-normalized waveform examples of six Landau events for cluster LB03 (clustered with thresholds 0.7 and 0.8) and rather subcluster LB03a (additionally clustered with threshold 0.9) at stations TMO50, TMO53 and TMO54 for all three components (Z, N-S, E-W). Note the time and magnitude bandwidth of the events. See text for details regarding the individual cluster names. Time range is in seconds after the corresponding source time.

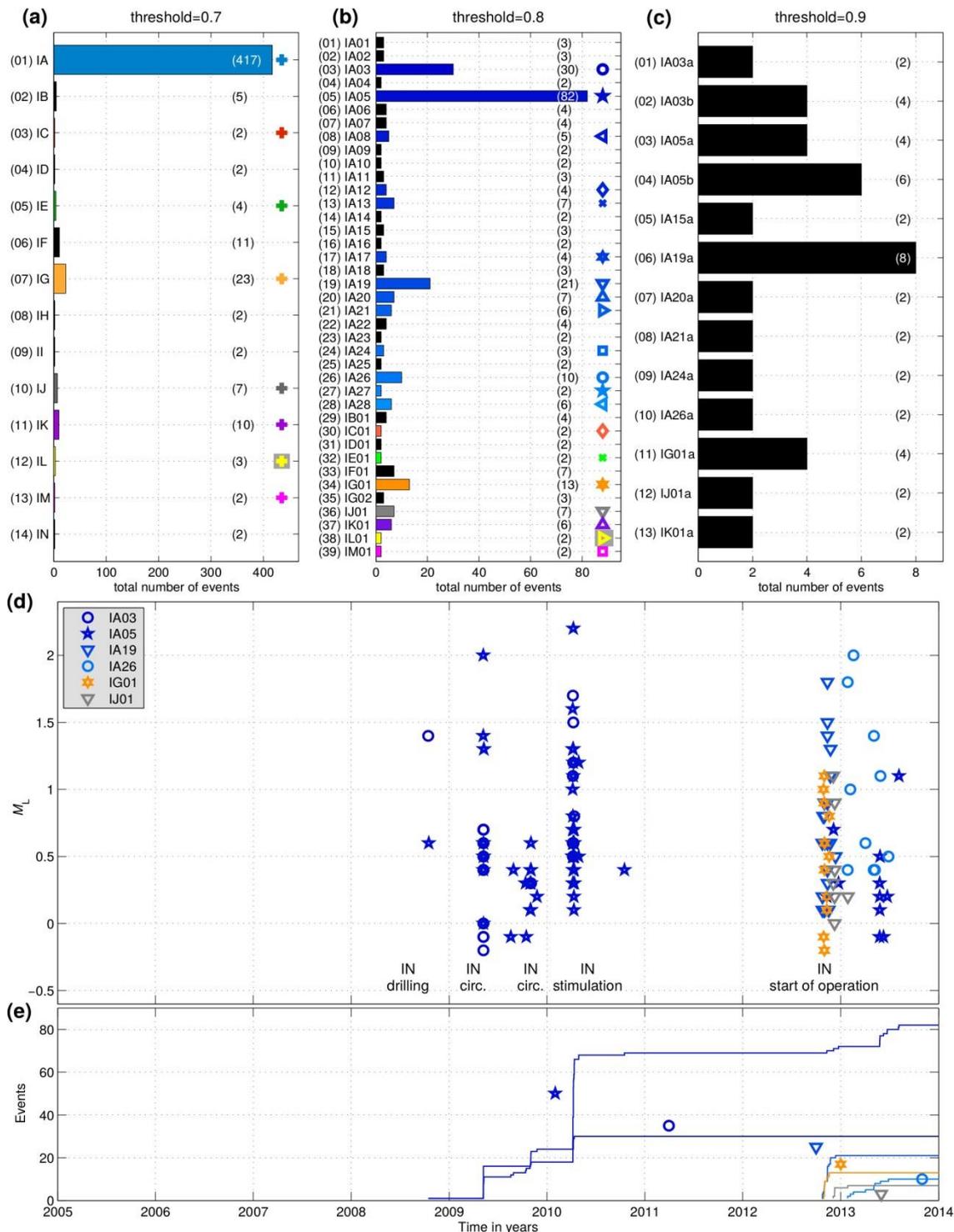


Figure 14: Detailed SNR-weighted clustering results for the Insheim area: Separately for each threshold a): 0.7, b): 0.8 and c): 0.9, the number of the clusters is plotted with respect to the number of events in each cluster. Colors indicate different clusters for thresholds of 0.7 and symbols indicate subclusters for thresholds of 0.8 for which relative locations of several events are available (see Fig. 17). Note the different axis scales in a)-c). Magnitude distributions and event size evolution of the six largest clusters are illustrated in d) and e) for a threshold of 0.8.

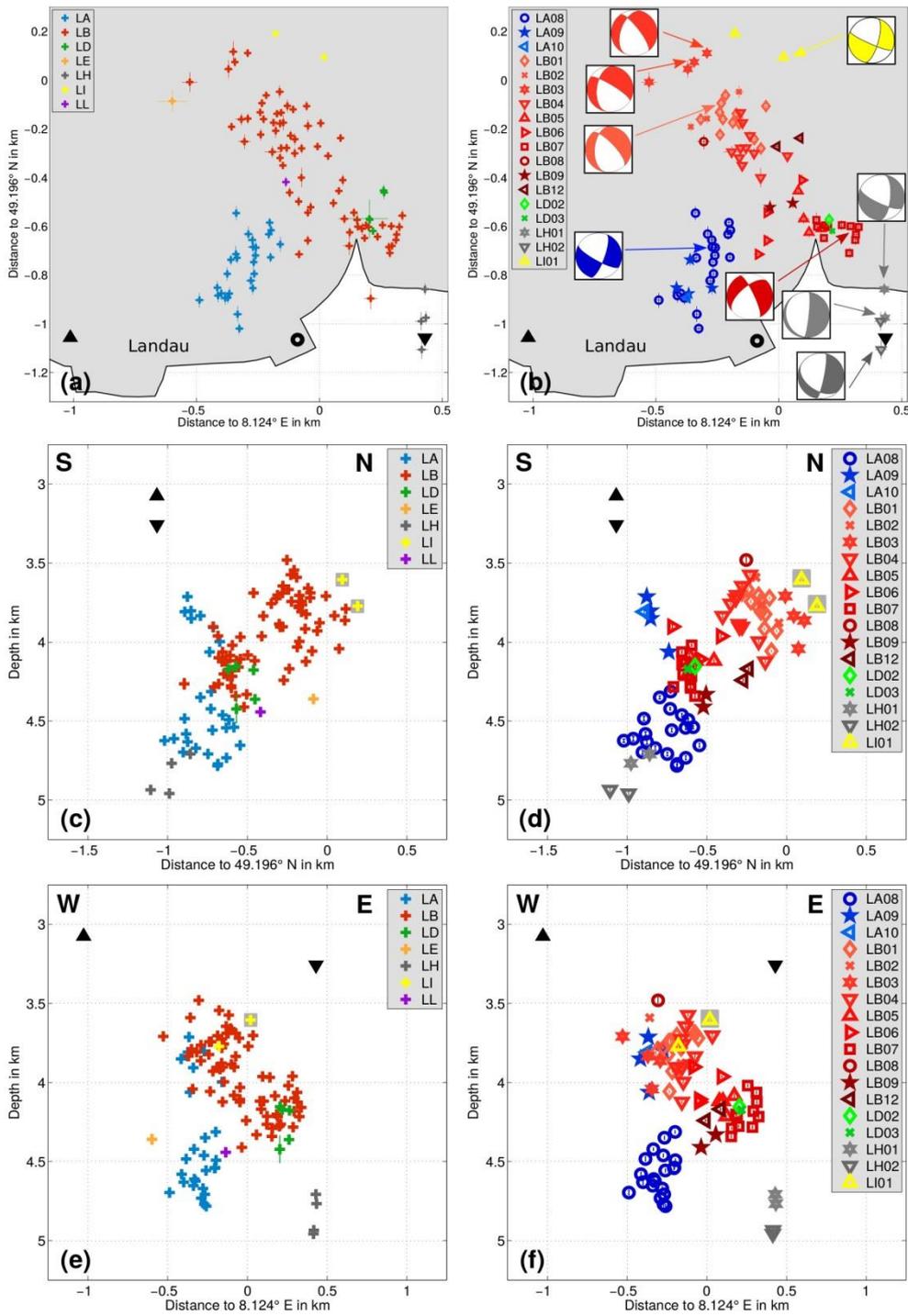


Figure 15: Spatial cluster distribution for the Landau area. a) and b): relative epicenter distribution for the thresholds 0.7 (left, LA-LL) and 0.8 (right, LA08-LI01). c) and d): corresponding depth sections in direction N-S. e) and f): corresponding depth sections in direction E-W. Relative epicenter locations from Zeiß (2013). The urban area of the town Landau is shaded in light gray. The black circle in a) and b) marks the position of the geothermal powerplant Landau and black triangles in a)-f) the landing points of the production (upward) and reinjection (downward) wells.

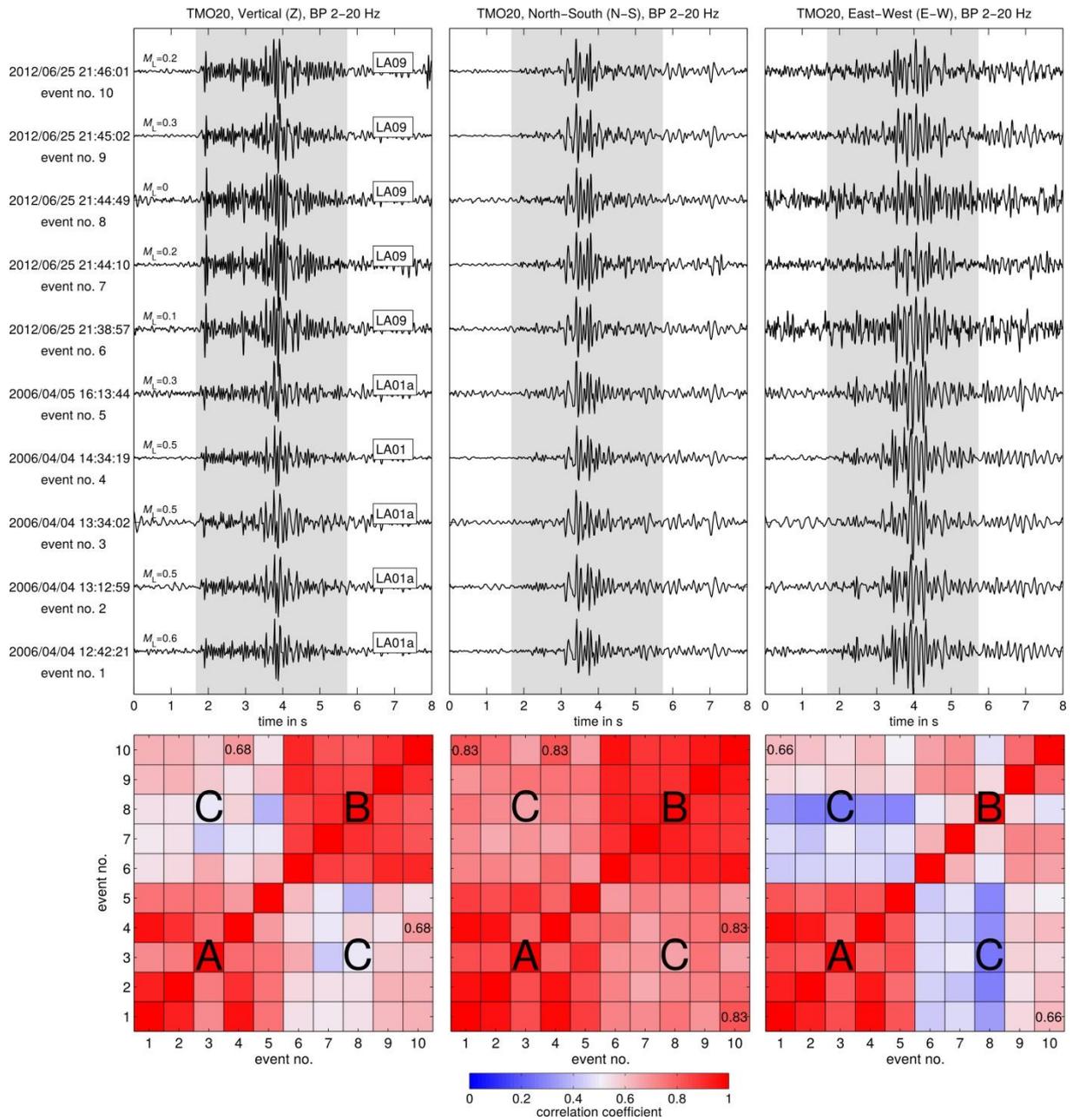


Figure 16: Amplitude-normalized waveforms and corresponding component matrices (Z, N-S, E-W) of selected Landau events at recording station TMO20 during the stimulation experiment 2006 (event no. 1-5, cluster LA01) and the event swarm on 25/06/2012 (event no. 6-10, cluster LA09). The gray shaded areas mark the 4 s waveform windows used for cross-correlation. The letters A-C in the component matrices indicate areas which are discussed in the text

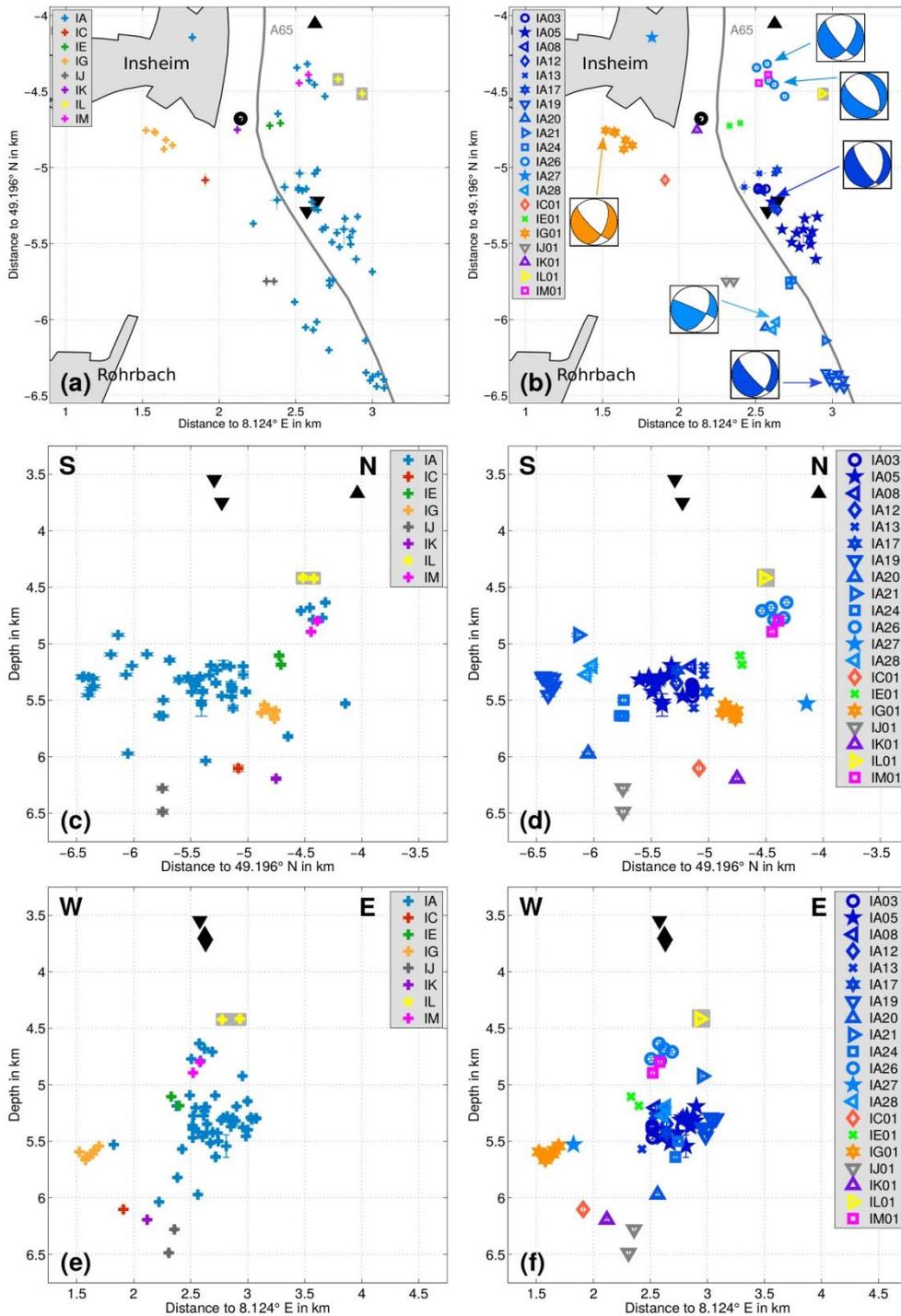


Figure 17: Spatial cluster distribution for the Insheim area. a) and b): relative epicenter distribution for the thresholds 0.7 (left, IA-IM) and 0.8 (right, IA03-IM01). c) and d): corresponding depth sections in direction N-S. e) and f): corresponding depth sections in direction E-W. Relative epicenter locations from Zeiß (2013). Settlements are shaded in light gray, the gray line shows the road A65. The black circle in a) and b) marks the position of the geothermal powerplant Insheim and black triangles in a)-f) the landing points of the production (upward) and reinjection (downward) wells.

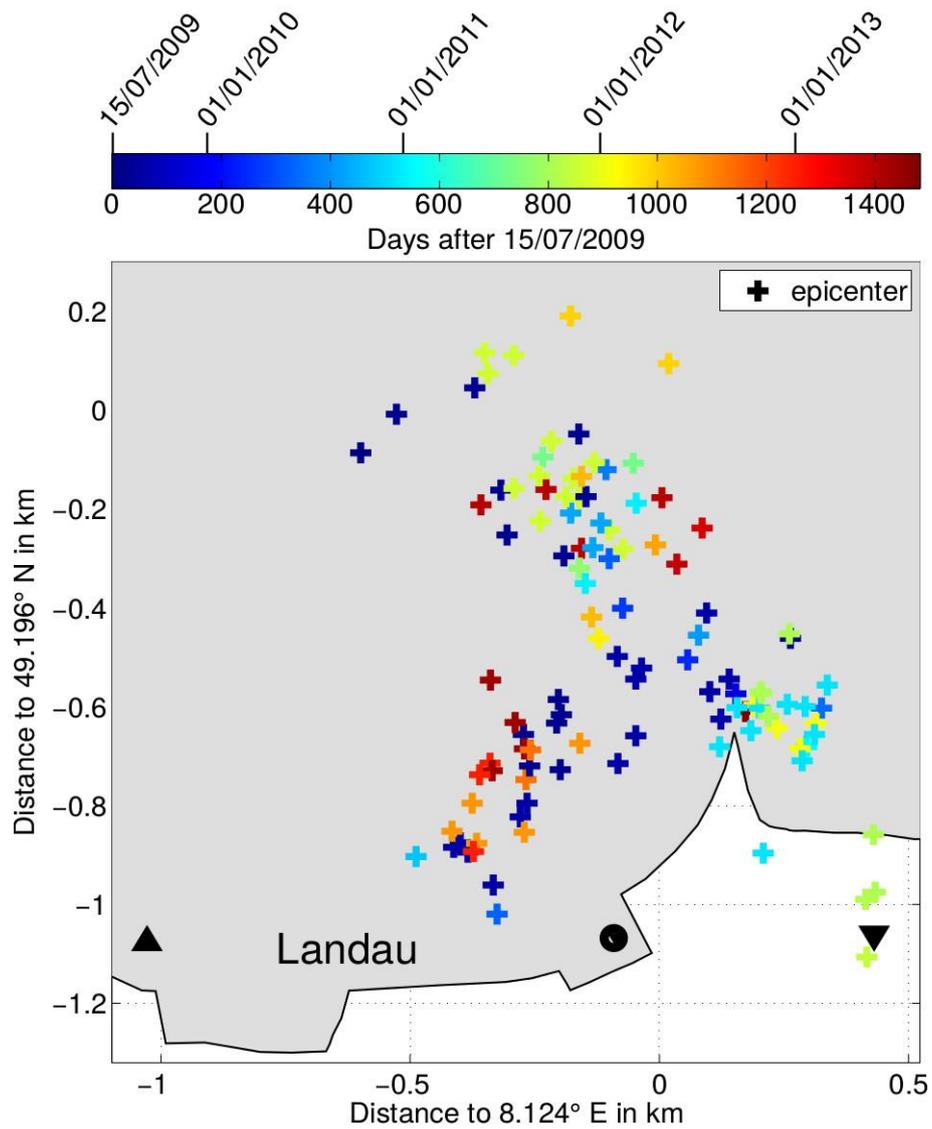


Figure 18: Temporal evolution of located events of the Landau area. The distribution corresponds to the event distribution shown in Fig. 15a for a threshold of 0.7. The urban area of the town Landau is shaded in light gray. The black circle marks the position of the geothermal powerplant Landau and black triangles the landing points of the production (upward) and reinjection (downward) wells.

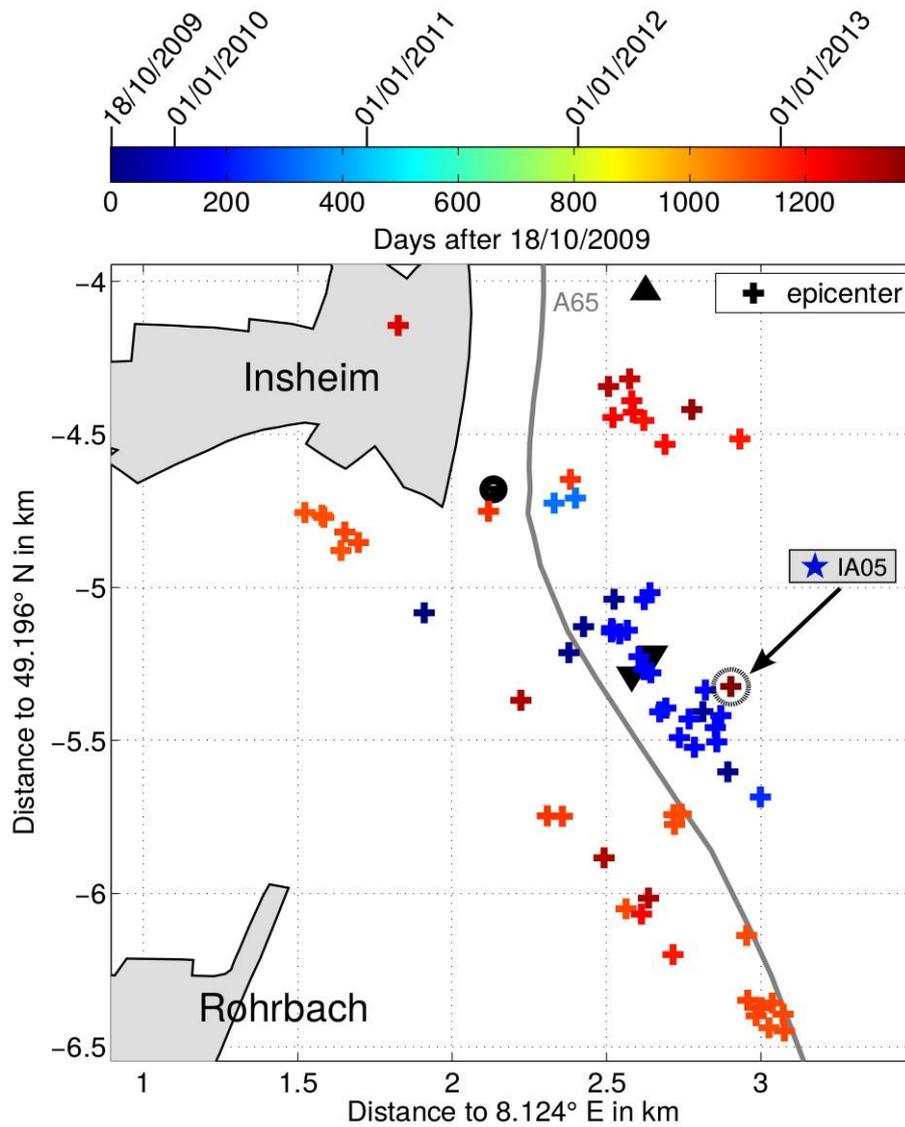


Figure 19: Temporal evolution of located events of the Insheim area. The distribution corresponds to the event distribution shown in Fig. 17a for a threshold of 0.7. The arrow marks the only located event ($M_L = 1.1$) of cluster IA05 which occurred after the initial operation of the powerplant on 06/08/2013. Other unlocated events related to IA05 that occurred after the initial operation are reported in Figs 14d and e. Settlements are shaded in light gray, the gray line shows the road A65. The black circle marks the position of the geothermal powerplant Insheim and black triangles the landing points of the production (upward) and reinjection (downward) wells.