A Consistent High-Resolution Catalog of Induced Seismicity in Basel Based on Matched Filter Detection and Tailored Post-Processing

Journal Article

Author(s):
Herrmann, Marcus ; Kraft, Toni ; Tormann, Thessa; Scarabello, Luca; Wiemer, Stefan

A Consistent High-Resolution Catalog of Induced Seismicity in Basel Based on Matched Filter Detection and Tailored Post-Processing

Marcus Herrmann[1], Toni Kraft[1], Thessa Tormann[1,2],
Luca Scarabello[1], and Stefan Wiemer[1]

[1] Swiss Seismological Service, ETH Zürich, Zürich, Switzerland
[2] Now at PartnerRe, Zürich, Switzerland

July 3, 2019

Key Points:

- We developed a sensitive single-station detection algorithm that deals robustly with quality deficits in the input data

- The 12-year long catalog features a consistent detection threshold and moment magnitude estimates spanning from $M_\mathrm{w}-1.5$ to 3.1

- We observed a nonlinear event size distribution during reservoir stimulation which leads to an overestimate of the seismic hazard

Abstract:

Seismic monitoring of the Basel Enhanced Geothermal System (EGS) has been running for more than a decade. Yet the details of the long-term behavior of its induced seismicity remained unexplored because a seismic event catalog with consistent detection sensitivity and magnitudes did not exist. This knowledge is essential for developing guidelines and mitigation procedures on how to safely operate and terminate injection activities. Only few observational data exist that cover all phases of such projects in a consistent manner. Here we describe a method that overcomes these deficiencies based on sensitive matched filter detection and a machine learning approach to remove false detections. With an emphasis on consistency, we create a catalog that contains more than $280\,000$ events down to $M_\mathrm{w}-1.5$. The much higher temporal resolution allows us to analyze induced microearthquakes in great detail and to gain new insights. We resolved temporal variations of seismicity parameters and, in the post-operational phase, a preferential temporal clustering of events. We find a breakdown in the Gutenberg–Richter scaling during reservoir stimulation, which may have physical reasons or could be caused by a method-independent detection limit during high event rates. The scaling breakdown has implications for the design of Adaptive Traffic Light Systems and may limit the potential of real-time mitigation strategies in future EGS projects. Nevertheless, our catalog gives the opportunity to study the temporal evolution of the sequence in unprecedented detail, which will help to better understand the physical processes in a geothermal reservoir, potentially not only in the Basel case.

Plain Language Summary:

Fluid injections into the deep underground, such as performed in geothermal projects, may cause earthquakes. These induced earthquakes provide important information about the involved physical processes, but can sometimes be stronger than acceptable and hinder a project to continue. Avoiding unacceptable earthquakes requires a better understanding of the immediate and long-term seismic response of the underground to such operations. However, inconsistencies of existing earthquake catalogs and their generally low resolution restrict our ability to understand these processes. Such catalog restrictions also affected the case of the Basel deep geothermal project. As a result, specifically the long-term behavior of its induced seismicity remained unexplored. To overcome these deficiencies, a consistent catalog with high resolution is needed. We reinvestigate the induced seismicity in Basel in detail over its whole life span (12 years). Using seismograms of known earthquakes, we search for similar earthquakes and detect an abundance of smaller ones that were previously unknown. To ensure catalog consistency, we further develop advanced techniques that provide robust magnitude estimates and maintain a high detection sensitivity. Like increasing the resolution of an image with a spyglass, the new catalog reveals previously unseen details of this particular sequence. In the injection period, for instance, we find deviations from the

2

expected behavior of earthquakes and their magnitude distribution. These findings make it necessary to rethink earthquake mitigation strategies in geotechnical projects.

1 Introduction

Deep geothermal energy offers an attractive energy resource due to its huge energy potential, but remains essentially unexploited. One of the many technical challenges to exploit deep geothermal energy, for example, with an enhanced geothermal system (EGS), is the high initial costs of exploration and drilling, as well as the control of induced seismicity. The Basel Deep Heat Mining project pursued the goal of creating one of the first commercial EGS plants in the world and providing electrical and thermal energy directly within the city of Basel (see Fig. 1). To create an EGS reservoir at depth, fluids are pumped under high pressure into the rock mass to cause hydroshearing, which eventually increases its permeability for subsequent water circulation. In Basel, this stimulation process started on 2 December 2006 with an injection of $11\,570\,\mathrm{m}^3$ of water from the near Rhine river over the course of 5 days at increasing flow rates [*Häring et al.* 2008]. The stimulation was accompanied by an increasing seismic activity, including a $M_\mathrm{L}2.6$ event, which prompted the operator to stop the injection on 8 December. Only a few hours later, a $M_\mathrm{L}3.4$ occurred, the largest event in the sequence. It was felt by the population up to $20\,\mathrm{km}$ away [*Edwards et al.* 2015], caused minor nonstructural damage within the city, lead to increased awareness of the public, and attracted international attention [e. g., *Kraft et al.* 2009; *Giardini* 2009]. In late 2009, a seismic risk assessment concluded an unacceptable risk for a continued geothermal operation [*Baisch et al.* 2009, SERIANEX risk study]. The public authorities suspended the project [*Giardini* 2009], and the well was closed in April 2011.

Although the Basel geothermal project failed, the collected data of the induced seismicity improved our understanding of EGS in several scientific studies that: investigated the reservoir structure and the orientation of fault planes [e. g., *Asanuma et al.* 2008; *Dyer et al.* 2008; 2010; *Deichmann et al.* 2014; *Kraft and Deichmann* 2014], performed statistical analyses for earthquake forecasting purposes [e. g., *Bachmann et al.* 2011; *Mena et al.* 2013; *Gischig and Wiemer* 2013; *Király-Proag et al.* 2018], modeled geomechanical properties [e. g., *Goertz-Allmann et al.* 2011; *Goertz-Allmann and Wiemer* 2013; *Bachmann et al.* 2012], studied the larger events in terms of their trigger mechanism [e. g., *Mukuhira et al.* 2013] and rupture propagation [e. g., *Folesky et al.* 2015], analyzed ground motion and macroseismic intensities [e. g., *Ripperger et al.* 2009; *Edwards et al.* 2015], and performed seismic risk analysis [e. g., *Baisch et al.* 2009; *Mignan et al.* 2015]. These studies benefited from the six-station borehole seismometer network (three stations used in this study are shown in Fig. 1) and a surface network with 30 stations in the area [*Deichmann and Ernst* 2009]. Previous work utilized an earthquake catalog with about 3500 locatable events [*Dyer et al.* 2008; 2010] but was limited to these earthquakes only.
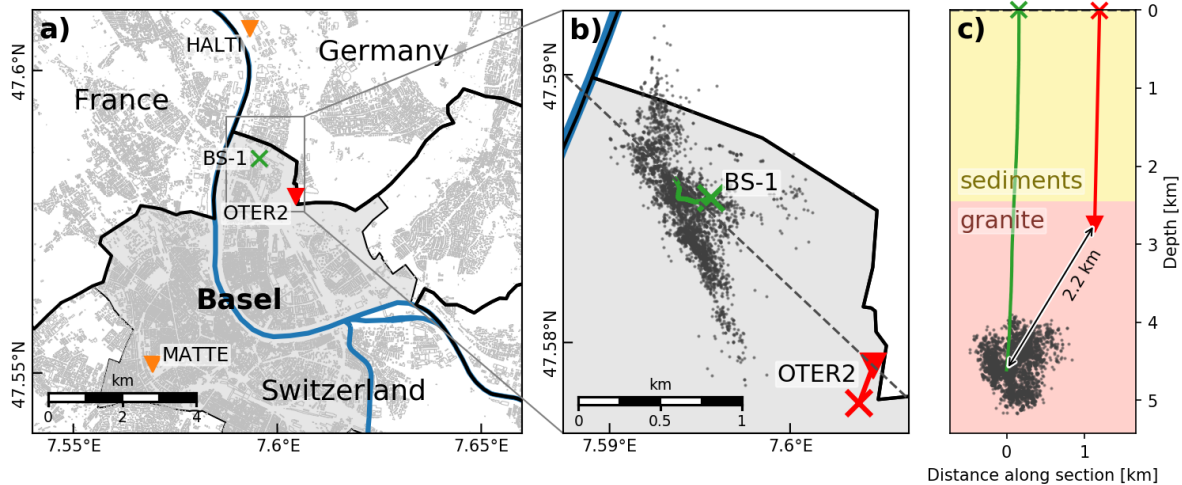
Figure 1: Overview of the study region with the injection well (BS-1, green) and the borehole seismometer used in this study for detecting events (OTER2, red triangle). (a) Map view with building locations and two more borehole stations used for additional analysis in this study (MATTE, HALTI; orange triangles). (b) Zoomed map view with earthquake locations (black dots) from the combined catalog (see text). (c) Depth section along the dashed line indicated in (b) with all earthquakes projected onto that plane. The transition from sedimentary rocks to the granite basement is at ~2.5 km depth [*Häring et al.* 2008; *Bethmann et al.* 2012]. Boundary, building, and river data originate from the OpenStreetMap project [`www.openstreetmap.org`].

The hydraulic stimulation caused an abundance of microearthquakes, which reflect small shear ruptures on pre-existing and newly generated fractures that generate new flow paths in the rock mass. These earthquakes contribute to creating the EGS reservoir, but are mostly too small to be detected on more than the closest seismic station (OTER2, see Fig. 1). The existing catalogs therefore reflect only a fraction of the actually detectable earthquakes. The seismicity covered by these catalogs decayed until mid-2010, whereupon no earthquake was reported by the operator of the borehole seismic network (Geothermal Explorers Ltd. [GEL]). In May 2012, the Swiss Seismological Service (SED) took over the responsibility of seismic monitoring and kept two borehole stations in operation (MATTE and OTER2, see Fig. 1). In the same month, a $M_L 1.2$ occurred [*Diehl et al.* 2013; *Deichmann et al.* 2014]. Since then, the SED registered an increase in seismicity, with magnitudes up to $M_L 1.9$ (October 2016) [*Diehl et al.* 2018]. Due to the change in seismic monitoring from 2012 onward, the currently existing catalogs do not cover the life span of the induced sequence consistently in terms of detection sensitivity, magnitudes, and locations. To provide a more consistent understanding of the evolution of the sequence, we aimed to generate a new catalog that covers the complete life span of the EGS reservoir (November 2006–2018) with a much higher detection sensitivity and homogeneous magnitude estimates.

The processes behind injection-induced seismicity are not yet understood well enough to make reliable forecasts of the likely seismic response of the underground to hydraulic stimulation. The poor understanding underlines the need for more data and better data quality. Recording

4

and analyzing the seismic response in much higher detail, in a consistent way, and over longer time periods may allow us to:

1. resolve seismogenic processes that were undetected before,

2. permit more detailed seismic and statistical analyses,

3. study the long-term behavior of these processes and induced seismicity in general,

4. provide timely detections of changes in the seismogenic behavior of geothermal reservoirs, and

5. improve seismicity forecasts and short-term hazard assessments.

Here we harness the potential of matched filter analysis and machine learning techniques to retrieve small previously missed microearthquakes that are hidden in the seismic records of the most sensitive seismometer station (OTER2). We explore the potential of our proposed procedure to improve our understanding of the governing processes behind injection-induced seismicity. We also discuss the advantages and pitfalls that may arise should such a procedure serve as an integral part in an Adaptive Traffic Light System (ATLS), which ultimately aims to reduce the induced seismic risk in future EGS projects [e. g., *Douglas and Aochi* 2014; *Gischig et al.* 2014; *Trutnevyte and Wiemer* 2017; *Mignan et al.* 2017; *Grigoli et al.* 2017; *Wiemer et al.* 2017b].

2 Data

The starting point of our matched filter analysis was a catalog of known earthquakes whose recorded waveforms were used as templates. We compiled such a catalog from several sources:

1. the catalog of the project operator GeoPower Basel AG based on the seismic network operated by GEL ("GEL"; *Dyer et al.* [2008, 2010]); borehole station network; 3664 events; covering November 2006 to June 2010; M_w0.15–3.00;

2. the catalog of N. Deichmann from the SED ("SED"; *Deichmann et al.* [2014]), which received regular updates (N. Deichmann, pers. comm., 2014–2018); surface station network; 226 events; covering November 2006 to December 2017; M_L0.5–3.4 (of N. Deichmann); M_w0.94–2.95 (of F. Bethmann, pers. comm., 2011); and

3. a catalog from a separate manual analysis performed in this study ("SED-μAn"); 31 events; covering May 2012 onward; M_L-0.04–0.82 ($\approx M_w$0.83–1.21)

The GEL catalog is based on detections of a six-station borehole network and contains only events that could be located [*Dyer et al.* 2008; 2010]. Events were initially detected with an amplitude-threshold-based trigger at station OTER2 (Fig. 1). During the injection period, 13 500 potential events were obtained in this manner [*Dyer et al.* 2008]. For 3555 events, moment magnitudes, M_w, were calculated by GEL in 2008 (T. Spillmann, pers. comm., 2015)

using the spectral method of *Abercrombie* [1995]. Additional 109 events without such a magnitude were detected after 2008 or were added in a later revision by *Dyer et al.* [2010]. For 979 events, the pick and source times in the GEL catalog had errors of up to a minute [*Kraft and Deichmann* 2014] due to a software bug in the operator's system. We assessed and corrected these timing errors with the procedure outlined in the Supporting Information (Text S1). During the review of the GEL events, we identified five duplicates and two events that were not originating from the reservoir; these were excluded from further analysis. *Kraft and Deichmann* [2014] performed a relocation analysis of this catalog, providing relative locations for 1982 events.

The SED catalog is based on detections of the national surface network of the SED and the borehole station OTER1, which is located about 2.2 km above OTER2. Until 2012, the SED catalog has 196 events in common with the GEL catalog; the remaining 30 events occurred from 2012 onward. Locations are based on master-event relocation provided by N. Deichmann (pers. comm., 2018). Moment magnitudes were obtained by F. Bethmann (pers. comm., 2011) as presented in *Bethmann et al.* [2011] for 195 events (until end of 2007); for the remaining 31 events, we converted local magnitudes, M_L—as obtained by N. Deichmann (pers. comm., 2018) using the median value of four station magnitudes [*Deichmann and Ernst* 2009; *Deichmann et al.* 2014]—into M_w using the scaling relation of *Bethmann et al.* [2011] ($M_w = 0.633 \cdot M_L + 0.766$).

The SED-μAn catalog contains additional events that were not detected automatically by the SED network, but that could be located manually. They were found by an early template matching analysis using only four templates. The absolute locations of these events have larger uncertainties than usually expected for the SED catalog due to uncertainties in the arrival times.

To compile a combined catalog of event waveforms, we appended to the GEL catalog all SED and SED-μAn events that occurred after 2010. For events before 2012, we preferred the locations of *Kraft and Deichmann* [2014], if available; otherwise, we took the original GEL locations [*Dyer et al.* 2008; 2010]. For events from 2012 onward, we could only use the locations of the SED catalog (*Deichmann et al.* [2014] and N. Deichmann, pers. comm., 2018); locations of the SED-μAn catalog were ignored. Because the magnitudes of the catalogs were inconsistent, we revised the magnitudes in our study (see section *Magnitude Estimation*). In total, the combined catalog (provided in the Supporting Information) contains 3723 events with associated waveforms at OTER2.

The deepest borehole station, OTER2 (a 4.5 Hz velocity sensor), is located at a depth of 2.74 km with a distance of 2.24 km from the injection point (Fig. 1). As OTER2 is situated in the crystalline basement (Fig. 1), it features a low attenuation of high-frequency signals (see Fig. 3a and *Bethmann et al.* [2012]). The 12-year-long recordings contained several data gaps lasting up to a month. These gaps were complemented with waveform data from another—yet less sensitive—borehole station, MATTE (Fig. 1), located at 0.55 km depth with a distance of 5.94 km from the injection point. Both stations were recording at a sampling rate of 1000 Hz from November 2006 until May 2012. In May 2012, when the SED took over the seismic monitoring and replaced the recording hardware, the sampling rate was changed to
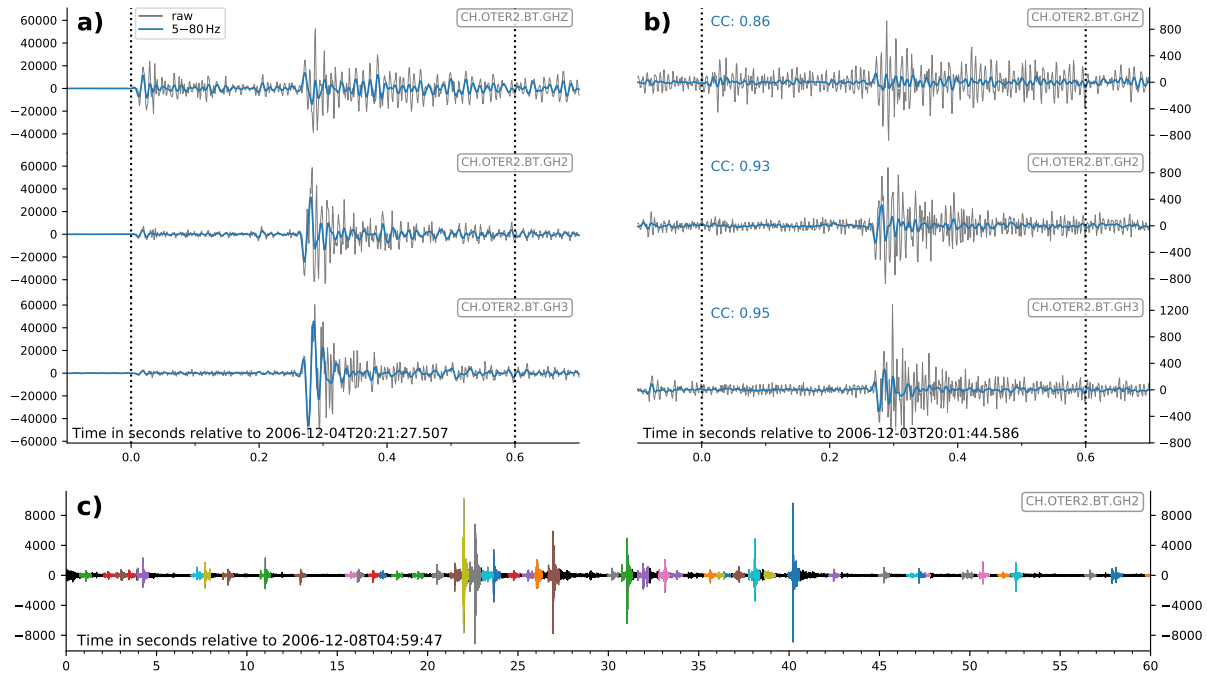
Figure 2: Example 1000 Hz waveforms (black, unfiltered, units in counts) recorded at the borehole station OTER2 and the 5–80 Hz band-passed waveforms (blue) as used for matched filter analysis. (a) Template waveform ($M_w 0.9$); the vertical dotted lines indicate the 0.6 s duration used for matched filter analysis. (b) Waveform of a detected event ($M_{wx} -0.6$) using the template waveform in (a). The cross-correlation (CC) similarity is indicated in the figure for each component. (c) One-minute time window during the highest rate of fluid injection (early morning of 8 December 2006) illustrating the high event rate; in this time window, 45 templates detected 50 events (highlighted by 10 different colors) ranging from $M_{wx} -0.88$ to 0.05. Smaller-amplitude signals that were not detected were either reservoir-unrelated (transient noise), or reservoir-related but below the magnitude of completeness, M_c, at that time ($\sim M_{wx} -0.5$, see Fig. 9c). (M_{wx} is our moment magnitude estimate for detected events.)

500 Hz [*Diehl et al.* 2013]. Unfortunately, the new digitizer could not be isolated well from electromagnetic noise at 50 and 150 Hz, which coupled stronger into the system and increased the noise level by a factor of ~2 compared to conditions before May 2012. The data availability over the whole study period (November 2006–2018) amounts to 96.5 % for OTER2, 98.4 % for MATTE, and 99.3 % for both combined.

The vertical component of OTER2 failed in June 2010, which is particularly unfortunate as it contains most of the *P*-wave energy of the reservoir events (channel GHZ; see Fig. 2a & b). The *S*-wave energy, instead, is visible on all components and very impulsive on the horizontal ones. We accepted that deficiency and did not switch to an alternative station because OTER2 has a much higher signal-to-noise ratio than all other stations. As our goal was to create a homogeneous and consistent catalog, we only used the two horizontal components for the analysis of the whole study period.
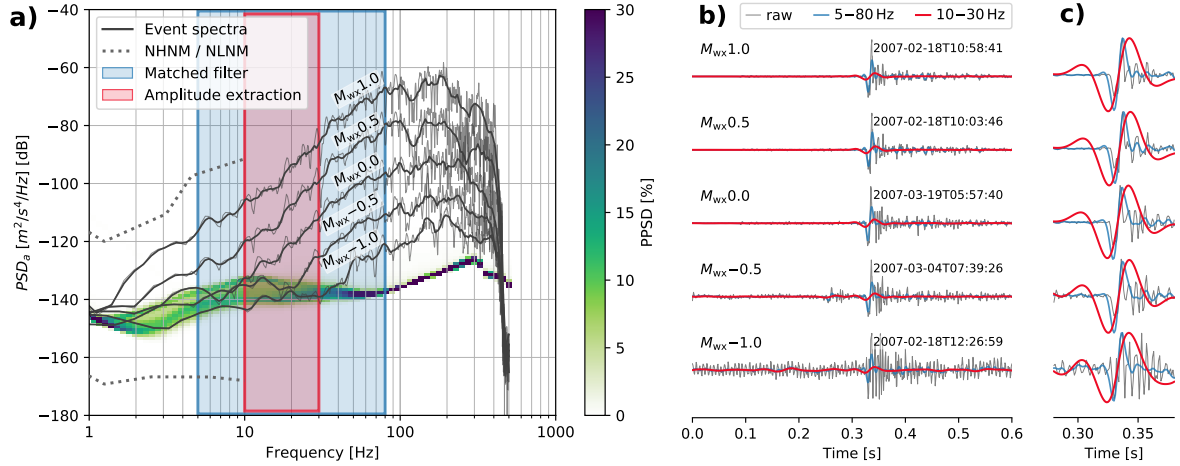
Figure 3: Event spectra and waveforms for five detected earthquakes of distinct magnitude: $M_{wx}-1.0$, -0.5, 0.0, 0.5, and 1.0, which all belong to the same template family. (a) Acceleration power spectral densities (PSD_a) of the S-phase spectra overlayed on the probabilistic power spectral density (PPSD, similar to *McNamara and Buland* [2004]) at station OTER2 (GH2 channel) for the period 2006–2012. For reference, the two dotted lines indicate the *Peterson* [1993] new high- and low-noise model (NHNM, NLNM). Two frequency bands indicate the passbands for matched filter detection (blue) and amplitude extraction (red). (b) Waveforms of the five events filtered in the two frequency bands with the colors corresponding to (a); the black waveforms are unfiltered. The blue and red waveforms were filtered in the frequency domain, thus having zero-phase lag (non-causal). (c) A zoom (0.1 s) into the S-phase with normalized waveform amplitudes.

3 Methods

3.1 Matched Filter Detection

Matched filter detection, or template matching, is an earthquake detection method based on waveform cross correlation, which proved to lower the detection threshold by about one magnitude unit compared to conventional methods (e. g., like the energy-based short-term average/long-term average technique) and can often recover ≳10 times more events [*Gibbons and Ringdal* 2006; *Peng and Zhao* 2009; *Schaff and Waldhauser* 2010]. Matched-filter-based detection has previously been applied in various seismological contexts to improve existing seismic catalogs, for example, of deep geothermal projects [e. g., *Plenkers et al.* 2013; *Vasterling et al.* 2017], hydraulic fracturing operations [e. g., *Holland* 2013; *Caffagni et al.* 2016], mining-induced seismicity [e. g., *Chambers et al.* 2015], waste-water injections [e. g., *Goebel et al.* 2016], earthquake swarms [e. g., *Shelly et al.* 2016], aftershock sequences [e. g., *Peng and Zhao* 2009; *Schaff and Waldhauser* 2010], nonvolcanic tremors [e. g., *Shelly et al.* 2007], and for nuclear explosion monitoring [e. g., *Carmichael* 2016]. Despite of recent advances in seismic event detection [e. g., *Carmichael* 2016; *Hammer et al.* 2012; *Perol et al.* 2018], we adhere to the more basic approach of matched filter detection, which we consider the better choice for the Basel case: Our multi-template matched filter detector can be made very

8

sensitive to the already known 3723 reservoir earthquakes, which densely delineate the extent of the geothermal reservoir [*Asanuma et al.* 2008; *Dyer et al.* 2010; *Kraft and Deichmann* 2014].

Matched filter detection has the potential to detect events with signal amplitudes even below the noise level [*Gibbons and Ringdal* 2006; *Schaff* 2008] and at times of high event rates when waveforms overlap [*Peng and Zhao* 2009; *Schaff and Waldhauser* 2010]. The method takes advantage of the high waveform similarity observed within seismic sequences [*Geller and Mueller* 1980]: seismic waves that originate from a similar source region caused by a similar source mechanism travel a similar path in the medium and will have similar waveform shapes at a receiver. Hence, with a waveform of a known event one can search for more events that have a similar, or the same, source. To find them, a matched filter continuously measures the similarity between the waveform of a known event and the available continuous data. It provides a cross-correlation (CC) trace that spans the extent of the available data (for details, see Text S2). A peak in the CC trace that exceeds a defined threshold triggers a new detection. A very high CC value indicates a repeating source, but the value degrades (i. e. the waveforms decorrelate) with growing separation distance [*Menke* 1999; *Baisch et al.* 2008; *Castellanos and* van der *Baan* 2015; *Hakso and Zoback* 2017], increasing noise level [*Gibbons and Ringdal* 2006; *Schaff* 2008; *Carmichael* 2016], increasing frequency bandwidth of the signal [*Baisch et al.* 2008], differing magnitude, and due to deviations in the focal mechanism and source-time function [*Harris* 2006; *Schaff and Waldhauser* 2010].

Configuration

We performed a single-station detection procedure at OTER2. The common frequency range of events above the noise level of OTER2 is in the range of 20–400 Hz (see Fig. 3a). But the signals need to be band-limited to a range where the similarity between a template and the unknown events is high, that is, below the corner frequencies of their source spectra [*Harris* 2006; *Schaff and Waldhauser* 2010]. On the one hand, the bandwidth should be large to reduce false detections [*Arrowsmith and Eisner* 2006; *Harris* 2006]. On the other hand, an increasing upper band limit degrades the waveform similarity of closely spaced hypocenters [*Baisch et al.* 2008], which makes a template waveform less tolerant to small variations at the source. We assessed the influence of different passbands on the number of detected reservoir events (during 1 day of reservoir stimulation, on 3 December) and false detections (during 1 day prior to the stimulation, on 1 December) and found a 5–80 Hz filter (4th-order Butterworth, blue frequency band in Fig. 3a, see also Fig. S1) to maximize the sensitivity to (small) reservoir events and resulting in no false detections. The 1000 Hz waveform data were resampled to the highest common sampling rate used over the whole study period (500 Hz).

Due to the absence of the discriminative vertical component, we implemented a strict trigger condition: The CC threshold had to be exceeded at both components at the same time. The minimum of both CC values was then taken as the similarity measure of a detection to its template.

9

A statistical threshold based on the 'median absolute deviation' ['MAD', e. g., *Shelly et al.* 2007] may generally lessen the variation of the detection limit, but not in our case: During the injection, the high event rate caused closely spaced maxima and, therefore, an elevated CC trace; as a result, the MAD threshold is more than twice as high (compared to before the injection), which would have impaired the event detection. To avoid this added inconsistency, we considered a fixed CC threshold the better choice.

To complement potentially missing events at times when OTER2 experienced data outages, we performed a separate matched filter analysis at station MATTE, however, with a less sophisticated setup and post-processing than outlined above for OTER2 (for details, see Text S3).

Algorithmic Implementation

We developed a Python-based framework which can perform the workflow presented in this study (e. g., template selection, waveform preprocessing, matched filter detection, post-processing, and result plotting). It was particularly designed for the single-station approach, because our interest was to use only the most sensitive station for earthquake detection. For waveform management and processing, we made use of `obspy` [*Beyreuther et al.* 2010] but replaced some of its routines for performance considerations (e. g., filtering and resampling in the frequency domain, see Text S4).

For performing event detection with thousands of templates over ~12 years of data in manageable time, we parallelized our matched filter routine to enable high-performance computing involving hundreds of processors. We made use of MPI [*Message Passing Interface Forum* 1997] and implemented a master/slave configuration, where one master process controls as many slave processes as processors are available. Each slave process either (1) reads waveform data from an archive and preprocesses it or (2) performs a match filter detection on preprocessed waveform data with one template. To avoid idle slave processes, the master process optimizes resource sharing and balances the computational load. The performance scales with the number of processors because the processing of the chunk–template pairs are independent from each other. See Text S5 and *Lewis* [1995] for further performance considerations concerning the algorithmic implementation of the matched filter operation.

Preprocessing the 50 Hz Noise Signal

The 50 Hz noise signal intensified after the digitizer replacement in 2012 and caused a generally higher noise level and a reduced detection threshold. Band-stopping this noise with a time domain filter would have removed too much frequency information from the signal and also would have altered the waveforms too much (e. g., "ringing" artifacts; see Fig. S2). Instead, we applied a narrow notch filter in the frequency domain (see Text S4 and Fig. S1) that cancels out only the undesired mono-frequent 50 Hz noise in the time domain signal.

3.2 Template Selection

The strength of template matching is simultaneously also its disadvantage: the selective sensitivity to the template waveform. In our case, we needed a large number of templates to adapt to the diverse set of event waveforms that were recorded at OTER2. One reason for this diversity was the proximity of OTER2 to the reservoir, recording event waveforms that were only weakly attenuated and varied in inclination angles (see Fig. 1c). Another reason was the complexity of the reactivated fault system in the Basel reservoir, which was characterized by fault planes of various orientations [*Deichmann et al.* 2014]. Many of the identified fault segments deviate significantly from the overall orientation of the seismic cloud, causing a substantial variation of the seismic radiation patterns observed at OTER2.

To create a template set from all known reservoir events, while excluding event waveforms that are similar to each other, we performed a waveform cluster analysis [e. g., *Aster and Scott* 1993; *Maurer and Deichmann* 1995]. As input, we used 0.6 s-long waveforms aligned by their *P*-wave onsets and applied the same 5–80 Hz band-pass filter used for matched filter detection. The short duration of 0.6 s preserved enough *S*-wave coda energy (see Fig. 2a), while it minimized the number of events that had to be excluded due to overlapping other events. Yet we had to exclude 257 event waveforms with bad signal quality (i. e. low signal-to-noise ratio or several events happening and overlapping within the 0.6 s) and 19 events with $M_\mathrm{w} > 2.0$ (i. e. larger events, which do not share much waveform similarity with the small events that we wanted to detect; *Deichmann* [2017]).

As a pair-wise similarity measure, we took the smaller of the two CC values measured at the sample, n_x, where the summed CC traces of the two horizontal components have their maximum:

$$S_{jk} = \min_{i=\{1,2\}} \left\{ CC^i_{jk}(n_x) \right\} \qquad \text{with } n_x = \arg\max_n \left\{ \sum_{i=1}^2 CC^i_{jk}(n) \right\}. \tag{1}$$

In equation (1), i is the channel index and $CC^i_{jk}(n)$ the running CC coefficient between two events j, k. (Taking the smaller of both CC values is consistent with the similarity measure used for matched filter detection.) Instead of taking the maximum or the average CC value at n_x, our choice makes the cluster forming more sensitive to the dissimilar component between two events and better separates truly dissimilar waveforms. To organize the similarity matrix S_{jk} into clusters, we employed agglomerative hierarchical clustering with the average linkage method [*Sokal and Michener* 1958]. The cluster cutoff was set at a rather strict similarity threshold of 0.9—a value typically used to group events that originate from the same source region and share the same source mechanism, that is, multiplets [*Arrowsmith and Eisner* 2006; *Deichmann et al.* 2014; *Castellanos and* van der *Baan* 2015; *Hakso and Zoback* 2017]. This threshold guaranteed an adequate coverage of the seismic cloud with a high detection sensitivity. The clustering led to 2274 clusters, of which 620 had more than one member (i. e. non-singleton clusters).

However, the impulsive *S*-phases of the events were dominating the cluster-forming, resulting in incorrect cluster associations with varying *S*-to-*P* traveltime differences of the cluster members
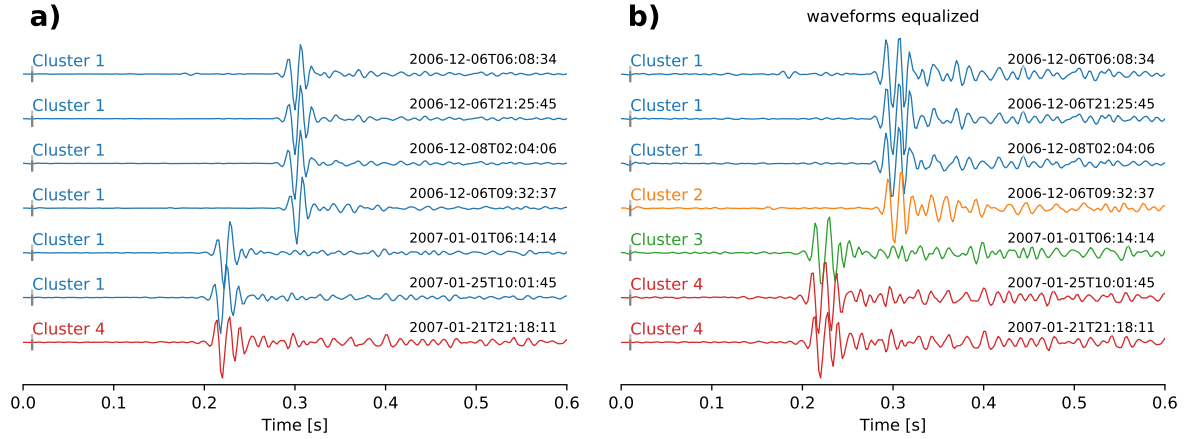
Figure 4: Example illustrating the effect of 'waveform equalizing' on the clustering of waveforms (clustering threshold: 0.9) for six example waveforms. Waveforms are aligned by their barely visible *P*-onset (indicated by vertical gray dashes) and were processed with a 4th-order, 5–80 Hz Butterworth band-pass filter. (a) Using original waveforms yields two clusters (colored blue and red). (b) Applying waveform equalizing on the same waveforms yields four redefined clusters (colors: blue, orange, green, and red) with better cluster-internal agreement of the overall waveform shapes.

(see Fig. 4a). To counteract this effect, we employed a processing step to scale down the amplitude of the *S*-phase relatively to the whole waveform. We call this procedure 'waveform equalizing' as it redistributes the waveform's energy more equally in the whole time window (see Fig. 4b). A waveform is equalized by multiplying the following window:

$$\mathbf{e} = 1 - \min\left\{6\,\frac{\text{Kurt}(\mathbf{x})}{N}, 0.75\right\}\frac{|\mathbf{x}_a|}{\max|\mathbf{x}_a|}, \tag{2}$$

in which $|\mathbf{x}_a|$ is the envelop of $\mathbf{x}$, $\text{Kurt}(\mathbf{x})$ is the kurtosis of $\mathbf{x}$, a measure of extreme values (outliers), and N is the sample length of the waveform. The window $\mathbf{e}$ can be interpreted as flipped normalized envelop of the signal scaled proportionally to $\text{Kurt}(\mathbf{x})$ but not more than 75 % of the *S*-wave amplitude to never notch away the *S*-phase completely. The kurtosis is multiplied by 6 and normalized by N to obtain a measure of extremity of the *S*-phase that ranged between 0 and slightly over 1 in our case.

Waveform equalizing reduced the weight of the *S*-phase in the similarity analysis and produced more consistent clusters (see Fig. 4b). The intrinsic waveform characteristics are captured much better after waveform equalizing because clustering is made more sensitive to the whole waveform shape and not only to the impulsive S-phase. A sole increase of the clustering threshold (e. g., to 0.95) without waveform equalizing would have resulted in more cluster splitting only. A total of 2508 clusters was produced; 560 were non-singleton and the largest had 16 members. Each cluster contributed one event to the template set; for non-singleton clusters, the member with the highest average similarity to all other cluster members was selected as template. Most of the template events were associated with a location, and their spatial distribution depicts the extension of the seismic cloud (see Fig. 5).

12
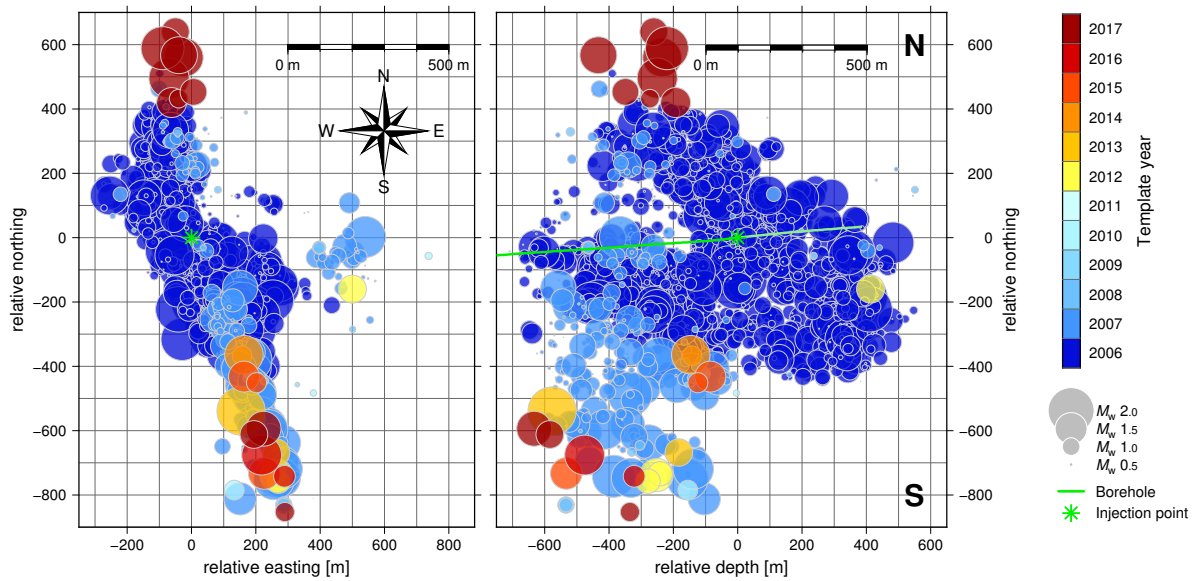
Figure 5: Locations of template events in map view (left) and in a North–South depth section (right) colored by year of occurrence. Symbol size is proportional to their magnitude (see legend). Locations until 2012 were provided by *Dyer et al.* [2008, 2010] and replaced with locations of *Kraft and Deichmann* [2014], if available. Locations for ≥2012 are based on the SED catalog (*Deichmann et al.* [2014] and N. Deichmann, pers. comm., 2018). Locations have different uncertainties depending on their origin. Template events from the SED-μAn catalog are not shown due to their high uncertainties. No template event occurred in the year 2011.

3.3 Post-Processing (of Detections)

After combining the detections of multiple templates, our detection catalog may refer multiple times to the same events. Therefore, we considered detections that were separated by less than 0.2 s as such duplicates and only kept the detection with the highest CC value (i. e. of the best-matching template) in that time interval. The separation tolerance of 0.2 s was chosen based on the observation that the time of the *S*-wave peak varied by 0.15 s among all templates relative to the start of the templates. As this impulsive *S*-wave is the main detection characteristic, the chosen value should guarantee the merging of duplicates while still allowing to detect overlapping events with a time difference of ≥ 0.2 s.

Removing False Detections with a Machine Learning Approach

Due to the impulsive character of the *S*-waves and the lack of P-wave energy in the template waveforms, we detected not only events originating from the Basel reservoir but also many unrelated signals that were impulsive as well, such as transient noise pulses, data artifacts (e. g., jumps, spikes, or gaps), or the *P*- and *S*-phases of non-reservoir events. Such false detections had rather high CC values to our templates (see Fig. 6b & c) because the impulsive signals were mainly transformed to the impulse response of the applied filter.
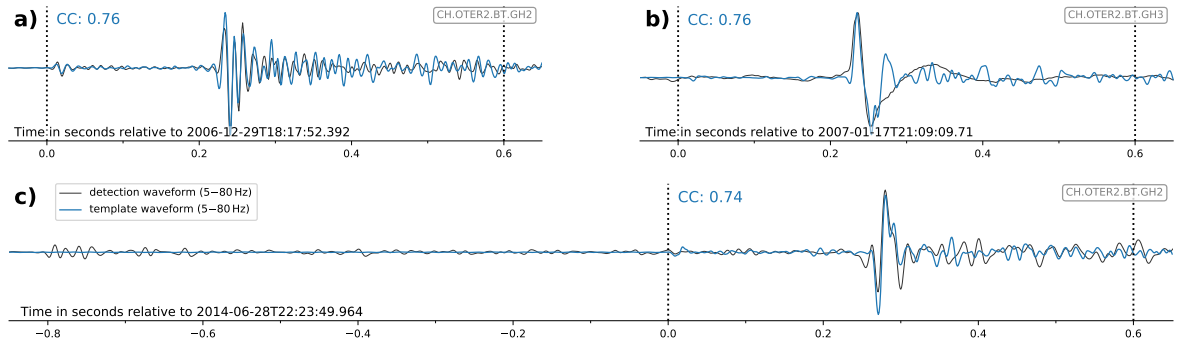
Figure 6: Filtered waveforms (5–80 Hz with 50 Hz noise removal) of example detections (black) as found by the overlayed template waveform (blue). All detections have CC-based similarities (CC, indicated in each subfigure) of around 0.75 for the 0.6 s-long template waveform (within the dashed lines), but are of completely different nature: (a) reservoir-related event; (b) impulsive noise; (c) distant earthquake with its *P*-phase about 0.8 s before the detection time.

Setting a fairly high similarity threshold to minimize the false detection rate would have excluded many reservoir events with lower similarities to our templates. Such low-similarity detections could either have a low signal-to-noise ratio [*Gibbons and Ringdal* 2006] or represent events that happen on fault patches for which no suitable templates existed in the combined catalog. A high similarity threshold would have therefore caused a higher incompleteness in our detection catalog. Alternatively, a multi-station detection approach could have reduced the false detection rate [*Slinkard et al.* 2014; *Gibbons and Ringdal* 2006], yet, with the cost of considerably lowering the overall detection sensitivity.

The discrepancy illustrated in Fig. 6 (i. e. that also false detections have high CC values) implies that the CC-based similarity measure in a limited frequency band, which we used for template matching, is not a good metric for characterizing the nature of the detections. Specifically, too much high-frequency information is filtered out for a robust false detection removal. The situation is further complicated by the fact that we can only use two of three components.

To overcome these difficulties, we initially set a low-similarity threshold of 0.45 and then specifically targeted the removal of false detections above that limit with a signal classification scheme. The threshold of 0.45 was found by visual inspection of the detected waveforms; below that value, a disproportionately high number of detections had low signal-to-noise ratios and we could not assess if they were reservoir-related. To classify the signals and separate false detections from the desired reservoir events, we employed a machine learning approach driven by waveform features (see Fig. 7). Waveform features are characteristic functions that extract specific information from the complete seismic signal [*Falsaperla et al.* 1996; *Beyreuther and Wassermann* 2008; *Hammer et al.* 2012], for example, by exploiting their spectral content. Machine learning techniques for the automatic classification of predetected events in seismic recordings have been applied numerously before [e. g., *Falsaperla et al.* 1996; *Del Pezzo et al.* 2003; *Masotti et al.* 2006; *Giacco et al.* 2009; *Mousavi et al.* 2016; *Provost et al.* 2017]. These approaches consist of a training stage, where a model is fit to a manually labeled training set,

14

and a classification stage, where the model estimates the classes of unknown detections based on what it inferred from the training set.

For creating the training set, we visually inspected detected waveforms and classified them into four categories: reservoir earthquakes ('*reservoir*'), transient noise pulses and data artifacts ('*noise*'), reservoir events detected by the wrong phase arrival, that is, S detects P or S detects a coda wave ('*time-err*'), and reservoir-unrelated earthquakes ('*outside*'). A total of 4006 waveforms were classified from different time periods, with about one half reservoir (2070) and one half non-reservoir (1936) detections. For feature extraction, we selected a subset of the features collected by *Beyreuther and Wassermann* [2008] and *Hammer et al.* [2012], all of which are implemented in the Python framework obspy [*Beyreuther et al.* 2010]: normalized envelop, centroid time, signal bandwidth, central frequency, dominant frequency, instantaneous bandwidth, instantaneous frequency, three cepstrum coefficients, and eight half-octave bands (Fig. 7; for details, see Text S6). Additionally, we incorporated the maximum signal amplitude as a feature. To account for their time dependence, each feature was computed at five points in time ('states'), with overlapping time windows of 256 samples (0.512 s). The states were chosen to cover different parts of the detection waveform: with focus on pre-S time window (t_1, Fig. 7), S-phase (t_3), coda (t_5), and two intermediate states t_2 and t_4 (see Fig. S3). In total, the time windows covered 1.4 s of data of each detection, including some data outside the 0.6 s template duration. To accentuate the relative changes between the states $t_1 - t_5$ of a feature, we added the differences between each consecutive state as additional features. In total, each detection waveform was represented by a 342-element feature vector (19 waveform features $\times$ [5 states + 4 differential states] $\times$ 2 channels). To assure that the feature vectors of the training set are in a comparable value range, they were jointly standardized to have zero mean and unit standard deviation [*James et al.* 2013].

We trained five established machine learning models for our multi-class discrimination task (see Fig. 7): *Logistic Regression* [*Cox* 1958], *Multilayer Perceptron* [*Hornik et al.* 1989], *Support Vector Machine* [*Vapnik* 1995], *Random Forest* [*Breiman* 2001], and *Gradient Tree Boosting* [*Friedman* 2001]. In Text S6, we present a short description of each algorithm and their use in literature for classifying seismic events. All of them are implemented in the Python framework *scikit-learn* [*Pedregosa et al.* 2011] and, after training them without further tuning, showed a classification performance of > 97 % accuracy.

To maximize the classification performance, we employed four training steps (see Fig. 7 and a detailed description of each step in Text S6). In each step, the best settings were found by evaluating the models in terms of their prediction performance with a 10-fold cross-validation: The training data are randomly split into 10 subsets of equal size, and the evaluation procedure—using nine splits for training and one split for testing—is performed for each fold and then averaged [*James et al.* 2013]. These splits were created with stratified sampling, that is, each fold approximately preserves the relative class frequencies of the whole training set. To optimize the classification performance for the 'reservoir' class, we quantified the prediction error in terms of the precision and recall score [*James et al.* 2013] of the 'reservoir' label and ignored the performances for the other classes (true negatives). The precision score is the fraction of correctly predicted 'reservoir' labels (true positives) to all 'reservoir' predictions
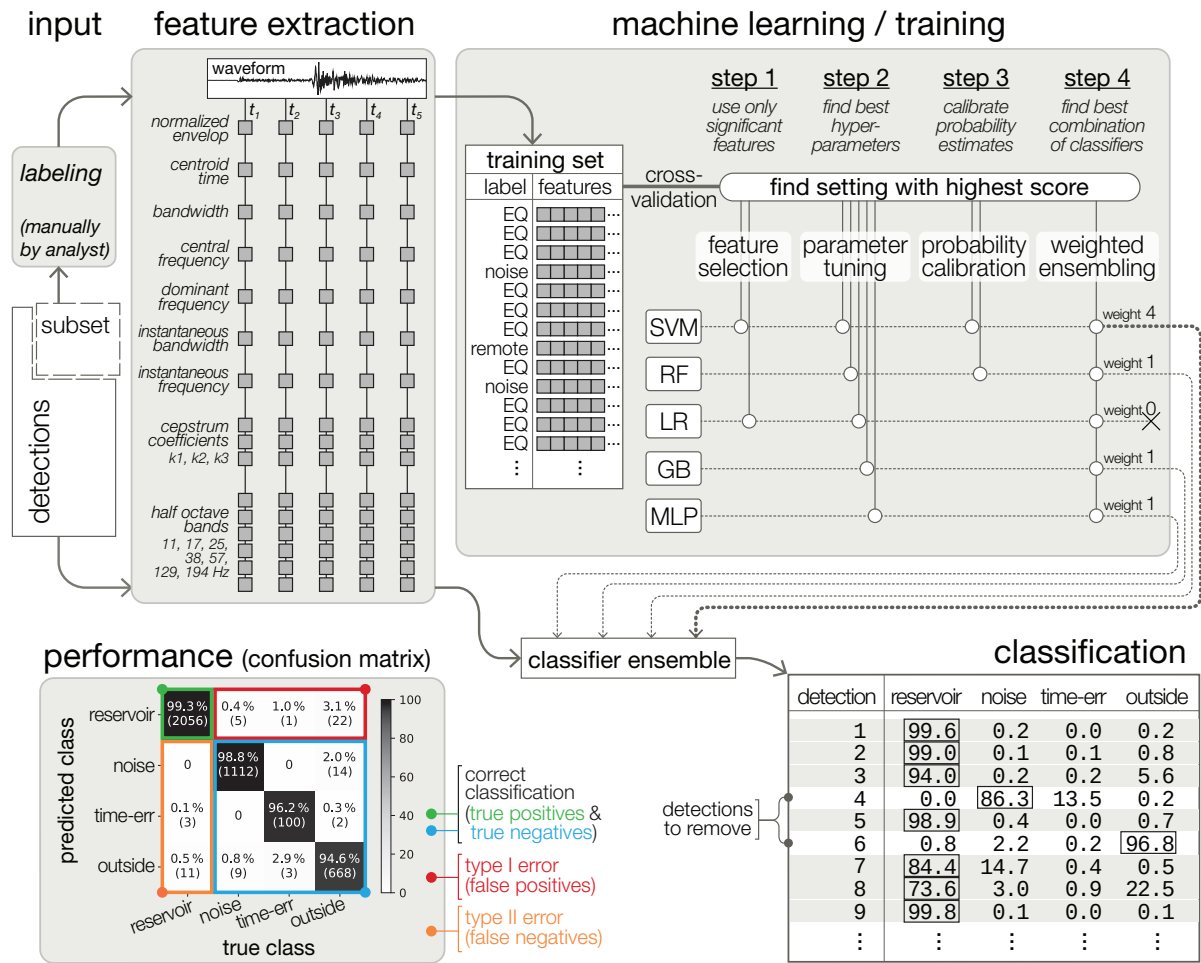
Figure 7: Flow-chart of our machine learning approach to classify detections. The classification of detections was used to remove false (non-reservoir) detections. *'input'*: A subset of the detections (4006) were manually classified to create a labeled training set. *'feature extraction'*: Many different waveform features were generated for the training set and the remaining detections at five defined points in time, t_{1-5} (small gray squares). *'machine learning / training'*: During the four processing steps (see Text S6 for detailed descriptions), five classification models were trained and optimized using different machine learning techniques: Support Vector Machine (SVM), Random Forest (RF), Logistic Regression (LR), Gradient Tree Boosting (GB), and Multilayer Perceptron (MLP). In the last step, the models were combined with different weights to a more robust classifier ensemble. *'classification'*: Once trained, the classifier ensemble estimated class-related probabilities for all detections based on their associated feature values. The detections were assigned to the class with the highest probability. *'performance'*: The confusion matrix visualizes the classification accuracy of our classifier ensemble when re-evaluating the training set via cross-validation (see Text S6); it compares the predicted class against the true class, revealing correctly classified (diagonal elements) and misclassified training detections (non-diagonal elements) for each class. Correct predictions of reservoir detections are considered as true positives (green) and correct predictions of all non-reservoir events (e. g., 'outside' predicted as 'noise') as true negatives (blue). False positives (type-I error) are in the upper right row (red frame); false negatives (type-II error) are in the lower left column (orange frame). The upper value in each cell is normalized by the class element size; the lower value is the absolute amount.

16

(accounting for false positives), while the recall score is the fraction of true positives to all 'reservoir' events (accounting for false negatives). Both scores were combined into a single measure using their harmonic average, the F1 score, which guided the model evaluation in the different training steps (see Text S6). The training steps ultimately led to a combination of all classifier models into a classifier ensemble, which improved the robustness over a single model and compensated their individual deficiencies by adding diversity [*Polikar* 2006; *Kuncheva* 2014].

The confusion matrix (also known as contingency table) in Fig. 7 (description in caption) reports the classification performance of the classifier ensemble with respect to the individual classes. The evaluation was based on cross-validated probability estimates obtained for each training element (out-of-fold estimates, see Text S6), which were averaged over ten random 10-fold cross-validations. Regarding the 'reservoir' class, the classifier ensemble produced more false positives (type-I error, red in Fig. 7) than false negatives (type-II error, orange in Fig. 7). Hence, the recall score was slightly higher (0.993) than the precision score (0.987).

Once trained on the full training set, the classifier ensemble learned the meaning of the waveform features to discriminate classes. The ensemble was then applied to the extracted feature vectors of all our detections and estimated their class memberships probabilistically (see Fig. 7, '*classification*'). These probabilities can be interpreted as class-related confidence values [*Polikar* 2006; *Kuncheva* 2014] and inform about the most likely class a detection belongs to.

Reassociating Templates and Detections with Equalized Waveforms

As the large weight of the *S*-phase dominates the waveform similarity, detections are associated with a template mainly based on the shape of their *S*-phase. Consequently, remaining waveform parts (especially information in the coda wave, see Fig. 4) have a minor influence, although they carry important information about the travel path and, therefore, the source location. Our previously presented 'waveform equalizing' procedure can counteract this effect. But because the window $\mathbf{e}$ (equation (2)) is unique at every waveform sample, the equalized template waveforms could not be used for template matching itself. Therefore, we re-employ waveform equalizing as follows to improve the template–detection associations: After equalizing all template and detection waveforms, we cross-correlate each detection with every template and reassociate it with the template that produces the highest CC value. For very small events that approach the noise level, the window $\mathbf{e}$ will lose its effect: With increasing noise level, $\mathrm{Kurt}(\mathbf{x})$ tends to zero leaving the original waveform unaltered. The recalculated CC values are generally smaller than the original ones because this procedure enhances the waveforms' heterogeneity. To remove detections with low CC values but account for the similarity change, a new threshold of 0.40 was applied, which includes the average similarity change (-0.05) among all detections.

17

Magnitude Estimation

For any statistical analysis of seismicity, it is essential to have an earthquake catalog with uniform and consistent magnitudes. To calculate such magnitudes for our detection catalog, we followed the procedure detailed below.

First, we consistently recalculated moment magnitudes for all the events in the combined catalog (3723 events) using the spectral fitting method proposed by *Abercrombie* [1995]. We performed this procedure with event recordings of the borehole stations that were operated over a longer time (OTER2, MATTE, and HALTI; see Fig. 1). After correcting the recorded signals for the instrument response, we estimated the far-field displacement spectra of *S*-waves with an adaptive multitaper method [*Prieto et al.* 2009; *Krischer* 2016]. Then, we fit these observed spectra with theoretical far-field source spectra using a nonlinear least-squares optimization with boundary constraints [implemented in the `scipy` python package, *Jones et al.* 2001]. As proposed by *Abercrombie* [1995], we chose the source model of *Boatwright* [1980] instead of a 'ω-square model' [*Aki* 1967; *Brune* 1970; 1971] to obtain a better fit to the sharp corners caused by the close hypocentral distances in our study. *Deichmann* [2017] observed that the shear wave attenuation of induced earthquakes in the Basel reservoir recorded at OTER2 can be modeled using a quality factor, Q_s, between 100 and 200 and suggested to use $Q_{s,\text{base}} = 140$ for the basement and $Q_{s,\text{sedi}} = 80$ for the sedimentary layers. Hence, we use a Q_s of 140 for OTER2, and the mean of $Q_{s,\text{base}}$ and $Q_{s,\text{sedi}}$ (Q_s of 110) for both HALTI and MATTE. The high-frequency falloff exponent, n, was allowed to vary between 1 and 6. The remaining free fitting parameters were the low-frequency plateau, Ω_0, and the corner frequency, f_c, of the displacement spectra. Ω_0 is related to the seismic moment, M_o [e.g., *Aki and Richards* 2002, Chapter 10], and can be subsequently converted into a moment magnitude, M_w [*Hanks and Kanamori* 1979]. For the conversion between Ω_0 and M_o, we replicated the physical parameters used by GEL (density $\rho = 2650\,\text{kg m}^{-3}$, *S*-velocity $\beta = 3450\,\text{km/s}$, and average radiation pattern of *S*-waves $FS = 0.63$; T. Spillmann, pers. comm., 2015). Our M_w estimates were averaged over all three borehole stations and compare well to GEL's M_w estimates (3552 events, Fig. S4a) and to the M_w estimates calculated by *Bethmann et al.* [2011] for a subset of the SED catalog (195 events, Fig. S4c). For the sake of completeness, we document here that the M_w estimates of GEL and *Bethmann et al.* [2011] are also in good agreement with each other (Fig. S4d).

Unfortunately, the majority of our detected earthquakes was too small to estimate their moment magnitudes robustly. Yet the maximum amplitude of the *S*-phase observed at OTER2, A, could be determined reliably even for the tiniest events when most of the signal was already hidden in noise (see Fig. 3). In a next step, we therefore established individual template-family-based relations between $\log A$ and our newly calculated M_w, which we used to estimate moment magnitudes, M_{wx}, for all earthquakes in our catalog. To do so, we made use of the results of *Deichmann* [2017], who showed that for earthquakes with source-spectral corner frequencies $f_c > f_{\text{max}} \rightarrow \log A \propto \log M_o$ and therefore

$$M_w = \frac{2}{3} \log A + C. \tag{3}$$

Here f_{max} is the high-frequency band limitation of the radiated seismic wave field introduced by seismic attenuation [*Hanks* 1982], and C is a constant. *Deichmann* [2017] found that induced earthquakes recorded at the shallow borehole station MATTE have $f_{max} \approx 20$ Hz and comply with $f_c > f_{max}$ if $M_w \leq 2.1$. Consequently, co-located earthquakes in this magnitude range have identical source-pulse widths and identical normalized waveforms independent of their magnitude [*Deichmann* 2017]. This fact implies that waveforms of these events are amplitude-scaled copies of each other and scale proportional to their seismic moment.

At station OTER2, f_{max} is much higher than at MATTE (f_{max}(OTER2) > 100 Hz; see *Kraft and Deichmann* [2014, Fig. 3]) and, consequently, equation (3) holds for much smaller earthquakes than at MATTE. To use equation (3) for as many earthquakes in our catalog as possible, we apply a 4th-order Butterworth low-pass filter with a corner frequency $f_{LP} = 30$ Hz to the event recordings at OTER2. (f_c was larger than 30 Hz for the vast majority of the analyzed earthquakes, see Fig. S6.) Hence, f_{LP} at OTER2 mimics the effect of the smaller f_{max} at MATTE. Consequently, all co-located earthquakes with $f_c > f_{LP}$ have identical normalized waveforms and obey equation (3). Even though *Deichmann* [2017] proofs equation (3) for displacement amplitudes, it is also valid for velocity or acceleration amplitudes in the case of event families with identical normalized waveforms (see proof in Text S7).

Before measuring A, we additionally applied a 10 Hz Butterworth high-pass filter (4th-order) to remove low-frequency signals, which would otherwise mask the S-phase of smaller events. The resulting 10–30 Hz passband is indicated in red in Fig. 3a. Although it appears that very small events contain no energy above the noise level in this frequency range (represented by the probabilistic power spectral density [PPSD] calculation over 6 years), the red waveforms in Fig. 3c show that the S-phase peak can still be perceived. Based on these band-pass filtered velocity seismograms, we measure A as the maximum of the quadratic mean of OTER2's two horizontal components in the 0.6 s event window: $A = \max \sqrt{\mathbf{x}_1^2 + \mathbf{x}_2^2}$.

The large number of catalog events had allowed us to form many template families in the earlier waveform cluster analysis. For these, we established individual template-family-based relations between $\log A$ and our newly calculated M_w with linear regression analysis. We fit equation (3) with respect to the intercept C (Fig. 8), while also accounting for the uncertainties in our M_w estimates. If a template family is represented by more than one member, these $\log A$–M_w relations account more robustly for the remaining aleatory scatter [*Deichmann* 2017] of our M_w estimates.

Although many of the 2508 template families were singletons (65 %), the 560 non-singleton families (20 of them shown in Fig. 8) showed generally little in-cluster deviation from the $1:\frac{2}{3}$ scaling (mean 1σ: 0.011; max. 1σ: 0.069). This small scattering implied that (1) the slope of $\frac{2}{3}$ is appropriate to constrain the amplitude–magnitude regressions and (2) the intercept C is representative for the individual template-family-based relations. We postulate that the generally small error in C also applies to the remaining template singletons. The different offsets in C represent amplitude differences at OTER2 that are caused by different hypocentral distances as well as different radiation patterns of individual events [see *Deichmann et al.* 2014]. Hence, the individual regressions account for both effects because our previous
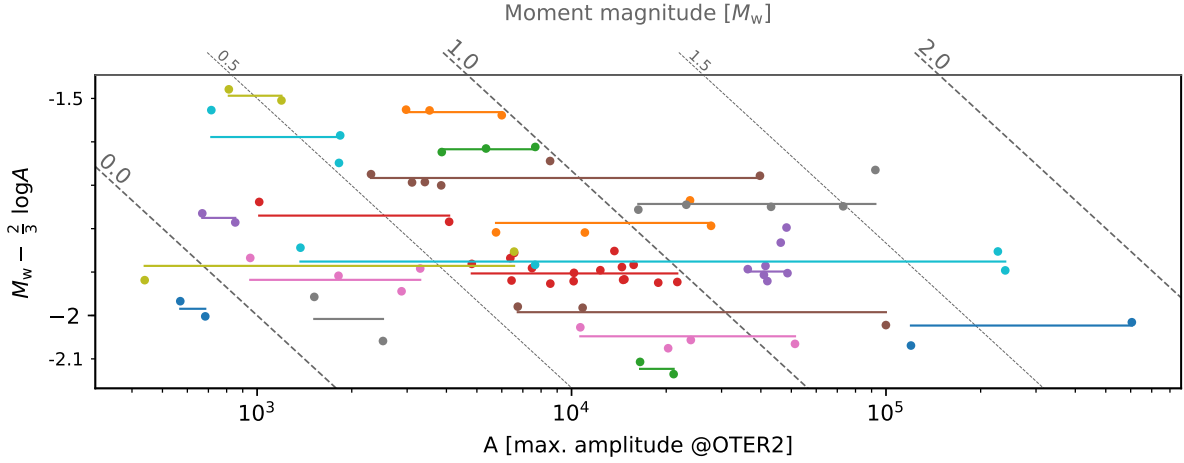
Figure 8: Reduced magnitude versus amplitude plot showing 20 of more than 500 example clusters (i. e. template families) with more than one member. Corresponding regression lines for each family have a fixed $1:\frac{2}{3}$ scaling between M_{w} and $\log A$. The y-axis corresponds to the intercept C in equation (3), which is determined for each family separately via regression analysis. The family memberships (represented by different colors) were determined with waveform cluster analysis using all catalog events. The families in this figure were selectively chosen to show the full amplitude range, full magnitude range, the largest cluster, and also clusters with larger in-cluster deviation.

waveform clustering ensures that cluster members represent closely spaced events with similar focal mechanisms [e. g., *Geller and Mueller* 1980].

Finally, we used the template-family-based $\log A$–M_{w} relations to estimate M_{w} for our detected events template-family-wise from the measured $\log A$. For events with $M_{\mathrm{wx}} \geq 2.0$, f_{c} is generally smaller than f_{LP} at OTER2, and equation (3) is not valid anymore. For these larger events, we replaced M_{wx} with our M_{w} estimates obtained from spectral fitting. Additionally and for the same reason, magnitude-dependent differences in the waveform shapes of these larger events begin to dominate, and their template association becomes less reliable (see drop in similarities in Fig. S11).

As a side note, the spectral fitting also allowed us to estimate the corner frequencies of the catalog events and eventually their Brune stress drops, $\Delta\sigma$. We determined a mean $\Delta\sigma$ of ~15 MPa (see Fig. S6) using the same assumptions as *Goertz-Allmann et al.* [2011], that is, the relations for a circular rupture [*Eshelby* 1957], a shear-velocity of 3450 m/s, and a corner frequency parameter for *S*-waves of 0.375 [*Brune* 1970]. In comparison, *Goertz-Allmann et al.* [2011] obtained a median $\Delta\sigma$ of 2.3 MPa using a subset of the events.
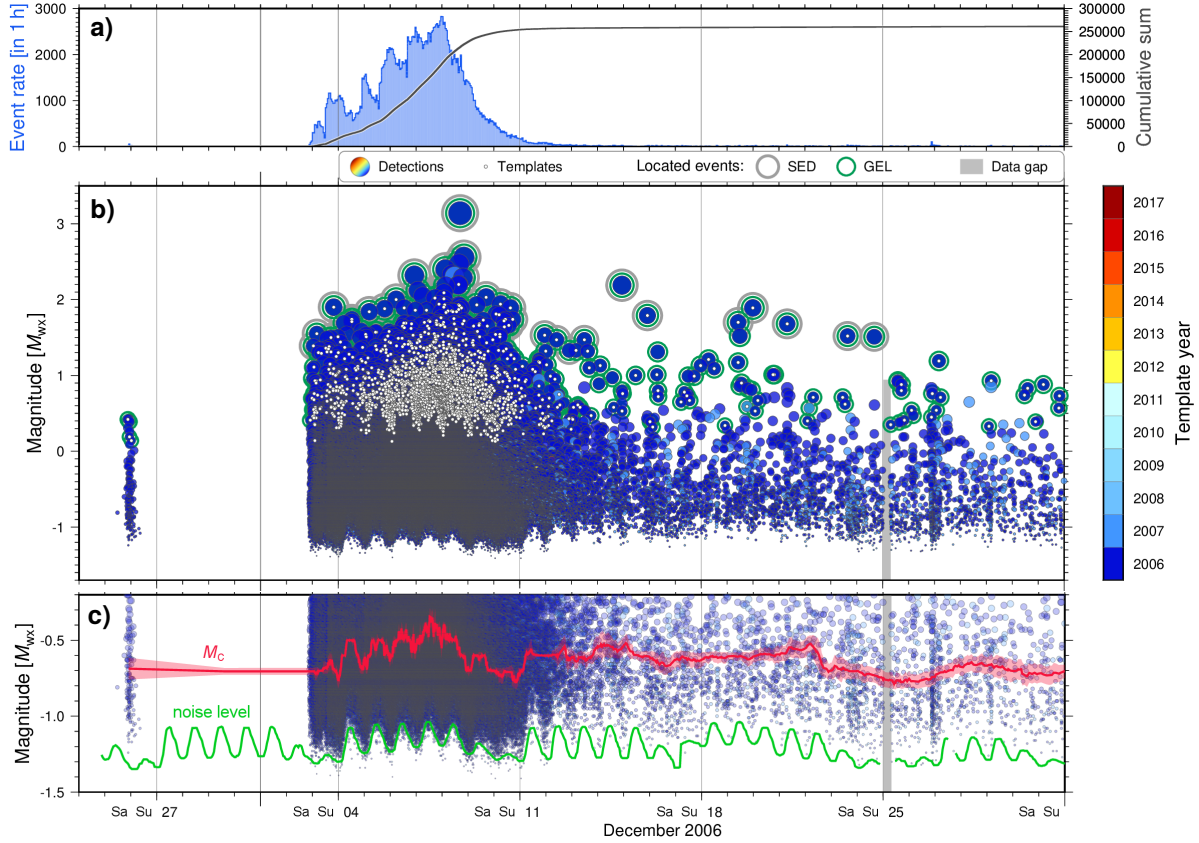
Figure 9: Time history for the period November/December 2006. (a) Histogram (blue) and cumulative rate (gray curve) of detected events illustrating the enormous activity during the injection period. (b) Time versus magnitude; coloring of events as in Fig. 5, which indicates to which template an event is most similar. Existing catalog events are highlighted with a gray (SED) and green ring (GEL); white dots represent template events. (c) A zoom into the lower magnitude range ($M_{wx}-0.2$ to -1.5) showing the varying noise level of the borehole station OTER2 (green curve) and magnitude of completeness (M_c, red curve). M_c was estimated for varying window sizes (see text) and plotted at the center of each window rather than at its end to align better with the noise level. Bootstrap resampling provided a M_c distribution shown by its arithmetic mean (solid red curve) and standard deviation (1σ, red shading). A correlation between M_c and the noise level is shown in Fig. S7. The indicated data gap (gray bar) on 25 December (both at OTER2 and MATTE) reaches up to the detection limit of the surface network ($M_{wx}0.9$). The 'Sa Su' labels on the time axis indicate the weekends.

4 Results

4.1 The New Catalog: Induced Seismicity in Basel in High Resolution

The new catalog consists of 280 941 events with magnitudes down to $M_{wx}-1.5$ ($\approx M_L-3.4$). Of these, 42 events were added by the separate scan at MATTE to complement OTER2 data gaps (detection limit at MATTE $\approx M_{wx}0.1 \,\widehat{=}\, M_L-1.0$). Not all data gaps could be complemented by MATTE, as 46 % of the data gaps overlapped at both stations. At these times, the detection

limit of the SED catalog applied ($\approx M_{wx}0.9 \; \hat{=} \; M_L0.5$). The total number of events in our catalog increased 80-fold compared to the GEL catalog and 1200-fold to the SED catalog. The total seismic moment release covered by the new catalog amounts to 4.96×10^{14} N m ($\approx M_w3.73$). In comparison, the GEL catalog covers 3.01×10^{14} N m ($\approx M_w3.59$) and the SED catalog 1.86×10^{14} N m ($\approx M_w3.45$). This comparison shows that the many newly detected events account for a significant proportion (39 %) of the detectable total seismic moment release in the reservoir.

The First Weeks During and After Stimulation

For the low-rate injection test conducted on 25/26 November, prior to the main reservoir stimulation [*Häring et al.* 2008], we detected 147 events between $M_{wx}-1.2$ and $M_{wx}0.42$ (Fig. 9), which corresponds to $M_L-2.9$ and $M_L-0.6$. The operators detected a similar number of events (146, the GEL catalog contains four of them) due to their initially very low trigger level. Their event list is based on a subsequent manual inspection to remove false detections. (In the main stimulation, the operator increased this threshold dramatically to cope with the manual analysis of detected seismic events.)

The main stimulation started on the evening of 2 December at 18:14 UTC, and the first detected event ($M_{wx}-1.1 \; \hat{=} \; M_L-2.6$) occurred at 20:55, 2.5 hours after injection start. Up to that time, 2 m^3 of water had been injected and the well-head pressure was at 7.4 MPa (the downhole pressure was not recorded at that time). The first event in the GEL catalog ($M_{wx}0.40 \; \hat{=}$ $M_L-0.6$) occurred 70 minutes later at 22:05 (3.7 m^3 of water injected and well-head pressure at 8.6 bar). Within the next 2 hours (6 hours after the start of the injection), the catalog spans over 2.5 orders of magnitude as the sequence reached event magnitudes of up to $M_{wx}1.1$ ($\hat{=}$ $M_L0.4$). Over the course of the following 5 days, the seismicity increased steadily both in rate (Fig. 9a) and magnitude (Fig. 9b) in response to the rising pressure and flow rate. When the seismic activity reached an unexpectedly strong level ($M_{wx}2.4 \; \hat{=} \; M_L2.6$ on 8 December, at 03:06), the injection was reduced 1 hour later and halted 8 hours later at 11:33 with the well being closed (shut-in) [*Häring et al.* 2008]. As a consequence, the event rate dropped but was still at a high level ($M_{wx} > 2.0 \; \hat{=} \; M_L > 2.0$), including the widely-felt $M_{wx}3.1 \; \hat{=} \; M_L3.4$ event (8 December, at 16:48). Shortly after this largest event, the well was reopened (bleed-off) at 18:46, and the event rate dropped further.

Despite a decaying seismicity, our catalog reveals that the reservoir was continuously seismically active. When considering GEL events only (green circled in Fig. 9b), the seismicity appears much more sporadic with magnitudes mainly above $\approx M_{wx}0.3 \; \hat{=} \; M_L-0.8$. Compared to the GEL catalog, our catalog adds events that are almost two magnitude units smaller and that were previously missed. For whole December 2006, it now contains more than 260 000 events and resolves event rates that are as high as 2800 events per hour (see Fig. 9a), that is, on average almost one detectable event per second. At such high rates, events overlapped (see Fig. 2c) but could often be detected thanks to the high sensitivity of the matched filter approach.

**Time-Varying Noise Level**

Our new catalog also reveals a daily variation of the lower detection limit, especially during injection. At night, smaller events could be detected than during the day. At weekends, this lower detection limit lasted longer than during the week. These patterns suggest that the time-varying noise level in the city affects the capability of the matched filter to detect the smallest events. The varying noise level can also be seen in the PPSD (similar to *McNamara and Buland* [2004]) in Fig. 3a. To estimate a noise-dependent lower magnitude limit over time, we (1) extracted the noise amplitude via a PPSD analysis (5 %-percentile to focus on noise rather than event signals) in the same frequency band in which we determined event amplitudes (10–30 Hz, red in Fig. 3a), (2) converted this acceleration amplitude into a root mean square velocity amplitude [*Bormann* 1998], and (3) subsequently into magnitude units using the template-family-based log A–M_w relation with the smallest intercept (Fig. 8). The retrieved "noise magnitude", M_{wn} (green curve in Fig. 9c), follows the lowest detectable magnitudes very closely over time and shows the daily and weekly variation of the noise level—even highlighting the Christmas holidays (25/26 December).

We investigated whether also the magnitude of completeness, M_c, was subject to these variations. We determined M_c for varying event window sizes (initially 147 events to capture the prestimulation as separate bin, then 3000 until 10 December 18:00, 1000 until 13 December 00:00, and 500 afterward) with the maximum curvature method [*Wiemer and Wyss* 2000] including an additional correction of +0.2 magnitude units [*Wössner and Wiemer* 2005]. The red curve in Figure 9c shows that the M_c level also exhibited these temporal variations. The strong correlation between M_{wn} and M_c can especially be resolved during the stimulation (Pearson's R of 0.80, see Fig. S7). The template set shows the same temporal patterns during the stimulation (white dots in Fig. 9b) since also the GEL catalog was subjected to a daily variation of its smallest magnitudes.

Long-Term Decay and Renewed Increase of Seismicity

In the following months and years, seismicity continued to decay until 2012 (Fig. 10). Compared to previously existing catalogs, our new catalog reveals that larger events are accompanied by a (temporal) clustering behavior of seismicity, for example, on 6, 12, and 15/16 of January (Fig. 11). Generally, many events occur in clusters, for example, on 5, 8, 14, 17/18, 22, 25, 28 of January, throughout 2007 (not shown), as well as in April 2008, July 2009, and April/June 2010 (Fig. 10). Between 2008 and mid-2010, the previously existing catalogs showed only infrequent seismic activity (mean inter-event time: 28 days), which is now resolved as an almost permanent activity in our catalog (mean inter-event time: 1.1 days). For the period from mid-2010 to mid-2012, no events were previously reported, but our results show that the sequence was indeed still active at a very low level (on average ~11 events per month). In 2012, however, almost 6 years after the injection, seismicity suddenly increased again with magnitudes up to $M_{wx} 2.1 \cong M_L 1.9$ [*Wiemer et al.* 2017a; *Diehl et al.* 2018] and continued to occur preferentially in episodic sequences until 2018.
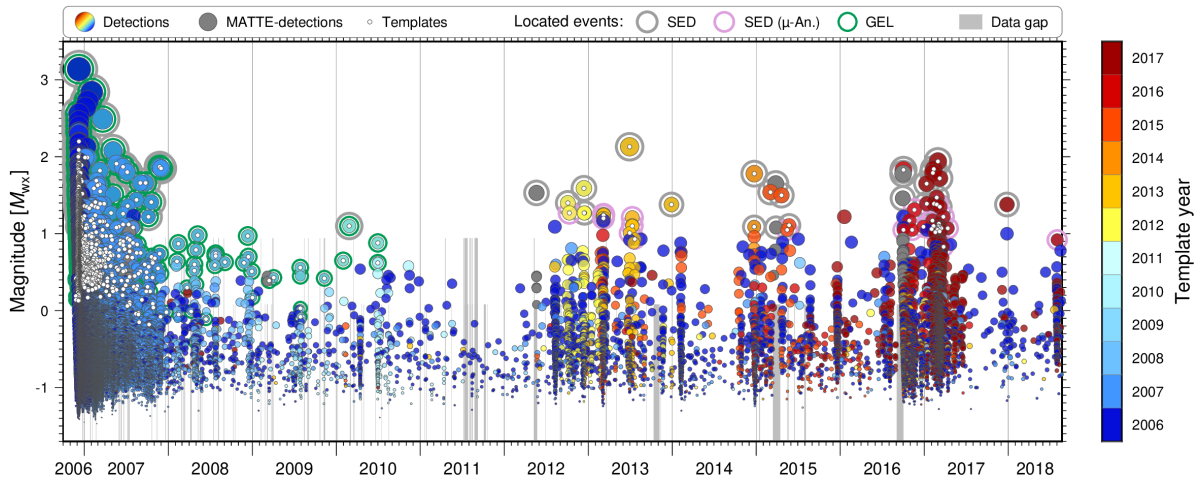
Figure 10: Time history of earthquake detections for the whole study period. The coloring is the same as in Figs. 5, 9, and 11. Gray bars (data gaps of OTER2) indicate two alternative detection limits (by their height): $M_{wx}0.1$, when MATTE was operational, and $M_{wx}0.9$—the detection limit of the surface network—when also MATTE had an outage. Only data gaps larger than 1 hour are shown.
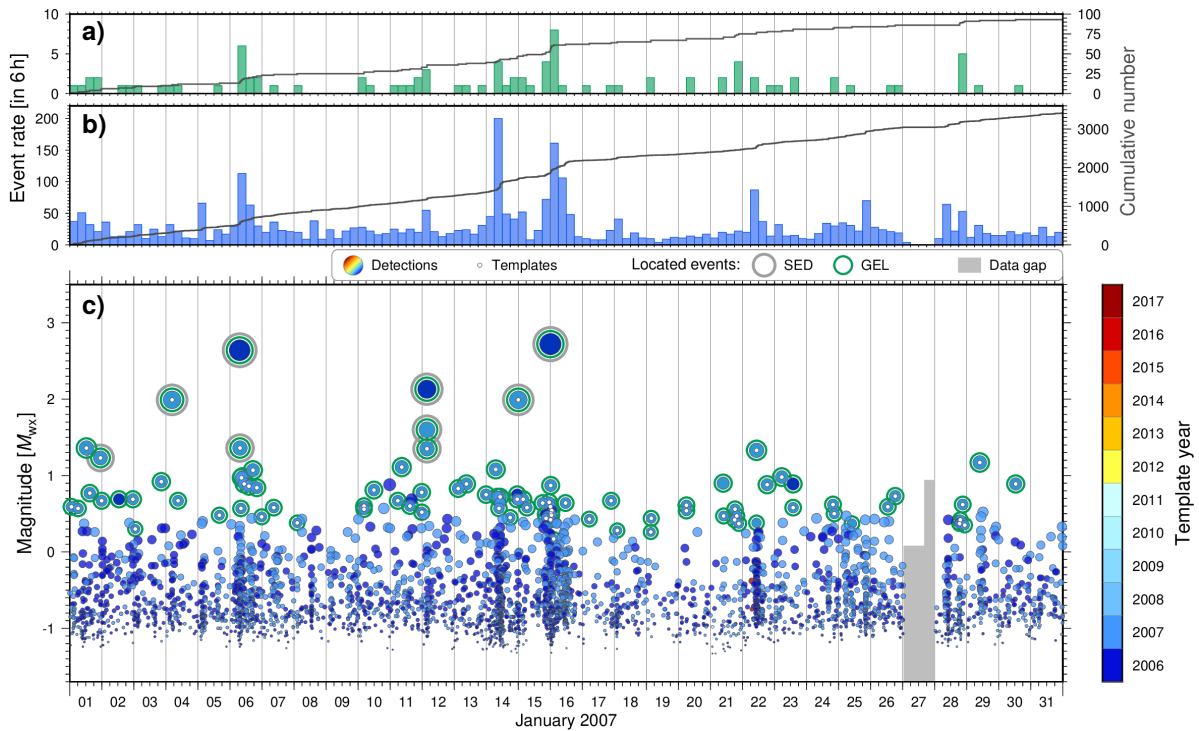
Figure 11: Time history for a zoom into January 2007 to illustrate the higher resolution of temporal event clusters. (a, b) Comparison of the resolved event rate by the GEL catalog (a) and our catalog (b); the gray curves represent the cumulative number of detected events. (c) The detected events (time versus magnitude) are colored as in Figs. 5 and 9. Data gaps (gray bars) on 25 January indicate two different detection limits, as explained in the caption of Fig. 10.

24

Template associations change gradually over the years (see color coding in Fig. 10) from dark blue (early templates) over yellow to red (later templates). Assuming a close spatial proximity of the detections to their templates, this transition corresponds to an expansion of the seismic cloud, which was already observed in previous studies [e. g., *Dyer et al.* 2008; 2010; *Deichmann and Ernst* 2009; *Deichmann et al.* 2014]. Especially clusters that occur after 2012 are predominantly associated with templates from the same year indicating that seismicity in these years occurs on faults that had not been active before. But the general presence of associations with early templates throughout the study period indicates that the central part of the reservoir remained active.

4.2 Details on the Event Classification

From inspecting the classification results, we found the target detections (i. e. the 'reservoir' class) to be well-separated from the non-target (i. e. false) detections. Only during the high-rate injection period in December 2006 we noticed an overproportional number of non-target events per time unit (~7400 'outside' events, and ~4000 'time-err' detections). Most of them were very small events and misclassified due to the extremely high event rate that caused events to overlap and waveform features to not represent a single event anymore. In such cases, the trained classifier had problems to discriminate event classes properly and was overconfident toward non-'reservoir' classes. To correct for this overconfidence, we took advantage of the probabilistic classification estimates (see Fig. 7) and reclassified non-'reservoir' detections as 'reservoir' events if their associated 'reservoir' class probability exceeded a certain threshold (see Text S6). In this way, non-target detections only kept their class if they had a very low confidence level of belonging to the 'reservoir' class. This adjustment recovered ~9200 'reservoir' events.

For the whole study period, 280 899 of the total 388 614 detections were identified as 'reservoir' events. The majority of these events occurred in December 2006 during the injection period (~90 %, green curve in Fig. 12). The 'time-err' detections correlated strongly with the occurrence of 'reservoir' events as they represented wrong phase detections, but at a much lower rate (~3000 in total, blue curve in Fig. 12). 'Noise' class detections were omnipresent, but preferentially occurred in swarms from 2012 onward for some unknown reason (gray curve in Fig. 12). 'Outside' events instead occur at an almost constant rate (red curve in Fig. 12), suggesting that they relate to a constant background seismicity in the vicinity of the reservoir (~185 events per year).

4.3 First Statistical Insights

Event Size Distributions

Figure 13 compares our catalog with the existing catalogs in three different periods in terms of their frequency–magnitude distribution (FMD). To characterize the FMDs quantitatively,
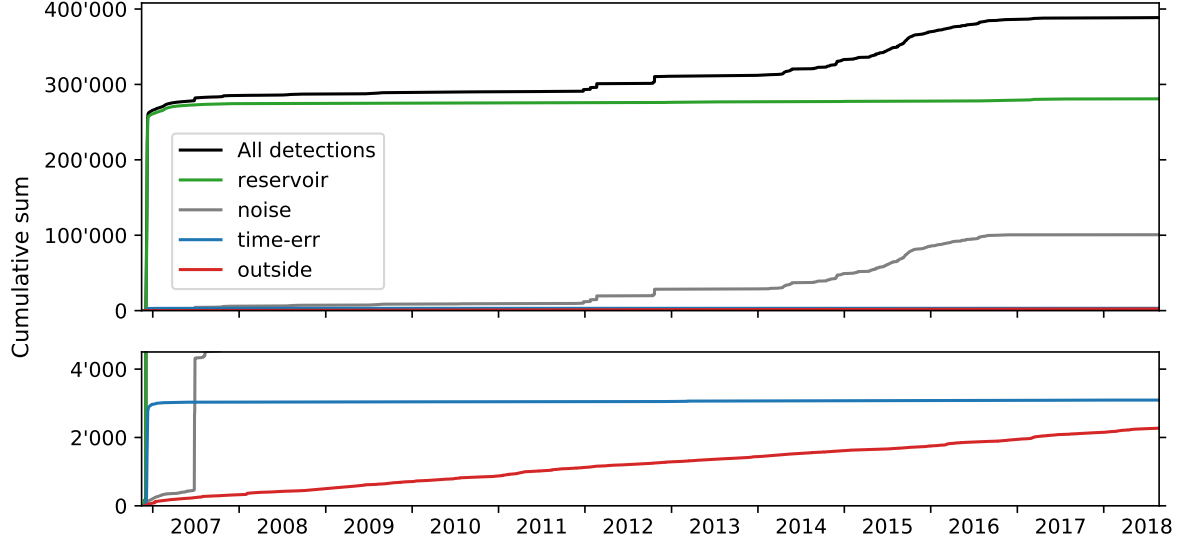
Figure 12: Cumulative histograms of all detections (black) and separated event classes (green, gray, and red; see legend) after applying our machine learning approach. The lower plot is a vertical zoom to show the constant event rate of the 'outside' earthquake class (red) for 2007 onward.

we assumed a distribution according to the Gutenberg–Richter (GR) relation [*Gutenberg and Richter* 1949] and estimated their *b*-values (the slope of the cumulative FMD expressing the ratio of small to large earthquakes) and *a*-values (the intercept at $M = 0.0$) based on the formulas of *Tinti and Mulargia* [1987] and *Marzocchi and Sandri* [2003]. For calculating the completeness magnitude, M_c, we mainly used the Median-Based Analysis of the Segment Slope method [MBASS, *Amorèse* 2007], which is more suited for gradually curved FMDs below M_c (A. Mignan, pers. comm., 2018), with a conservative correction of $+2\sigma$, hereinafter referred to as $M_c^{MBASS}(+2\sigma)$. For a reference, we also considered two alternative estimation methods: (1) maximum curvature [MAXC, *Wiemer and Wyss* 2000] including an additional correction of $+0.2$ magnitude units [*Wössner and Wiemer* 2005], hereinafter referred to as $M_c^{MAXC}(+0.2)$; and (2) *b*-value stability [*Cao and Gao* 2002] (referred to as "MBS" by [*Wössner and Wiemer* 2005; *Mignan* 2012]). MAXC and MBS estimates are indicated in Fig. 14 for the different periods, but not further interpreted. Parameter estimates for 10 000 bootstrap samples of each catalog provided a distribution for each parameter given as arithmetic mean and 1σ-uncertainty. For estimating $M_c^{MBASS}(+2\sigma)$ and associated distributions for the *a*- and *b*-value, we performed two rounds of bootstrapping: first to obtain $M_c^{MBASS}(+2\sigma)$ itself, and a second time for the arithmetic mean and 1σ-uncertainty for the *a*- and *b*-values.

For the first period (Nov./Dec. 2006, Fig. 13a), the FMD shows an overall nonlinear distribution of magnitudes, including an apparent break at $\sim M_{wx} 1.0$. The *b*-value as function of M_{co} [*Cao and Gao* 2002; *Mignan and Woessner* 2012] increased continuously until this point ($\sim M_{wx} 0.96$, see Fig. 14a), where the *b*-value of 1.68 ± 0.05 for our catalog approximately agrees with the *b*-value at the apparent breaking point of the GEL catalog (1.62 ± 0.05 at $M_w 0.90$). (Note that above $M_w 0.85$, our M_w estimates have a shift of $+0.07$ magnitude units compared to

26
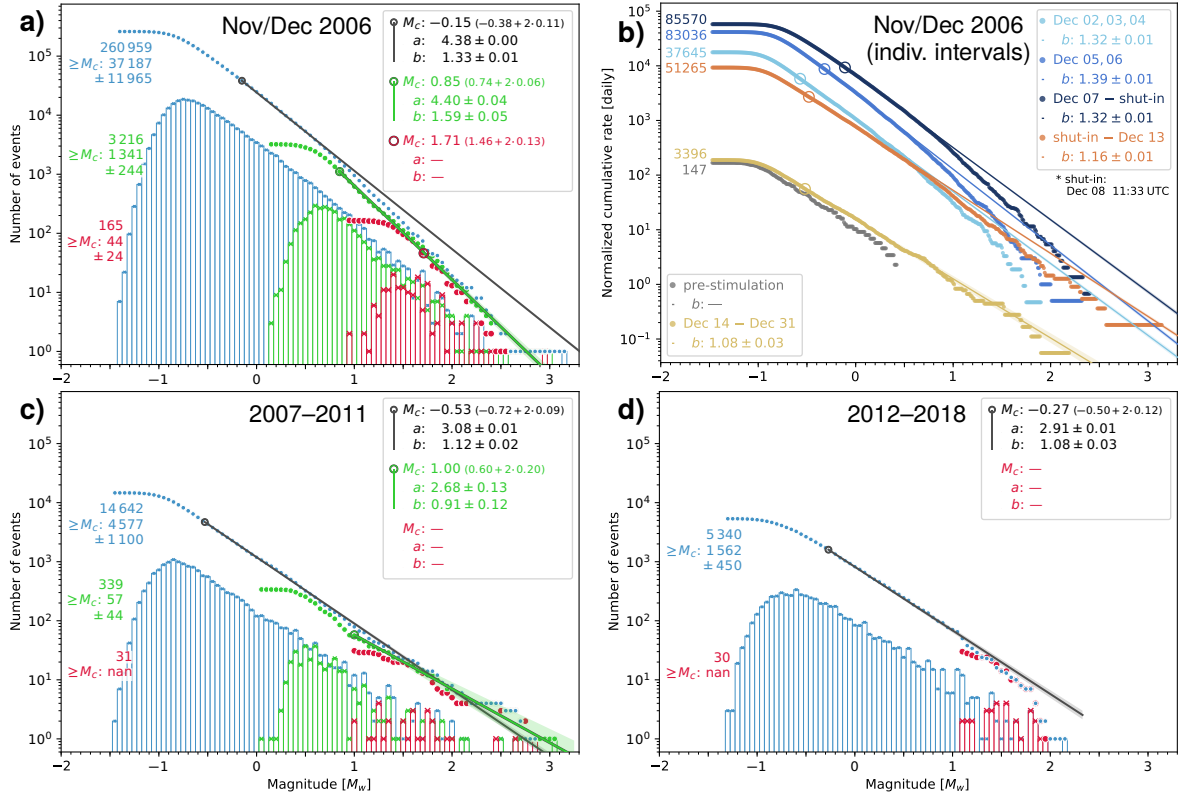
Figure 13: Frequency–magnitude distributions (FMDs) for different time periods. (a, c, d) Comparison of FMDs of our detection catalog (blue), the GEL borehole-network catalog (green), and the SED catalog (red) for three time periods indicated by the headline in each plot. The FMDs are given in terms of a discrete (bars) and cumulative (dots) distribution. (b) Comparison of cumulative FMDs for individual subintervals (see legend) of the November/December 2006 time period based on our catalog; note the normalized y-axis to a daily rate. The number of contributing events is given in the top left of each FMD. (a–d) For each FMD, we determined the magnitude of completeness with $M_c^{\mathrm{MBASS}}(+2\sigma)$ and fit the Gutenberg–Richter relation (*a*- and *b*-value, see legend).

the GEL catalog; see Fig. S4b.) The SED catalog did not contain enough events above its $M_c^{\mathrm{MBASS}}(+2\sigma)$ to determine GR parameter values; however, with $M_c^{\mathrm{MAXC}}(+0.2) = 1.45$ we determined a *b* value of 1.56 ± 0.18 for this time period. For our catalog, all methods estimate M_c in the nonlinear part, below the apparent break at $\sim M_{\mathrm{wx}}0.96$ (see Fig. 14a). As a result, the estimate of $M_c^{\mathrm{MBASS}}(+2\sigma) = -0.15$, for instance, causes the GR fit to overestimate the occurrence of larger magnitudes (see Fig. 13a). The scaling break could not be previously seen, because it appears just above the M_c of the GEL catalog ($M_c^{\mathrm{MBASS}}(+2\sigma) = M_{\mathrm{w}}0.85$). The strong deviation from the GR relation above $\sim M_{\mathrm{wx}}0.96$ is only observed in this high-rate seismicity period. We present possible explanations in the *Discussion* section.

To inspect when the nonlinearity including the scaling break occurs, we additionally determined FMDs for six subintervals (one for the prestimulation, three during the main stimulation, and two for the post-shut-in phase, see Fig. 13b). See Fig. 14b for corresponding *b*-value sensitivity analyses. Apparently, the FMD becomes more nonlinear at higher event rates during the
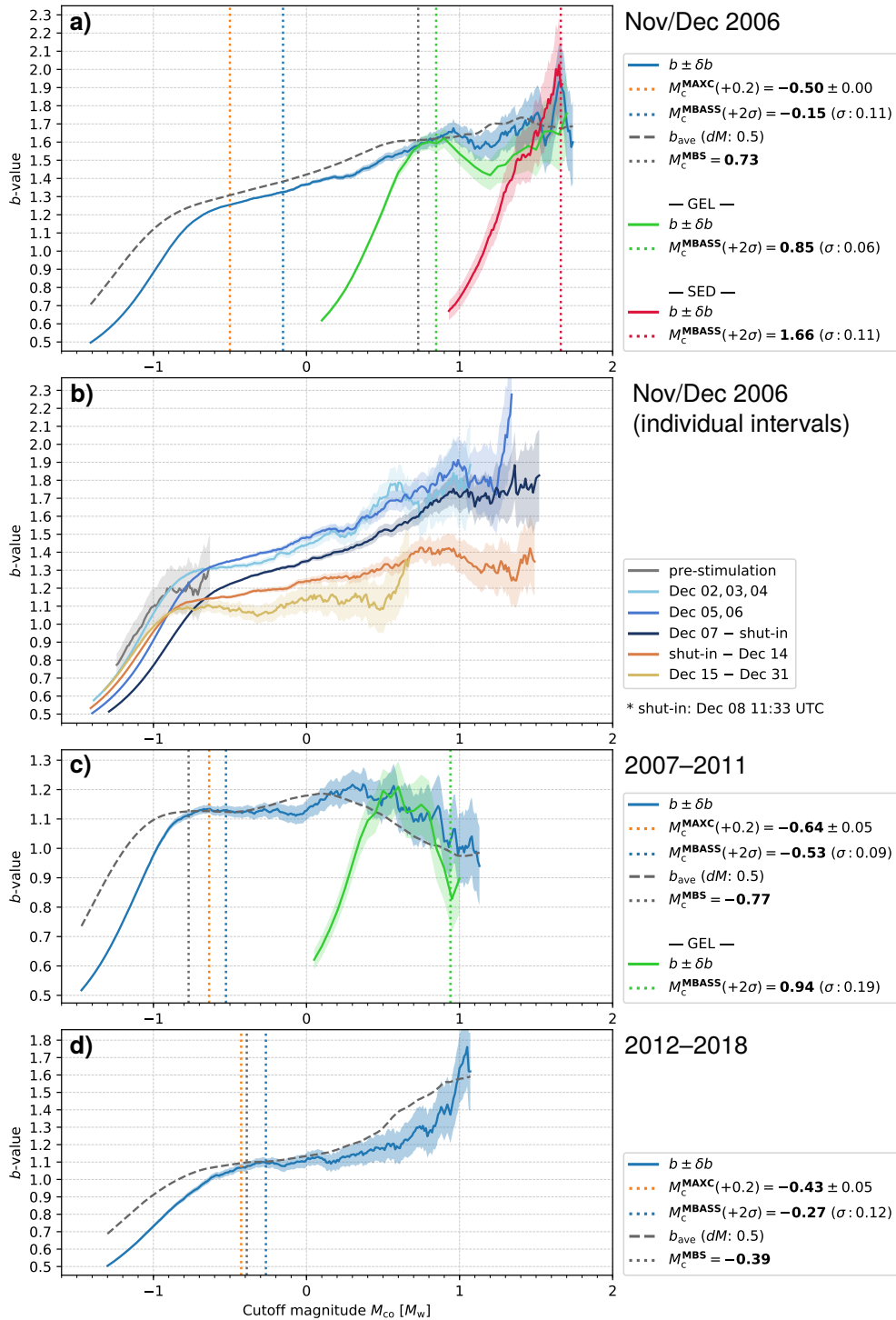
Figure 14: *b*-value sensitivity analyses for the time periods indicated by the headline in the top right of each plot (corresponding to Fig. 13 with the same color scheme). The *b*-value was estimated as a function of cutoff magnitude, M_{co}, with the bias-free maximum likelihood method of *Tinti and Mulargia* [1987] (1σ-uncertainty depicted as shading). Vertical dotted lines indicate estimates of M_c based on three methods (see text): MBASS for every catalog (blue, green, and red), MAXC (orange, only for our catalog), and MBS (gray, only for our catalog). The gray dashed curve is the arithmetic mean, b_{ave}, from *b*-values of successive cutoff magnitudes M_{co} in half a magnitude range ($dM = 0.5$) and used for the estimation of M_c^{MBS}.

stimulation (blue curves in Figs. 13b and 14b). After the shut-in (orange and gold curve), the nonlinearity starts to disappear, also visible as a less pronounced increase of the b-value as function of M_{co} (Fig. 14b). For most intervals in this period, the FMD breaks in scale between $M_w 0.5$–1.0 (see Fig. 13b), which is noticeable by a more rapid increase of the b-value as function of M_{co} in Fig. 14b. Only in the interval after 15 December (gold curve), this scaling break disappears, and the more linear FMD agrees better with the GR relation. The b-value decreased from 05/06 Dec. onward until the end of this period.

In the post-stimulation period (2007–2011, Fig. 13c), the FMD of our catalog is linear over 3 orders of magnitude down to $M_c^{MBASS}(+2\sigma) = -0.53$. The improvement in M_c compared to the existing catalogs is larger than in the previous period: 1.5 orders of magnitude relative to GEL ($M_c^{MBASS}(+2\sigma) = 1.00$), and more than 2 orders relative to the SED catalog (estimate only with $M_c^{MAXC}(+0.2) = 1.70$ due to the few events). The obtained b-value estimate of 1.12 ± 0.02 is much higher than for the GEL catalog (0.91 ± 0.12). The reason is a breaking point in the FMD of the GEL catalog at $M_w 0.95$, which is coincidentally close to its $M_c^{MBASS}(+2\sigma)$ used for the GR fit. This breaking point is not visible in our catalog and indicates missing events in the GEL catalog specifically between $M_w 0.8$ and 1.4. The SED catalog did not contain enough events also in this period to reliably estimate GR parameter values.

In the latest period (2012–2018, Fig. 13d) the FMD of our catalog showed linearity only over 2 orders of magnitude, because of a much more gradual curvature at low magnitudes and a tapering toward higher magnitudes. The estimated $M_c^{MBASS}(+2\sigma) = -0.27$ is slightly higher than in the previous period due to the stronger gradual curvature in the lower magnitudes.

Temporal Variation of Seismicity Parameters

To demonstrate the high resolution of our catalog, we illustrate the temporal variation of seismicity parameters in the period of November/December 2006 using short event windows. Surprisingly, the FMDs in most windows were more angular-shaped [see *Mignan* 2012] and less gradually curved than the overall FMD for the November/December 2006 period (see movie of short-term FMDs and corresponding GR-fits in the Supporting Information). Hence, estimating M_c with the MAXC method is a reasonable choice ($M_c^{MAXC}(+0.2)$, see orange curve in Fig. 15). We choose windows of varying lengths that move through the catalog: initially 147 events to capture the prestimulation as separate bin, then 3000 until 10 December 18:00, 1000 until 13 December 00:00, and 500 afterward. The event windows were moved in steps of $\sim \frac{1}{30}$ of their lengths. We performed a separate continuous analysis using a fixed completeness magnitude $M_c^{FIX} = 0.96$ and a smaller window length of 250 events (purple curves in Fig. 15) to only consider events above the apparent scaling break in the overall FMD (see Fig. 13a). The b-value is estimated again using *Tinti and Mulargia* [1987], which provides bias-free estimates also for small data sets [*Marzocchi and Sandri* 2003]. To reduce fluctuations of the parameter estimates and assess their uncertainties, we performed the analysis for 1000 bootstrap samples in each window. In Fig. 15, the bootstrap distributions are shown as the arithmetic mean (solid curves) and the standard deviation (shadings; 1σ-uncertainty). All parameter estimates are
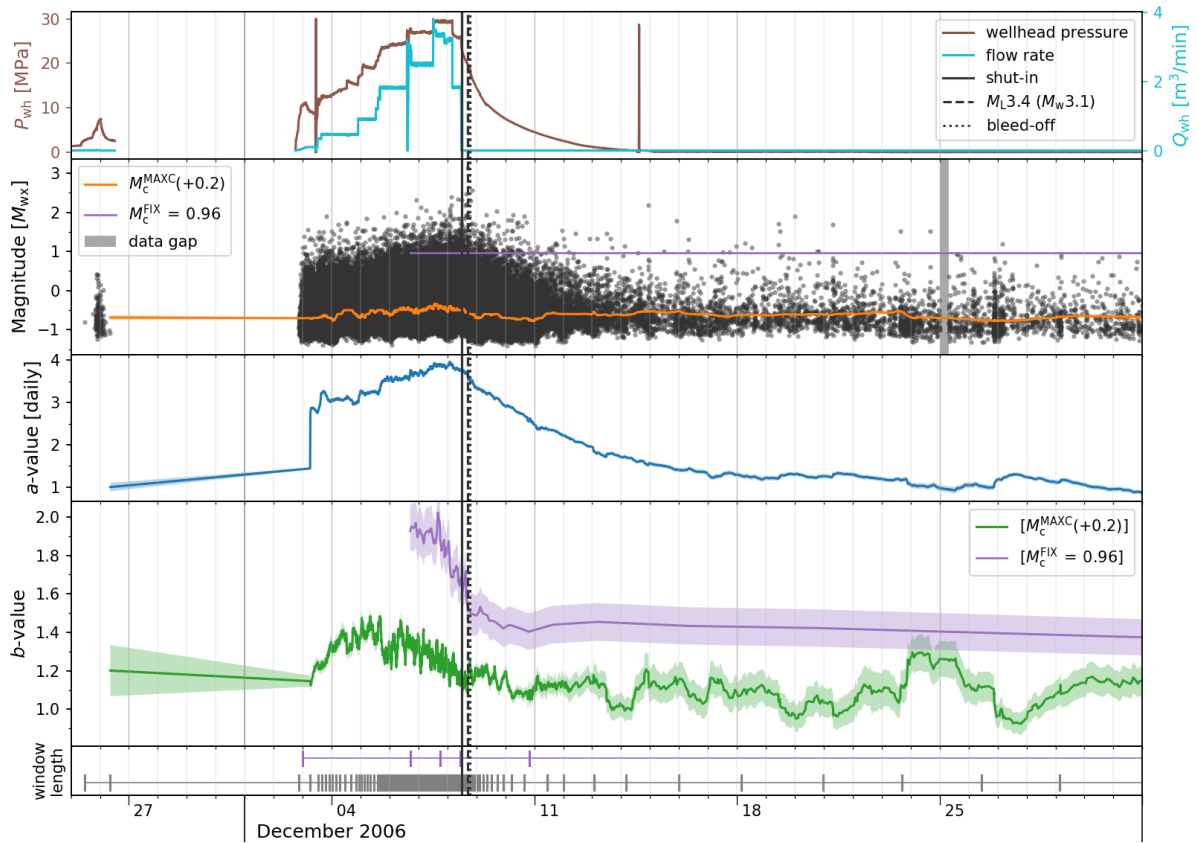
Figure 15: Time history of statistical seismicity parameters of our catalog for the period November/December 2006: time-varying magnitude of completeness $M_{\mathrm{c}}^{\mathrm{MAXC}}(+2\sigma)$ (orange) and corresponding Gutenberg–Richter a- and b-values (dark blue and green, respectively). The purple curves relate to a comparative analysis using a fixed $M_{\mathrm{c}} = 0.96$, which is the apparent scaling break in the overall FMD for this period (see Fig. 13a). For comparison, the top plot shows hydraulic data measured at the wellhead: pressure (brown) and flow rate (cyan). For the continuous analysis, we adjusted the event window lengths to the activity rate (see text). Their time spans are indicated in the lower plot for the two separate analyses.

plotted at the end of each time window to maintain causality and base the estimates only on past processes.

As already noted and shown in Fig. 9c, $M_{\mathrm{c}}^{\mathrm{MAXC}}(+0.2)$ shows a daily and weekly oscillation. The daily a-value (see Fig. 15), however, does not show this oscillation as it corresponds to the activity level, that is, the intercept of the GR-fit at $M0.0$. With ongoing injection, the a-value increased, and progressively larger events occurred. After injection stop and immediate shut-in, the a-value declined in a similar fashion as the well-head pressure, even though large events still occurred. The a-value correlates very well with the wellhead pressure (Pearson's R 0.92). For comparison, the flow rate correlates with the wellhead pressure with a slightly lower Pearson's R of 0.87.

The time-varying b-value estimates (green curve in Fig. 15) based on the time-varying $M_c^{MAXC}(+0.2)$ show a distinct change during the stimulation period: They initially increased from the start of the main stimulation and then decreased to the initial level, while the injection was still continuing and intensifying. The decrease could also be resolved using only the events above $M_c^{FIX} = 0.96$ (violet curve in Fig. 15), but with a delay due to fewer available events and with a more sudden change, that is, not until ~24 hours before the eventual shut-in. The b-value estimates for $M_c^{MAXC}(+0.2)$ are consistently lower than for M_c^{FIX} due to the scaling break at $M_w 0.96$. After the stimulation, the $M_c^{MAXC}(+0.2)$-based b-value remained at a low level for 2 weeks followed by another increase and subsequent decrease between 23–26 December. Note the higher uncertainties due to the decreased event window length from 10 December 18:00 onward.

The time-varying analysis recovers much more events for statistical analysis to a total of 129 480 detected events (50 % of the 260 959 events) above the time-varying $M_c^{MAXC}(+0.2)$, while the overall a- and b-value estimates for November/December 2006 (Fig. 13a) are based only on 15 % of the events. This difference is because the overall M_c depends on the period with the highest incompleteness. Even if the time-varying analysis is based on $M_c^{MBASS}(+2\sigma)$ estimates (not shown), 45 % of the 260 959 events are still above the time-varying completeness.

4.4 Template Statistics

The large number of templates allows us to analyze the systematic change of the magnitude range of the detected events as a function of template magnitude. Figure 16a shows that this range increases slightly with template magnitude. While the largest detectable magnitudes (red) show some correlation with template magnitude (Pearson's $R = 0.71$), the smallest magnitudes (blue) show almost no correlation ($R = 0.25$). We speculate that the former are influenced by the differences in the source-spectral corner frequencies (between the template and the detection), while the latter are dominated by the noise level or the upper band-pass limit. However, we did not find a satisfactory explanation yet.

In Text S8, we detail an additional analysis to investigate retrospectively the detection capability of the template set as a function of its size (see Figs. S9 and S10).
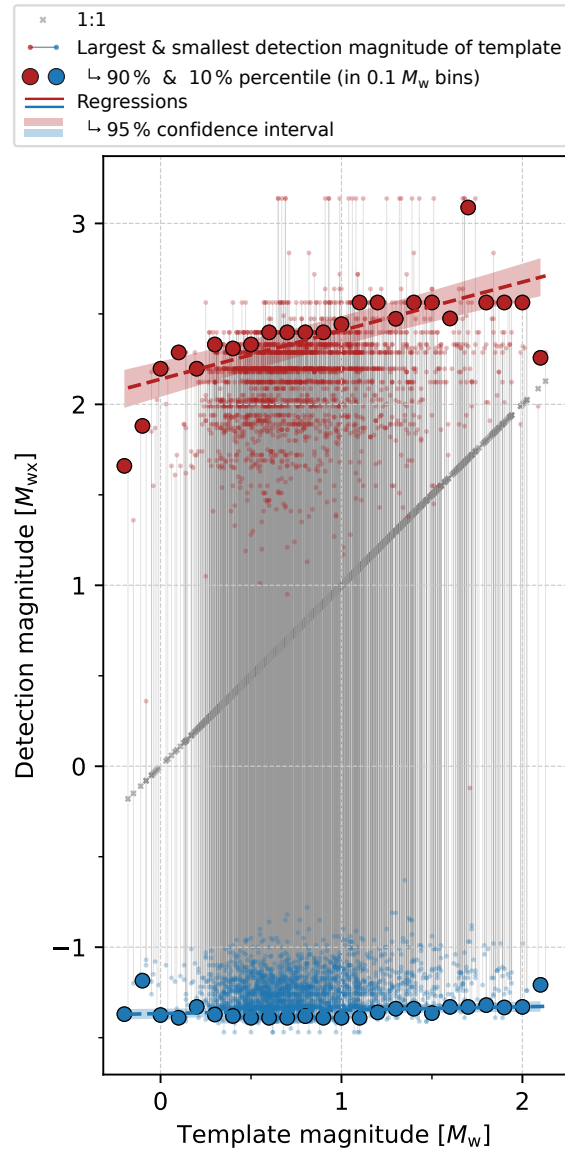
5 Discussion

5.1 Characteristics of the New Catalog

Detection Limit

As expected for any seismicity detector, the detection sensitivity of our detector depended on the noise level of the scanned station. The enhanced visibility of the phenomenon during, and

Figure 16: "Capture range" of templates: indicating the magnitude range of detection magnitudes (red: largest, blue: smallest) as function of template magnitude. Note that the analysis is based on the detectability of all events with respect to each template (and not just the best-matching template association like in the catalog). The smaller circles, connected vertically by a thin gray line, represent the detected magnitude range (maximum and minimum) for each individual template. The distributions of the largest and smallest detection magnitudes were binned to 0.1 M_w intervals and the 90 % and 10 % percentiles were taken, respectively (bigger circles). The gray crosses resemble a 1:1 slope and indicate when the detection magnitude equals the template magnitude.

shortly after, the main hydraulic stimulation can be explained by the strongly elevated event rates as follows: Although generally only a fraction of the events below M_c are detectable, the event rates in this period were so high that there is a much greater chance of detecting events below M_c. These events populate the magnitude range down to the detection level and,

due to their abundance, eventually illuminate the diurnal variation of the noise level. At later times, the seismicity rate declined dramatically, and not enough small events were detected to highlight the day–night variation. The day-to-night difference in detection sensitivity and M_c up to 0.3 M_{wx} units is remarkable, considering that OTER2 is installed at 2.7 km depth below the anthropocentric noise sources of the city. *Bachmann et al.* [2011] already noticed a temporal variation of M_c in the GEL catalog but did not quantitatively attribute it to the diurnal variation of the noise level.

Attempts to Explain the Nonlinearity and Scaling Break in the FMD

For the December 2006 period, we observed an overall nonlinear FMD already in the early phase of the study with a small template set (Fig. S8). This unexpected behavior was the main motivation to use many more and shorter template waveforms, because we initially suspected completeness, detection, or magnitude problems. As the nonlinearity remained also after processing 2508 templates, applying sophisticated post-processing, and improving magnitude determination, we are confident that the nonlinearity is a peculiarity of the Basel sequence. Despite the nonlinearity in the FMD, which is particularly evident during the injection period (see Fig. 13a, b), the FMDs of short time intervals of the catalog (1–5 hours) often show more linearity for small events below the scaling break at M_{wx}0.5–1.0 (see movie in the Supporting Information). It is the occasional nonlinearity of the FMDs in some short time intervals that eventually causes nonlinearity in the overall FMD of the injection period (Fig. 13a) or of longer subintervals (1.5–4 days, see Fig. 13b).

The overall nonlinearity in the FMD might result from a limited detection capability at high seismicity rates when event waveforms are more likely to overlap. In particular, small events are then more difficult to detect, as they are hidden in the coda of larger events. Consequently, from all events that cluster around the same time, only the largest ones can be detected. Note that also small events, for example, M_{wx}0.0, can mask even smaller events. Especially at higher rates, events will therefore be hidden overproportionally with decreasing magnitude. Note that this limitation would apply to any other detection method. Applying a very strict criteria on FMD-linearity, M_c would be much larger for the injection period than we determined here with some established methods.

An alternative explanation for a nonlinear FMD can be given based on currently proposed models of earthquake nucleation. From rate-and-state friction [*Dieterich* 1979] and elastic crack models [*Das and Scholz* 1981], it is expected that rupture starts as slow, aseismic slip, which has to overcome a critical slip patch size before it can accelerate to seismogenic speeds. Only the latter part of this process can be detected seismologically. The critical patch size, which depends on material properties and may vary along the fault plane, would then correspond to a minimum magnitude, M_{min}, below which no microearthquakes can occur. Consequently, the FMD would be depleted around and below M_{min} and exhibit a nonlinear shape at small magnitudes. *Ellsworth and Imanishi* [2012] have argued in this way to explain nonlinear FMDs of microearthquakes in the creeping section of the San Andreas Fault near Parkfield.

The dominant scaling break between $M_{wx}0.5$–1.0 could be caused by the limited size of the stimulated reservoir volume. *Shapiro et al.* [2011] proposed a model where only faults with a sufficiently large segment inside this volume can be seismically activated by the hydraulic stimulation. Consequently, this geometric effect would reduce the probability of large earthquakes and cause a tapered FMD toward higher magnitudes. The scaling break may also be caused by sampling a too short time period. *Howell* [1985] subdivided a catalog that obeys the GR relation and found that its subsets have FMDs predominately of the tapered type. From these perspectives, not the smaller but the larger events break in scale due to their underrepresentation in the FMD. Without further, more detailed analyses, it remains unclear which mechanism or effect causes the nonlinearity and scaling break in the FMD in the injection period. We want to point out that a nonlinearity as observed here has serious consequences for the design of ATLSs (see *Conclusions*).

Systematic errors in the post-processing (e. g., magnitude estimation) can be excluded as a cause for the nonlinearity and scaling break, as these effects disappear at later times after the stimulation. The smoother gradual curvature in the period 2012–2018 (Fig. 13d) below M_c ($\sim M_{wx}-1.3$) is possibly caused by the extension of the seismic cloud (i. e. the influence of the varying hypocentral distance by about 1 km when the seismic cloud reached its maximum extension). This smoothing effect is in addition to the influence of the diurnal change of the noise level on the FMD, which is present in all periods. An additional explanation for the smooth gradual curvature might be that smaller events tend to have lower similarities to the templates (see Fig. S11). These events might then be associated with the wrong template family, resulting in less accurate magnitude estimates, which causes a stretching of the curvature.

Change of the b-value During the Stimulation

The initially increasing and later decreasing b-value during the injection might be related to the pore-pressure perturbations caused by the hydraulic operations. Close to the injection point, where seismicity occurred during hydraulic stimulation, pore pressure is high. According to the effective stress law [e. g., *Hubbert and Rubey* 1959], high pore pressures can reactivate faults with lower differential stress, whereas in low pore-pressure regions, only favorably oriented faults, with tendentiously higher differential stress, can be triggered. If the b-value relates inversely to the differential stress—as proposed, for example, by *Scholz* [2015]—high-pore-pressure regions would promote higher b-values than low-pore-pressure regions [*Goertz-Allmann and Wiemer* 2013]. The growth of the seismic cloud from high to low pore-pressure areas would then explain the b-value drop over time. The b-value drop could also be explained by a distinct change in the geometry of the seismic cloud as it expanded. In the early phase of the injection, the seismic cloud grew in a nearly spherical manner from the injection point. Eventually, the seismicity started to migrate preferentially along a planar fault zone (see Fig. 5 and also *Asanuma et al.* [2008], *Häring et al.* [2008], *Dyer et al.* [2010], and *Deichmann et al.* [2014]), and the b-value decreased to a more typical tectonic value closer to 1.0. We found that the linear expansion started on 5 December, corresponding with the ending of the b-value

increase and the start of its subsequent drop (see Fig. 15). Such behavior would be expected if the *b*-value depended on the fractal dimension of the seismic cloud [e. g., *Turcotte* 1997].

Long-Term Behavior

We demonstrated that our approach can overcome catalog heterogeneity, which is one of the major limitations for analyzing the long-term behavior of seismicity. Inconsistencies in decade-long routine catalogs arise from changes in network geometry, analysis methods, analyst experience and diligence, or when catalogs are combined from different agency contributions. In our case, we eliminated these inconsistencies with a single-station approach for most of the 12-year-long study period when OTER2 was operational. Some longer outages of station OTER2 after 2012 were unfortunate, but they do not disturb the overall interpretability of the seismicity at the site. With more available stations, the remaining catalog inconsistencies in some subperiods can be reduced and quantified, as demonstrated in our case with borehole station MATTE. The backup scan at MATTE complemented data gaps down to its detection level of $M_{wx}0.1$ and completeness level M_c of $\sim M_{wx}0.6$, so that we may have missed only small events. When also MATTE was not functional, the completeness level increased to $\sim M_{wx}1.6$. However, the duration of gaps is small compared to the long operating time of the borehole stations (3.5 % no OTER2 data and 0.7 % neither OTER2 nor MATTE data), and we expect that they do not significantly influence the statistical interpretation of the catalog. With this approach, we enabled a detailed analysis of the renewed seismicity increase that began 6 years after the hydraulic stimulation and could develop mitigation measures for the operator [*Wiemer et al.* 2019].

The temporal seismicity clustering uncovered by our catalog occurred around larger events in aftershock-like fashion, which could indicate that earthquake–earthquake interactions played an important role in this injection-induced sequence. This behavior was already postulated by *Catalli et al.* [2016] based on Coulomb stress analysis of the catalog of *Kraft and Deichmann* [2014]. But we observed also swarm-like seismicity bursts without distinct mainshocks, which may be related to episodic fluid movements in the basement rock as postulated by *Goertz-Allmann et al.* [2017].

Template associations indicate that the inner part of the seismic cloud remained active throughout the 12-year study period, which suggests the existence of a loading mechanism in the Basel reservoir that compensates partly for the Kaiser effect (i. e. stress memory). Yet this observation might be biased by the fact that 85 % of the templates occurred in 2006 which together have a higher chance of being selected as the best-matching template.

5.2 Creating Consistency with Machine Learning

For manually labeling the detections above the threshold of 0.45, we could not base our judgment solely on the waveform shape itself. Due to their often low similarities with our templates, we considered detections as reservoir-related when they had similar *S*-to-*P* arrival

time differences and similar frequency content as the known catalog events, as well as when they temporally coincided with reservoir-related activities and events such as: (1) during the high-rate seismicity of the main stimulation or shortly afterward, (2) within larger temporal clusters that contain known reservoir events, or (3) within moderate-sized temporal clusters that contain detections with high similarities to our template events. Yet, for a number of events in our catalog, we could not substantiate a reservoir-related origin with definite certainty. Still, the probability that such events originated from the reservoir is quite high as it was by far the most active volume intersected by the spherical shell of possible event locations around OTER2 for the observed S-to-P arrival time differences.

The machine-learning-based classification could discriminate the different event classes exceptionally well and added consistency to the high detection sensitivity of our detector. This two-step approach essentially uses the matched filter detector for pretriggering and the machine learning to clean up the detection catalog—similar to how the energy-based detector has been used in numerous studies before as a pretrigger [e. g., *Plenkers et al.* 2013].

An interesting result of the event classification was found for the event class 'outside', whose larger S-to-P traveltime differences (e. g., Fig. 6c) hint to an origin in the wider vicinity of the Basel reservoir. These events can be interpreted as part of the background seismicity in the Basel area. Although their spatial completeness is not known, the fact that their cumulative rate increased rather constantly over the past 12 years allowed us to estimate a lower bound for the background seismicity (185 events per year). About 50 of the ~2300 'outside' events had shorter S-to-P traveltime differences than the 'reservoir' events, indicating that a different source volume, closer to OTER2 than to the reservoir, repeatedly ruptured.

Regarding the origin of the 'noise' detections, which showed a particular temporal occurrence pattern, we could only speculate. The majority of 'noise' detections had low amplitudes and were preferentially detected at night (see Fig. S12, gray line), which suggests a noise-level-dependent detection of signals from a permanently active nearby noise source. A possible location of this noise source might be the industrial harbor area above the Basel reservoir. The 'noise' class also included detections of recording errors (e. g., data spikes and other artifacts), but their amount was much lower than the transient noise detections.

6 Conclusions

We produced a catalog that spans 12 years of induced seismicity in Basel and contains more than 280 000 events (provided in the Supporting Information). It features a consistent detection threshold and reliably estimated magnitudes. We improved upon the detection limit of the GEL catalog by 1.5 moment magnitude units down to $M_{wx}-1.5$ ($\approx M_L-3.4$; ~2.5 M_L units improvement) and now cover seismicity over more than 4 orders of magnitude (in terms of M_L units almost 7 orders). To our knowledge, such a large magnitude span has not been previously achieved for induced seismicity studies of geothermal reservoir stimulations. In this way, our catalog also contributes to complementing the observational scarcity of seismic processes

between laboratory-scale experiments and tectonic seismicity studies. Matched filter detection proved to be a powerful technique to detect new events down to the anthropocentric seismic noise level, even though its application was rather challenging in our case. We faced three main difficulties:

1. the complex and extended fault geometry forced us to use a large template set (2508 events) to reduce the chance of missing events;

2. the impulsive waveform shape and the failing vertical component led to many reservoir-unrelated detections, which we removed with a machine-learning-based classification approach; and

3. to obtain meaningful magnitudes for detections recorded at a single seismometer, we employed template-family-based amplitude–magnitude relations, which implicitly account for the hypocentral distance and radiation pattern.

These advanced techniques made it possible to create a consistent high-resolution catalog. Our approach might prove useful also in other cases with similar obstacles and preconditions. We could significantly reduce the magnitude of completeness M_c down to $M_w-0.15$ ($\approx M_L-1.57$) and perform seismicity analysis in unprecedented detail. We discovered new patterns that have not been identified before:

1. a preferential clustering of seismicity observed throughout the sequence,

2. a daily variation of the detection and completeness level during the stimulation phase, and

3. a b-value decrease while the injection was still ongoing and increasing, which we explained with a change in the expansion of the seismic cloud.

We also observed a nonlinear event size distribution during the hydraulic stimulation, which deviated from the GR relation. This deviation could not be seen in previous catalogs due to their higher M_c. We showed that established methods to determine M_c result in an underestimate of the b-value and consequently in an overestimate of the occurrence probability of larger events. ATLSs in future geothermal projects will guide the hydraulic reservoir stimulation using seismic hazard forecasts that are based on the observed induced seismicity. Hence, ATLS must carefully consider potential scaling breaks or nonlinearities in the FMD during high-rate seismicity periods related to hydraulic stimulation. Otherwise, they may respond unnecessarily conservative and prevent a successful hydraulic stimulation.

This last conclusion does however not imply that advanced detection methods, like the one discussed here, are of no use in the successful implementation of an ATLS to mitigate unacceptable induced seismicity. The improved consistency and sensitivity of our detection algorithm allowed us to identify rapid changes in seismicity rate, earthquake size distribution, and migration. Such changes could indicate potentially hazardous modifications in the seismogenic behavior of a reservoir, may it be under stimulation, production, or after a shutdown. To understand how these seismicity variations can be used to mitigate unacceptable induced seismicity, advanced detection methods need to be applied routinely in these geotechnical operations. We

demonstrated that our approach can also overcome catalog heterogeneity, which makes such methods essential to study and interpret the long-term behavior of induced seismicity.

Acknowledgments

References

Abercrombie, R. E. (1995). "Earthquake source scaling relationships from -1 to 5 M_L using seismograms recorded at 2.5-km depth". In: *Journal of Geophysical Research: Solid Earth* 100.B12, pp. 24015–24036. DOI: 10.1029/95JB02397.

Aki, K. (1967). "Scaling law of seismic spectrum". In: *Journal of Geophysical Research* 72.4, pp. 1217–1231. DOI: 10.1029/JZ072i004p01217.

Aki, K. and P. G. Richards (2002). *Quantitative Seismology*. 2nd ed. Sausalito, California, U.S.A.: University Science Books. URL: https://www.uscibooks.com/aki.htm.

Amorèse, D. (2007). "Applying a change-point detection method on frequency-magnitude distributions". In: *Bulletin of the Seismological Society of America* 97.5, pp. 1742–1749. DOI: 10.1785/0120060181.

Arrowsmith, S. J. and L. Eisner (2006). "A technique for identifying microseismic multiplets and application to the Valhall field, North Sea". In: *Geophysics* 71.2, pp. V31–V40. DOI: `10.1190/1.2187804`.

Asanuma, H., Y. Kumano, H. Niitsuma, U. Schanz, and M. O. Häring (2008). "Interpretation of reservoir structure from super-resolution mapping of microseismic multiplets from stimulation at Basel, Switzerland in 2006". In: *Geothermal Resources Council Transactions* 32, pp. 65–70. URL: `https://www.geothermal-library.org/index.php?mode=pubs&action=view&record=1028296`.

Aster, R. C. and J. Scott (1993). "Comprehensive characterization of waveform similarity in microearthquake data sets". In: *Bulletin of the Seismological Society of America* 83.4, pp. 1307–1314.

Bachmann, C. E., S. Wiemer, B. P. Goertz-Allmann, and J. Woessner (2012). "Influence of pore-pressure on the event-size distribution of induced earthquakes". In: *Geophysical Research Letters* 39.9. DOI: `10.1029/2012GL051480`.

Bachmann, C. E., S. Wiemer, J. Wössner, and S. Hainzl (2011). "Statistical analysis of the induced Basel 2006 earthquake sequence: Introducing a probability-based monitoring approach for Enhanced Geothermal Systems". In: *Geophysical Journal International* 186.2, pp. 793–807. DOI: `10.1111/j.1365-246X.2011.05068.x`.

Baisch, S., L. Ceranna, and H.-P. Harjes (2008). "Earthquake cluster: What can we learn from waveform similarity?" In: *Bulletin of the Seismological Society of America* 98.6, pp. 2806–2814. DOI: `10.1785/0120080018`.

Baisch, S. *et al.* (2009). *Deep Heat Mining Basel: Seismic risk analysis*. SERIANEX Group, Departement für Wirtschaft, Soziales und Umwelt des Kantons Basel-Stadt, Basel.

Bethmann, F., N. Deichmann, and P. M. Mai (2011). "Scaling relations of local magnitude versus moment magnitude for sequences of similar earthquakes in Switzerland". In: *Bulletin of the Seismological Society of America* 101.2, pp. 515–534. DOI: `10.1785/0120100179`.

Bethmann, F., N. Deichmann, and P. M. Mai (2012). "Seismic wave attenuation from borehole and surface records in the top 2.5 km beneath the city of Basel, Switzerland". In: *Geophysical Journal International* 190.2, pp. 1257–1270. DOI: `10.1111/j.1365-246X.2012.05555.x`.

Beyreuther, M., R. Barsch, L. Krischer, T. Megies, Y. Behr, and J. Wassermann (2010). "ObsPy: A Python toolbox for seismology". In: *Seismological Research Letters* 81.3, pp. 530–533. DOI: `10.1785/gssrl.81.3.530`.

Beyreuther, M. and J. Wassermann (2008). "Continuous earthquake detection and classification using discrete Hidden Markov Models". In: *Geophysical Journal International* 175.3, pp. 1055–1066. DOI: `10.1111/j.1365-246X.2008.03921.x`.

Boatwright, J. (1980). "A spectral theory for circular seismic sources; simple estimates of source dimension, dynamic stress drop, and radiated seismic energy". In: *Bulletin of the Seismological Society of America* 70.1. URL: `http://www.bssaonline.org/content/70/1/1`.

Bormann, P. (1998). "Conversion and comparability of data presentations on seismic background noise". In: *Journal of Seismology* 2.1, pp. 37–45. DOI: `10.1023/A:1009780205669`.

Breiman, L. (2001). "Random forests". In: *Machine Learning* 45.1, pp. 5–32. DOI: `10.1023/A:1010933404324`.

Brune, J. N. (1970). "Tectonic stress and the spectra of seismic shear waves from earthquakes". In: *Journal of Geophysical Research* 75.26, pp. 4997–5009. DOI: `10.1029/JB075i026p04997`.

Brune, J. N. (1971). "Correction to "Tectonic stress and the spectra, of seismic shear waves from earthquakes"". In: *Journal of Geophysical Research* 76.20, pp. 5002–5002. DOI: `10.1029/JB076i020p05002`.

Caffagni, E., D. W. Eaton, J. P. Jones, and M. van der Baan (2016). "Detection and analysis of microseismic events using a Matched Filtering Algorithm (MFA)". In: *Geophysical Journal International* 206.1, pp. 644–658. DOI: 10.1093/gji/ggw168.

Cao, A. and S. S. Gao (2002). "Temporal variation of seismic *b*-values beneath northeastern Japan island arc". In: *Geophysical Research Letters* 29.9, pp. 48-1–48-3. DOI: 10.1029/2001GL013775.

Carmichael, J. D. (2016). "A waveform detector that targets template-decorrelated signals and achieves its predicted performance, Part I: Demonstration with IMS data". In: *Bulletin of the Seismological Society of America* 106.5, pp. 1998–2012. DOI: 10.1785/0120160047.

Castellanos, F. and M. van der Baan (2015). "Waveform similarity for quality control of event locations, time picking, and moment tensor solutions". In: *GEOPHYSICS* 80.6, WC99–WC106. DOI: 10.1190/geo2015-0043.1.

Catalli, F., A. P. Rinaldi, V. Gischig, M. Nespoli, and S. Wiemer (2016). "The importance of earthquake interactions for injection-induced seismicity: Retrospective modeling of the Basel Enhanced Geothermal System". In: *Geophysical Research Letters*. DOI: 10.1002/2016GL068932.

Chambers, D. J., K. D. Koper, K. L. Pankow, and M. K. McCarter (2015). "Detecting and characterizing coal mine related seismicity in the Western U.S. using subspace methods". In: *Geophysical Journal International* 203.2, pp. 1388–1399. DOI: 10.1093/gji/ggv383.

Cox, D. R. (1958). "The regression analysis of binary sequences". In: *Journal of the Royal Statistical Society* 20.2, pp. 215–242. DOI: 10.2307/2983890.

Das, S. and C. H. Scholz (1981). "Theory of time-dependent rupture in the Earth". In: *Journal of Geophysical Research: Solid Earth* 86.B7, pp. 6039–6051. DOI: 10.1029/JB086iB07p06039.

Deichmann, N. (2017). "Theoretical basis for the observed break in M_L / M_w scaling between small and large earthquakes". In: *Bulletin of the Seismological Society of America* 107.2, pp. 505–520. DOI: 10.1785/0120160318.

Deichmann, N. and J. Ernst (2009). "Earthquake focal mechanisms of the induced seismicity in 2006 and 2007 below Basel (Switzerland)". In: *Swiss Journal of Geosciences* 102.3, pp. 457–466. DOI: 10.1007/s00015-009-1336-y.

Deichmann, N., T. Kraft, and K. F. Evans (2014). "Identification of faults activated during the stimulation of the Basel geothermal project from cluster analysis and focal mechanisms of the larger magnitude events". In: *Geothermics* 52, pp. 84–97. DOI: 10.1016/j.geothermics.2014.04.001.

Del Pezzo, E., A. Esposito, F. Giudicepietro, M. Marinaro, M. Martini, and S. Scarpetta (2003). "Discrimination of earthquakes and underwater explosions using neural networks". In: *Bulletin of the Seismological Society of America* 93.1, pp. 215–223. DOI: 10.1785/0120020005.

Diehl, T. *et al.* (2013). "Earthquakes in Switzerland and surrounding regions during 2012". In: *Swiss Journal of Geosciences* 106.3, pp. 543–558. DOI: 10.1007/s00015-013-0154-4.

Diehl, T. *et al.* (2018). "Earthquakes in Switzerland and surrounding regions during 2015 and 2016". In: *Swiss Journal of Geosciences* 111.1-2, pp. 221–244. DOI: 10.1007/s00015-017-0295-y.

Dieterich, J. H. (1979). "Modeling of rock friction: 1. Experimental results and constitutive equations". In: *Journal of Geophysical Research* 84.B5, pp. 2161–2168. DOI: 10.1029/JB084iB05p02161.

Douglas, J. and H. Aochi (2014). "Using estimated risk to develop stimulation strategies for Enhanced Geothermal Systems". In: *Pure and Applied Geophysics* 171.8, pp. 1847–1858. DOI: 10.1007/s00024-013-0765-8.

Dyer, B. C., U. Schanz, F. Ladner, M. O. Häring, and T. Spillman (2008). "Microseismic imaging of a geothermal reservoir stimulation". In: *The Leading Edge* 27.7, pp. 856–869. DOI: 10.1190/1.2954024.

Dyer, B. C., U. Schanz, T. Spillmann, F. Ladner, and M. O. Häring (2010). "Application of microseismic multiplet analysis to the Basel geothermal reservoir stimulation events". In: *Geophysical Prospecting* 58.5, pp. 791–807. DOI: 10.1111/j.1365-2478.2010.00902.x.

Edwards, B., T. Kraft, C. Cauzzi, P. Kästli, and S. Wiemer (2015). "Seismic monitoring and analysis of deep geothermal projects in St Gallen and Basel, Switzerland". In: *Geophysical Journal International* 201.2, pp. 1022–1039. DOI: 10.1093/gji/ggv059.

Ellsworth, W. and K. Imanishi (2012). *Gutenberg-Richter breakdown and the smallest earthquakes at the San Andreas Fault Observatory at Depth*. Oral Presentation. ECGS Workshop 2012: Earthquake source physics on various scales. URL: http://www.ecgs.lu/source2012/.

Eshelby, J. D. (1957). "The determination of the elastic field of an ellipsoidal inclusion, and related problems". In: *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 241.1226, pp. 376–396. DOI: 10.1098/rspa.1957.0133.

Falsaperla, S., S. Graziani, G. Nunnari, and S. Spampinato (1996). "Automatic classification of volcanic earthquakes by using multi-layered neural networks". In: *Natural Hazards* 13.3, pp. 205–228. DOI: 10.1007/BF00215816.

Folesky, J., J. Kummerow, and S. A. Shapiro (2015). "Microseismic rupture propagation imaging". In: *Geophysics* 80.6, WC107–WC115. DOI: 10.1190/GEO2014-0572.1.

Friedman, J. H. (2001). "Greedy function approximation: A gradient boosting machine". In: *The Annals of Statistics* 29.5, pp. 1189–1232. DOI: 10.1214/aos/1013203451.

Geller, R. J. and C. S. Mueller (1980). "Four similar earthquakes in central California". In: *Geophysical Research Letters* 7.10, pp. 821–824. DOI: 10.1029/GL007i010p00821.

Giacco, F., A. M. Esposito, S. Scarpetta, F. Giudicepietro, and M. Marinaro (2009). "Support Vector Machines and MLP for automatic classification of seismic signals at Stromboli volcano". In: *Frontiers in Artificial Intelligence and Applications* 204, pp. 116–123. DOI: 10.3233/978-1-60750-072-8-116.

Giardini, D. (2009). "Geothermal quake risks must be faced". In: *Nature* 462.7275, pp. 848–849. DOI: 10.1038/462848a.

Gibbons, S. J. and F. Ringdal (2006). "The detection of low magnitude seismic events using array-based waveform correlation". In: *Geophysical Journal International* 165.1, pp. 149–166. DOI: 10.1111/j.1365-246X.2006.02865.x.

Gischig, V., S. Wiemer, and A. Alcolea (2014). "Balancing reservoir creation and seismic hazard in Enhanced Geothermal Systems". In: *Geophysical Journal International* 198.3, pp. 1585–1598. DOI: 10.1093/gji/ggu221.

Gischig, V. S. and S. Wiemer (2013). "A stochastic model for induced seismicity based on non-linear pressure diffusion and irreversible permeability enhancement". In: *Geophysical Journal International* 194.2, pp. 1229–1249. DOI: 10.1093/gji/ggt164.

Goebel, T. H. W., S. M. Hosseini, F. Cappa, E. Hauksson, J. P. Ampuero, F. Aminzadeh, and J. B. Saleeby (2016). "Wastewater disposal and earthquake swarm activity at the southern end of the Central Valley, California". In: *Geophysical Research Letters* 43.3, pp. 1092–1099. DOI: 10.1002/2015GL066948.

Goertz-Allmann, B. P., S. J. Gibbons, V. Oye, R. Bauer, and R. Will (2017). "Characterization of induced seismicity patterns derived from internal structure in event clusters". In: *Journal of Geophysical Research: Solid Earth* 122.5, pp. 3875–3894. DOI: 10.1002/2016JB013731.

Goertz-Allmann, B. P., A. Goertz, and S. Wiemer (2011). "Stress drop variations of induced earthquakes at the Basel geothermal site". In: *Geophysical Research Letters* 38.9. DOI: 10.1029/2011GL047498.

Goertz-Allmann, B. P. and S. Wiemer (2013). "Geomechanical modeling of induced seismicity source parameters and implications for seismic hazard assessment". In: *Geophysics* 78.1, KS25–KS39. DOI: 10.1190/geo2012-0102.1.

Grigoli, F. *et al.* (2017). "Current challenges in monitoring, discrimination, and management of induced seismicity related to underground industrial activities: A European perspective". In: *Reviews of Geophysics* 55.2, pp. 310–340. DOI: 10.1002/2016RG000542.

Gutenberg, B. and C. F. Richter (1949). *Seismicity of the earth*. Princeton, New Jersey, U.S.A.: Princeton University Press. URL: https://archive.org/details/seismicityofthee009299mbp.

Hakso, A. and M. Zoback (2017). "Utilizing multiplets as an independent assessment of relative microseismic location uncertainty". In: *The Leading Edge* 36.10, pp. 829–836. DOI: 10.1190/tle36100829.1.

Hammer, C., M. Beyreuther, and M. Ohrnberger (2012). "A seismic-event spotting system for volcano fast-response systems". In: *Bulletin of the Seismological Society of America* 102.3, pp. 948–960. DOI: 10.1785/0120110167.

Hanks, T. C. (1982). "f_{max}". In: *Bulletin of the Seismological Society of America* 72.6, pp. 1867–1879.

Hanks, T. C. and H. Kanamori (1979). "A moment magnitude scale". In: *Journal of Geophysical Research* 84.B5, pp. 2348–2350. DOI: 10.1029/JB084iB05p02348.

Häring, M. O., U. Schanz, F. Ladner, and B. C. Dyer (2008). "Characterisation of the Basel 1 Enhanced Geothermal System". In: *Geothermics* 37.5, pp. 469–495. DOI: 10.1016/j.geothermics.2008.06.002.

Harris, D. B. (2006). *Subspace detectors: Theory*. Tech. rep. Livermore, CA: Lawrence Livermore National Laboratory (LLNL). DOI: 10.2172/900081.

Holland, A. A. (2013). "Earthquakes triggered by hydraulic fracturing in south-central Oklahoma". In: *Bulletin of the Seismological Society of America* 103.3, pp. 1784–1792. DOI: 10.1785/0120120109.

Hornik, K., M. Stinchcombe, and H. White (1989). "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5, pp. 359–366. DOI: 10.1016/0893-6080(89)90020-8.

Howell Jr., B. F. (1985). "On the effect of too small a data base on earthquake frequency diagrams". In: *Bulletin of the Seismological Society of America* 75.4, pp. 1205–1207.

Hubbert, M. K. and W. W. Rubey (1959). "Role of fluid pressure in mechanics of overthrust faulting: I. Mechanics of fluid-filled porous solids and its application to overthrust faulting". In: *Geological Society of America Bulletin* 70.2, pp. 115–166. URL: https://pubs.geoscienceworld.org/gsa/gsabulletin/article/70/2/115/5071.

Hunter, J. D. (2007). "Matplotlib: A 2D graphics environment". In: *Computing in Science and Engineering* 9.3, pp. 99–104. DOI: 10.1109/MCSE.2007.55.

James, G., D. Witten, T. Hastie, and R. Tibshirani (2013). *An introduction to statistical learning*. 8th ed. Vol. 103. Springer Texts in Statistics. New York, NY: Springer New York. DOI: 10.1007/978-1-4614-7138-7.

Jones, E., T. Oliphant, P. Peterson, *et al.* (2001). *SciPy: Open source scientific tools for Python*. URL: http://www.scipy.org.

Király-Proag, E., V. Gischig, J. D. Zechar, and S. Wiemer (2018). "Multicomponent ensemble models to forecast induced seismicity". In: *Geophysical Journal International* 212.1, pp. 476–490. DOI: 10.1093/gji/ggx393.

Kraft, T. and N. Deichmann (2014). "High-precision relocation and focal mechanism of the injection-induced seismicity at the Basel EGS". In: *Geothermics* 52, pp. 59–73. DOI: 10.1016/j.geothermics.2014.05.014.

Kraft, T. *et al.* (2009). "Enhanced Geothermal Systems: Mitigating risk in urban areas". In: *Eos, Transactions American Geophysical Union* 90.32, pp. 273–274. DOI: 10.1029/2009EO320001.

Krischer, L. (2016). *mtspec Python wrappers 0.3.2*. DOI: 10.5281/zenodo.321789.

Kuncheva, L. I. (2014). *Combining pattern classifiers*. Second. Hoboken, NJ, USA: John Wiley & Sons, Inc. DOI: 10.1002/9781118914564.

Lewis, J. P. (1995). "Fast normalized cross-correlation". In: *Vision Interface, Canadian Image Processing and Pattern Recognition Society*. URL: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.6062.

Marzocchi, W. and L. Sandri (2003). "A review and new insights on the estimation of the b-value and its uncertainty". In: *Annals of geophysics* 46.December, pp. 1271–1282. DOI: 10.4401/ag-3472.

Masotti, M., S. Falsaperla, H. Langer, S. Spampinato, and R. Campanini (2006). "Application of Support Vector Machine to the classification of volcanic tremor at Etna, Italy". In: *Geophysical Research Letters* 33.20. DOI: 10.1029/2006GL027441.

Maurer, H. and N. Deichmann (1995). "Microearthquakes cluster detection based on waveform similarities, with an application to the Swiss Alps". In: *Geophys. J. Int.* 123, pp. 588–600. DOI: 10.1111/j.1365-246X.1995.tb06873.x.

McNamara, D. E. and R. P. Buland (2004). "Ambient noise levels in the continental United States". In: *Bulletin of the Seismological Society of America* 94.4, pp. 1517–1527. DOI: 10.1785/012003001.

Mena, B., S. Wiemer, and C. Bachmann (2013). "Building robust models to forecast the induced seismicity related to geothermal reservoir enhancement". In: *Bulletin of the Seismological Society of America* 103.1, pp. 383–393. DOI: 10.1785/0120120102.

Menke, W. (1999). "Using waveform similarity to constrain earthquake locations". In: *Bulletin of the Seismological Society of America* 89.4, pp. 1143–1146. URL: https://pubs.geoscienceworld.org/ssa/bssa/article/89/4/1143/103090/.

Message Passing Interface Forum (1997). *MPI-2: Extensions to the Message-Passing Interface*. Tech. rep. University of Tennessee, Knoxville, Tennessee. URL: https://www.mpi-forum.org/docs/.

Mignan, A. (2012). "Functional shape of the earthquake frequency-magnitude distribution and completeness magnitude". In: *Journal of Geophysical Research: Solid Earth* 117.B8. DOI: 10.1029/2012JB009347.

Mignan, A., M. Broccardo, S. Wiemer, and D. Giardini (2017). "Induced seismicity closed-form traffic light system for actuarial decision-making during deep fluid injections". In: *Scientific Reports* 7.1. DOI: 10.1038/s41598-017-13585-9.

Mignan, A., D. Landtwing, P. Kästli, B. Mena, and S. Wiemer (2015). "Induced seismicity risk analysis of the 2006 Basel, Switzerland, Enhanced Geothermal System project: Influence of uncertainties on risk mitigation". In: *Geothermics* 53, pp. 133–146. DOI: 10.1016/j.geothermics.2014.05.007.

Mignan, A. and J. Woessner (2012). "Estimating the magnitude of completeness for earthquake catalogs". In: *Community Online Resource for Statistical Seismicity Analysis* April. DOI: 10.5078/corssa-00180805..

Mousavi, S. M., S. P. Horton, C. A. Langston, and B. Samei (2016). "Seismic features and automatic discrimination of deep and shallow induced-microearthquakes using neural network and logistic regression". In: *Geophysical Journal International* 207.1, pp. 29–46. DOI: 10.1093/gji/ggw258.

Mukuhira, Y., H. Asanuma, H. Niitsuma, and M. O. Häring (2013). "Characteristics of large-magnitude microseismic events recorded during and after stimulation of a geothermal reservoir at Basel, Switzerland". In: *Geothermics* 45, pp. 1–17. DOI: 10.1016/j.geothermics.2012.07.005.

Pedregosa, F. *et al.* (2011). "Scikit-learn: Machine Learning in Python". In: *Journal of Machine Learning Research* 12, pp. 2825–2830. DOI: 10.1007/s13398-014-0173-7.2.

Peng, Z. and P. Zhao (2009). "Migration of early aftershocks following the 2004 Parkfield earthquake". In: *Nature Geoscience* 2.12, pp. 877–881. DOI: 10.1038/ngeo697.

Perol, T., M. Gharbi, and M. Denolle (2018). "Convolutional neural network for earthquake detection and location". In: *Science Advances* 4.2. DOI: 10.1126/sciadv.1700578.

Peterson, J. (1993). *Observations and modeling of seismic background noise*. USGS Open File Report 93-322. U.S. Geological Survey. DOI: 10.3133/ofr93322.

Plenkers, K., J. R. R. Ritter, and M. Schindler (2013). "Low signal-to-noise event detection based on waveform stacking and cross-correlation: application to a stimulation experiment". In: *Journal of Seismology* 17.1, pp. 27–49. DOI: 10.1007/s10950-012-9284-9.

Polikar, R. (2006). "Ensemble based systems in decision making". In: *IEEE Circuits and Systems Magazine* 6.3, pp. 21–45. DOI: 10.1109/MCAS.2006.1688199.

Prieto, G. A., R. L. Parker, and F. L. Vernon (2009). "A Fortran 90 library for multitaper spectrum analysis". In: *Computers and Geosciences* 35.8, pp. 1701–1710. DOI: 10.1016/j.cageo.2008.06.007.

Provost, F., C. Hibert, and J.-P. Malet (2017). "Automatic classification of endogenous landslide seismicity using the Random Forest supervised classifier". In: *Geophysical Research Letters* 44.1, pp. 113–120. DOI: 10.1002/2016GL070709.

Ripperger, J., P. Kästli, D. Fäh, and D. Giardini (2009). "Ground motion and macroseismic intensities of a seismic event related to geothermal reservoir stimulation below the city of Basel—observations and modelling". In: *Geophysical Journal International* 179.3, pp. 1757–1771. DOI: 10.1111/j.1365-246X.2009.04374.x.

Schaff, D. P. (2008). "Semiempirical statistics of correlation-detector performance". In: *Bulletin of the Seismological Society of America* 98.3, pp. 1495–1507. DOI: 10.1785/0120060263.

Schaff, D. P. and F. Waldhauser (2010). "One magnitude unit reduction in detection threshold by cross correlation applied to Parkfield (California) and China seismicity". In: *Bulletin of the Seismological Society of America* 100.6, pp. 3224–3238. DOI: 10.1785/0120100042.

Scholz, C. H. (2015). "On the stress dependence of the earthquake b value". In: *Geophysical Research Letters* 42.5, pp. 1399–1402. DOI: 10.1002/2014GL062863.

Shapiro, S. A., O. S. Krüger, C. Dinske, and C. Langenbruch (2011). "Magnitudes of induced earthquakes and geometric scales of fluid-stimulated rock volumes". In: *GEOPHYSICS* 76.6, WC55–WC63. DOI: 10.1190/geo2010-0349.1.

Shelly, D. R., G. C. Beroza, and S. Ide (2007). "Non-volcanic tremor and low-frequency earthquake swarms". In: *Nature* 446.7133, pp. 305–307. DOI: 10.1038/nature05666.

Shelly, D. R., W. L. Ellsworth, and D. P. Hill (2016). "Fluid-faulting evolution in high definition: Connecting fault structure and frequency-magnitude variations during the 2014 Long Valley Caldera, California, earthquake swarm". In: *Journal of Geophysical Research: Solid Earth* 121.3, pp. 1776–1795. DOI: 10.1002/2015JB012719.

Slinkard, M., D. P. Schaff, N. Mikhailova, S. Heck, C. Young, and P. G. Richards (2014). "Multistation validation of waveform correlation techniques as applied to broad regional monitoring". In: *Bulletin of the Seismological Society of America* 104.6, pp. 2768–2781. DOI: 10.1785/0120140140.

Sokal, R. R. and C. D. Michener (1958). "A statistical method for evaluating systematic relationships". In: *The University of Kansas Science Bulletin* 38, part 2.22, pp. 1409–1438. URL: https://archive.org/details/cbarchive_33927_astatisticalmethodforevaluatin1902.

Tinti, S. and F. Mulargia (1987). "Confidence intervals of b values for grouped magnitudes". In: *Bulletin of the Seismological Society of America* 77.6, pp. 2125–2134. URL: http://www.bssaonline.org/content/77/6/2125.

Trutnevyte, E. and S. Wiemer (2017). "Tailor-made risk governance for induced seismicity of geothermal energy projects". In: *Geothermics* 65, pp. 295–312. DOI: `10.1016/j.geothermics.2016.10.006`.

Turcotte, D. L. (1997). "Seismicity and Tectonics". In: *Fractals and Chaos in Geology and Geophysics*. 2nd ed. Cambridge University Press. Chap. 4, pp. 56–80. DOI: `10.1017/CBO9781139174695.006`.

van der Walt, S., S. C. Colbert, and G. Varoquaux (2011). "The NumPy array: A structure for efficient numerical computation". In: *Computing in Science and Engineering* 13.2, pp. 22–30. DOI: `10.1109/MCSE.2011.37`.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. 1st ed. Springer, New York. DOI: `10.1007/978-1-4757-2440-0`.

Vasterling, M., U. Wegler, J. Becker, A. Brüstle, and M. Bischoff (2017). "Real-time envelope cross-correlation detector: Application to induced seismicity in the Insheim and Landau deep geothermal reservoirs". In: *Journal of Seismology* 21.1, pp. 193–208. DOI: `10.1007/s10950-016-9597-1`.

Wessel, P., W. H. F. Smith, R. Scharroo, J. Luis, and F. Wobbe (2013). "Generic Mapping Tools: Improved version released". In: *Eos, Transactions American Geophysical Union* 94.45, pp. 409–410. DOI: `10.1002/2013EO450001`.

Wiemer, S. and M. Wyss (2000). "Minimum magnitude of completeness in earthquake catalogs: Examples from Alaska, the Western United States, and Japan". In: *Bulletin of the Seismological Society of America* 90.4, pp. 859–869. DOI: `10.1785/0119990114`.

Wiemer, S., M. Herrmann, D. Karvounis, T. Tormann, T. Kraft, and N. Deichmann (2019). "Long-lasting effects of fluid injections into the crystalline basement: Lessons learned from the Basel deep heat mining project".

Wiemer, S., T. Tormann, M. Herrmann, D. Karvounis, T. Kraft, and M. Marti (2017a). *Induzierte Erdbeben im Nachgang des eingestellten Geothermieprojekts in Basel*. Report. (German). Schweizerischer Erdbebendienst (SED) an der ETH Zürich. DOI: `10.3929/ethz-b-000254199`.

Wiemer, S., E. Trutnevyte, T. Kraft, and P. Roth (2017b). *"Good Practice" guide for managing induced seismicity in deep geothermal energy projects in Switzerland*. Report. Swiss Seismological Service at ETH Zürich. DOI: `10.3929/ethz-b-000254161`.

Wössner, J. and S. Wiemer (2005). "Assessing the quality of earthquake catalogues: Estimating the magnitude of completeness and its uncertainty". In: *Bulletin of the Seismological Society of America* 95.2, pp. 684–698. DOI: `10.1785/0120040007`.