



**HAL**  
open science

# Monitoring induced microseismicity in an urban context using very small seismic arrays: The case study of the Vendenheim EGS project

Rémi Fiori, Jérôme Vergne, Jean Schmittbuhl, Dimitri Zigone

## ► To cite this version:

Rémi Fiori, Jérôme Vergne, Jean Schmittbuhl, Dimitri Zigone. Monitoring induced microseismicity in an urban context using very small seismic arrays: The case study of the Vendenheim EGS project. *Geophysics*, 2023, 88 (5), pp.WB71-WB87. 10.1190/GEO2022-0620.1 . hal-04307951

**HAL Id: hal-04307951**

**<https://hal.science/hal-04307951>**

Submitted on 26 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Monitoring induced microseismicity in an urban context using very small seismic arrays: The case study of the Vendenheim EGS project

Rémi Fiori<sup>1</sup>, Jérôme Vergne<sup>2</sup>, Jean Schmittbuhl<sup>2</sup>, and Dimitri Zigone<sup>2</sup>

## ABSTRACT

Monitoring the seismicity induced by fluid injections in deep geothermal reservoirs is often limited by the significant anthropogenic ambient noise level inherent to most enhanced geothermal system (EGS) projects in urban contexts. We report on the performance of a monitoring network made of three small aperture (72 m) seismic arrays composed of 21 3D nodes. We test the setup for four months (from 12 December 2020 to 8 April 2021) of the Strasbourg sequence of induced earthquakes (2019–2022) related to the EGS Georhin project (Vendenheim, France). The deployment starts a few days after the ML<sub>v</sub> = 3.6 induced earthquake of 4 December 2020 and covers the early shut-in period of the wells. We use a beamforming technique to characterize the main noise sources, which consist of slow

apparent velocities of surface waves emitted from mobile anthropogenic sources (motorway and railway traffic). We detect events with a phase-weighted stacking method, which is efficient when wavefronts illuminate the arrays with a high apparent velocity. Earthquakes associated with these detections are located using a matched field processing technique. The obtained catalog includes 216 seismic events, which represent four times more events than the reference catalog from Le Bureau Central Sismologique Français–Renass (the national academic agency in charge of seismicity monitoring in France), and a reduction of the completeness magnitude from 0.3 to –0.5. The clustering of the seismicity is analyzed using waveform correlation. The enriched catalog reveals intermittent seismic activity during the slow and continuous decrease in fluid pressure after shut-in.

## INTRODUCTION

The occurrence of induced earthquakes is one of the biggest issues in the development of deep geothermal energy (Wiemer et al., 2009; Zang et al., 2014; Foulger et al., 2018; Ellsworth et al., 2019; Kang et al., 2019; Yeo et al., 2020). The mechanisms behind the onset and the evolution of induced seismicity still must be better understood. For that purpose, the appeal of low-magnitude microseismicity monitoring is twofold. On the one hand, the study of very small induced earthquakes can significantly improve the knowledge of the structure of the reservoir and the physical processes occurring in response to fluid injection and circulation. On the other hand, traffic light systems generally assume that major seismic events are preceded by an increasing number of smaller ones (Mignan

et al., 2017). Accurate monitoring and characterization of induced seismicity in the environment of a geothermal project is therefore a key element in the monitoring of the reservoir evolution and the risk hazard assessments.

One major difficulty in microseismicity monitoring is the level of ambient seismic noise, which limits the detection of small earthquakes. This is particularly important in urban environments where human activities generate significant seismic noise in the frequency band traditionally used for the detection of local induced microseismicity. It has been shown that using borehole seismic stations (Eisner et al., 2010), which are less sensitive to surface noise and usually located closer to the events, drastically improves monitoring capabilities and helps to control the seismic rate during fluid injections (Kwiatek

Manuscript received by the Editor 16 October 2022; revised manuscript received 24 January 2023; published ahead of production 3 May 2023; published online 22 June 2023.

<sup>1</sup>Strasbourg University/CNRS, EOST/ITES, Strasbourg, France and Agence de l'environnement et de la Maîtrise de l'Energie, Angers Cedex 01 France. E-mail: r.fiori@unistra.fr (corresponding author).

<sup>2</sup>Strasbourg University/CNRS, EOST/ITES, Strasbourg, France. E-mail: jerome.vergne@unistra.fr; jean.schmittbuhl@unistra.fr; zigone@unistra.fr.

© 2023 The Authors. Published by the Society of Exploration Geophysicists. All article content, except where otherwise noted (including republished material), is licensed under a Creative Commons Attribution 4.0 International License (CC BY). See <https://creativecommons.org/licenses/by/4.0/>. Distribution or reproduction of this work in whole or in part commercially or noncommercially requires full attribution of the original publication, including its digital object identifier (DOI).

et al., 2019). However, such instrumentation is costly and often not deployed when not imposed by regulations. Therefore, it appears necessary to optimize and adapt the geometry of surface seismic networks to the high-noise conditions encountered in urban environments.

A surface seismic array consists of a network of sensors whose interdistances are sufficiently small to allow unaliased recording of the seismic wavefield in a specific frequency band (Capon, 1967; Rost and Thomas, 2009; Chmiel et al., 2019). Seismic arrays are classically used for the monitoring of natural earthquakes or artificial sources at local, regional, or teleseismic distances (Schweitzer et al., 2012). They allow computing the apparent slowness and back azimuth of various seismic phases and enhance their signal-to-noise ratio (S/N), by acting like a spatial filter.

Here, we test the ability to use three patch arrays having a small aperture to improve the detection and location of induced earthquakes in an urban area in comparison to the regionally distributed network used by Bureau Central Sismologique Français–Réseau national de surveillance sismique (BCSF–Renns), the national academic agency in charge of the seismicity monitoring in France. We illustrate this approach in the case of the Georhin enhanced geothermal system (EGS) project in Vendenheim (France), which has been associated with a long seismic sequence (from 2019 to 2022), with a dozen events felt by the population (Schmittbuhl et al., 2021) that led to the cessation of the project by the regional authorities in December 2020.

In this paper, we first introduce the case study of the Georhin EGS project and the related induced seismic sequence. Then, we describe the main characteristics of the seismic arrays temporarily deployed for four months to study the seismicity of the area. We quantify the ambient seismic noise level and compare it with the spectra of induced seismic events to assess the detection capability of individual stations and determine the frequency range of interest for the monitoring of local microseismicity. We characterize the main noise sources in these frequencies using beamforming techniques on sliding time windows. Then, we quantify the effect of the linear and phase-weighted stacking (PWS) (Schimmel and Paulssen, 1997) of the array seismic traces on the detection level of small-magnitude events. We locate most of these events with an array processing approach based on the matched field processing (MFP) technique (e.g., Baggeroer et al., 1993; Corciulo et al., 2012; Chmiel et al., 2019). The enriched catalog obtained with this approach is analyzed and interpreted in time — in terms of event density evolution, Gutenberg–Richter parameters, and subclusters formation — and space. Finally, we discuss the limitations of the proposed method in terms of detection and location capabilities and discuss the potential and challenge of working with S waves in addition to P waves.

## CASE STUDY PRESENTATION AND CONTEXT

### The Vendenheim EGS project

After the success of the Soultz-sous-Forêts (Bas-Rhin, France) European EGS project, which started at the end of the 1980s (Cuenot et al., 2008), several other sites in the same region have been investigated for their geothermal potential. Among them, there is our case study, which corresponds to the Vendenheim EGS project (Georhin) located in Vendenheim (France), a town belonging to the Strasbourg Eurometropole, 10 km to the north of the Strasbourg city center. The project consists of a deep geothermal doublet that targets a fault zone at approximately 5 km depth (oriented N015 E and dipping 70° westward), at the transition from the

sediment cover to the granitic basement. There was almost no seismicity in a radius of 10 km around the wells for the past 40 years before the project. The development of this project produced a sequence of largely felt-induced earthquakes, which led to its arrest (see Schmittbuhl et al., 2021).

This sequence began in March 2018 with a few earthquakes close to the GT1 injection well (Figure 1). In May 2019, seismicity close to the GT2 production well also appeared. This seismicity was relatively small, below magnitude 2 except for one event. In November 2019, a cluster of events was observed approximately 5 km south of the wells with a peak magnitude MLv3.0 earthquake on 12 November 2019 (MLv: a local magnitude computed from the maximum amplitude of the vertical component). The activity in this southern cluster continued for several months. Following new activities in the doublet from the end of August 2020, a large sequence of earthquakes developed including an MLv3.6 on 4 December 2020 largely felt by the population and 22 events above magnitude 2. Because of this crisis, the operator stopped all activities on site. A progressive bleed-off and a shut-in were performed from mid-December 2020 to 2 January 2021, which induced a decrease in fluid pressure over more than eight months (Terrier et al., 2022). The biggest earthquake of the sequence (MLv3.9) occurred on 26 June 2021. Since then, seismic activity has decreased substantially.

The earthquake catalog shown in Figure 1 has been produced by the BCSF–Renns based on waveforms from a composite network of 76 local and regional surface stations (blue triangles in Figure 1, from Schmittbuhl et al., 2021). Part of this network has been deployed specifically to monitor this induced seismicity, whereas the other part consists of permanent stations of the national seismic network (Réseau Sismologique Français). This catalog has been obtained by automatic detection using a SeisComp3 software pipeline with optimized short-term average/long-term average (STA/LTA) and detection thresholds for each station. It has been enriched with events detected by template matching. Locations were performed with the *scoutloc* module of SeisComp3 using the 1D velocity model traditionally used in northeastern France (Rothe and Peterschmitt, 1950), which is characterized by constant velocities  $V_p = 4.9$  km/s and  $V_s = 3.4$  km/s for depths lower than 20 km. The magnitudes have been computed within SeisComp3 (see the documentation: Helmholtz-Centre Potsdam–GFZ German Research Centre for Geosciences and *gempa GmbH*, 2008). Most of them are MLv magnitudes, except for the smallest events for which a duration magnitude was used. All the events were systematically checked by human analysts. The whole catalog obtained with this method has a magnitude of completeness of 1.1 (Schmittbuhl et al., 2021). From March 2018 to August 2022, it contains 576 events labeled as induced earthquakes.

### Instrumental setup

To temporarily improve the monitoring of this sequence, we tested the deployment of patch arrays with a small aperture and very small spatial sampling — compared with what is usually implemented for microseismicity monitoring (Sick and Joswig, 2017; Chmiel et al., 2019), as shown in Figure 2a. Each array is composed of 21 nodes, following three concentric circles 8, 24, and 72 m in diameter.

The design of an array dictates the range of wavelengths that can be analyzed (Capon, 1973; Aki and Richards, 1980; Bullen and Bolt, 1985; Buttkus, 2000; Kennett, 2002). For wavelengths much higher than the aperture, the whole array can be considered as a single station, leading to resolution limitation. Conversely, spatial

aliasing happens for wavelengths smaller than half the minimal interstation distance. Hence, the wavenumbers properly sampled are estimated between  $k_{\text{resolution}}$  and  $k_{\text{aliasing}}$  defined as

$$k_{\text{resolution}} = \frac{2\pi}{L}, \quad (1)$$

$$k_{\text{aliasing}} = \frac{2\pi}{2\delta x}, \quad (2)$$

with  $L$  being the aperture of the array and  $\delta x$  being the minimal interstation distance within the array.

To determine in more detail the frequencies within which we can work without meeting resolution or aliasing limitations, we compute the whole array transfer functions (Halldorsson et al., 2009; Schweitzer et al., 2011). When stacking all the sensors of an array after aligning their traces with respect to their relative position and to a plane-wave slowness vector, we obtain a beam whose energy depends on the plane-wave wavenumber. The array transfer func-

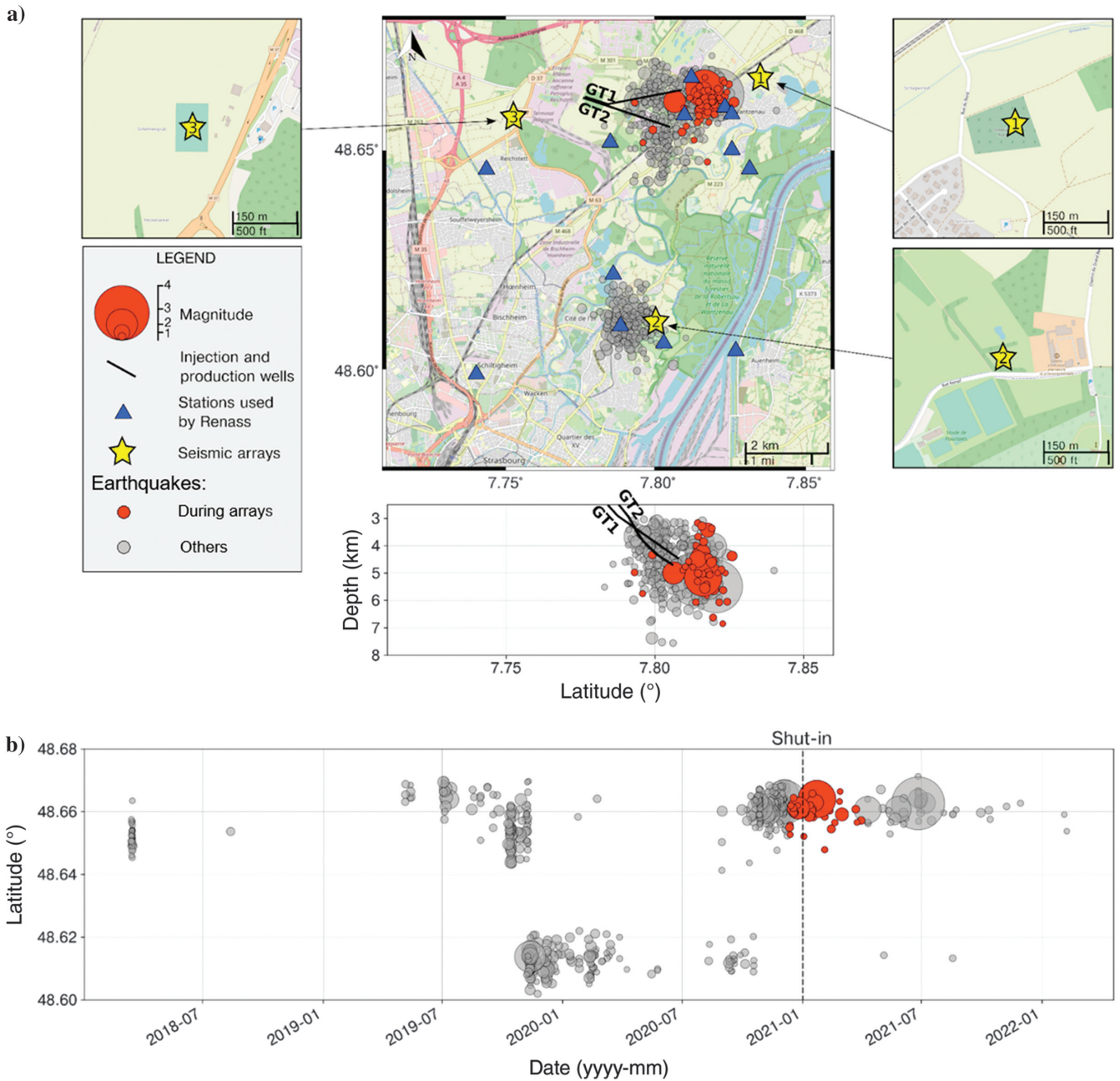


Figure 1. (a) Map of the Strasbourg Eurometropole area, with seismic array locations (the yellow stars), in the vicinity of the two seismic clusters recorded by the BCSF-Renass with the existing network (the blue triangles) operated since mid-2018. The two thick black lines are the production well (GT1) and injection well (GT2). The depth cross section shown under the map is for the seismicity of the northern cluster only. (b) Timeline of the whole seismic sequence until 1 July 2022 (gray) and in particular during the four months of acquisition with seismic arrays (red). The gray vertical dashed line represents the start of the progressive shut-in.

tion is the comparison of the maximal beam energy for each wavenumber considering the array geometry. Figure 2b shows the generic array transfer function computed using the Obspy method (array\_transff\_wavenumber function), for wavenumbers up to 2.4 rad/m with a step of 0.001 rad/m. In Figure 2, we discern a discoid central spot reaching 0 dB attenuation, hereafter called the “main lobe” (responsible for the resolution limitation), and some other decentered lobes reaching at least  $-3$  dB (the orange color in Figure 2 responsible for the aliasing), called “side lobes.” Wathelet et al. (2005) suggest considering the major aliasing peaks as the maximum usable wavenumber and considering half the width of the central peak as the minimum usable wavenumber. The  $-3$  dB value corresponds to about half the maximum value of the transfer function, which is the usual threshold used to delimit the peaks’ boundaries (e.g., Wathelet et al., 2008).

We find  $k_{\text{resolution}} = 0.066$  rad/m when taking the average radius of the circle formed by  $-3$  dB isoline in the main lobe and  $k_{\text{aliasing}} = 0.783$  rad/m when taking half the minimum value to the next  $-3$  dB isoline (reached for side lobes, the red circle in Figure 2b). These measured values are coherent with the theory of aliasing and resolution (equations 1 and 2). Also, unlike cross-shaped arrays that have a clear preferential orientation relative to the source, this kind of concentric circles array is less biased when analyzing front waves coming from different directions. The main lobe is made narrow by the array aperture and sharp

by the isotropic distribution of the sensors. The side lobes are reduced due to the variety of spatial intervals between sensors.

Using these aliasing and resolution bounds, we can transpose from the array response in the wavenumber domain to a frequency-velocity representation, using the relation

$$k = \frac{2\pi \cdot f}{V_{\text{app}}}, \quad (3)$$

where  $k$  is the wavenumber,  $f$  is the frequency, and  $V_{\text{app}}$  is the apparent velocity of the wave crossing the array. Figure 2c shows the aliasing and resolution limitations of the arrays, based on the values of  $k_{\text{resolution}}$  and  $k_{\text{aliasing}}$  previously given.

For our study, we deployed three such small seismic arrays (Figure 1a) three to six days after the 4 December 2020 magnitude MLv3.6 earthquake and during the shut-in period. The installation of such arrays requires free fields more than 72 m long in all directions. The three sites found were an unused area in the La Wantzenau cemetery (array 1), arable land in La Robertsau (array 2), and a disused soccer field in Reichstett (array 3). Together, they form a triangle encompassing the northern cluster whereas the southern cluster is more off-centered. These three arrays were operating for four months.

For each station of the arrays, we used velocity traces from three-component IGU-16HR SmartSolo nodes with a corner frequency of

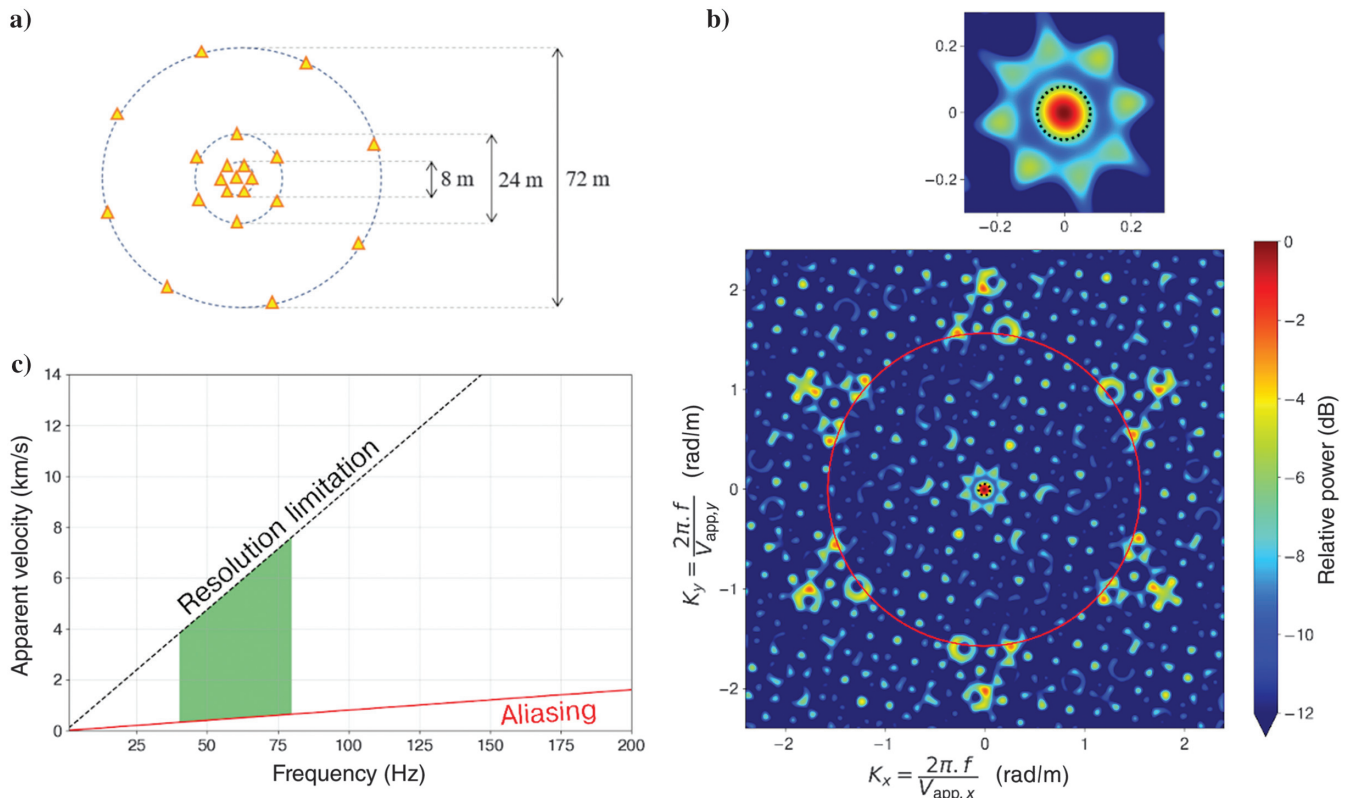


Figure 2. (a) Sensor configuration (the yellow triangles) within each of the three arrays and (b) theoretical array transfer function normalized with its maximum, in the wavenumber domain, for this geometry. The insert shown at the top is a magnification of the central part of the lower figure. The dashed black circle delineates the resolution limit and the red circle delineates the aliasing limit. (c) Resolution and aliasing boundaries in the frequency-velocity domain based on the values of  $k_{\text{resolution}}$  and  $k_{\text{aliasing}}$ . The green area corresponds to the 40–80 Hz interval used in this study.

5 Hz and set the sampling rate to 1 kHz. Sensors are fully buried (a few centimeters under the ground level) and covered by soil. They are north-oriented using a manual compass and verticalized by bubbling. Although the stations have an integrated GPS module that saves their coordinates, we measured their location more precisely using a differential GPS setup, allowing a horizontal precision of 0.04 m and a vertical precision of 0.07 m. This precision is all the more important as the stations are particularly close to each other. The internal clocks of those instruments are not synchronized with others. However, typical errors for atomic clocks in GPS satellites are within a few nanoseconds (McNeff, 2002), which is not significant considering the time delays we want to observe, even for the closest stations.

A diagram of the data completeness per array can be found in the supplemental material S1. One can notice short gaps in the data every 30 days for approximately 1 h, corresponding to each time we swapped the stations after the 30 days battery life of those instruments. The last swap was done too late, so all the nodes of arrays 1 and 3 were out of battery for some days. We also had a technical issue with a node of array 2, which did not initialize properly and recovered no data until the end of the acquisition. Except for those gaps, the instruments performed well and we recovered approximately 98% of the expected data set. In this study, we only use the vertical components, although the benefits of using the horizontal components will be discussed in the last section.

### Ambient noise characterization

The ambient seismic noise refers to all the signals continuously present in the environment not due to seismic events. We quantify the ambient seismic noise level by computing smoothed power spectral densities (PSD) using the Obspy module (obsypy.signal.PPSD) based on the algorithm of McNamara and Buland (2004). We compute PSD on 1 h time windows, making the classical assumption that seismic events represent a nonsignificant proportion of the signal. In our case, by estimating the duration of each earthquake signal from its detection until the return to the noise level, seismic activity represents at most a few dozen seconds out of 1 h of data during the period of the highest seismic rate, which is still less than 1%. Figure 3 shows the median (the blue line) and the 10th and 90th percentiles (the blue area) of the whole PSD set for the central node of array 1 (see supplemental material S2a and S2b for the two other arrays). At this site, we observe an important noise level between 1 and 10 Hz, approaching or even exceeding the new high-noise model (NHNM) (Peterson, 1993; McNamara and Buland, 2004). The median PSD reaches a peak of  $-105$  dB at 10 Hz, then decreases until  $-115$  dB at 40 Hz, and then reaches a second peak of  $-112$  dB at 200 Hz.

To evaluate the detection potential of this single node, we compare its noise level to PSD computed on earthquakes depending on their magnitude (the dashed green lines in Figure 3). Each line corresponds to a PSD representative of earthquakes of magnitude  $M = 0.5$ , 1.5, or 2.5, respectively. For each magnitude, we constitute a group by gathering all the events of magnitude  $M \pm 0.1$ . For each event, we extract a time window of 1 s for  $M = 0.5$ , 2 s for  $M = 1.5$ , and 3 s for  $M = 2.5$  around the P wave and compute its PSD as we did for noise. The final PSD is the median of all the PSD computed by the group.

We consider that a minimum S/N of 3 (i.e., approximately +5 dB) is needed to detect an earthquake, even if this threshold is arbitrary and highly depends on the context of data acquisition (Desimoni and Brunetti, 2015). Then, a single station should detect earth-

quakes of magnitude 1.5 or higher in any frequency range below 100 Hz. However, the PSD of magnitude 0.5 earthquakes is not sufficiently above the ambient noise PSD power for frequencies below 20 Hz and above 70 Hz at least 10% of the time, which is shown by the 90th percentile PSD. This illustrates that magnitude 0.5 is close to the limit of detection capability for a single station, even when recording as close as 1.5 km from the barycenter of the northern cluster epicenters.

To help determine the different origins of the recorded noise in the frequency band of local earthquake signals, we look at the PSD time series. Figure 4a shows a pseudospectrogram constructed with all the 1 h long PSD over four months of acquisition. To better quantify the different periodicity of the noise level, we compute the module of the fast Fourier transform (FFT) for each frequency of the spectrogram over the whole period (Figure 4b). We distinguish different behaviors depending on the frequency range. Between 0.1 and 1 Hz, we do not observe any noticeable periodicities, and the noise is mainly attributed to the oceanic secondary microseismic peak (McNamara and Buland, 2004). Between 1 and 60 Hz, we observe a clear and well-known daily (24 h) and weekly (168 h) periodicity with higher noise during the day compared with the night and during weekdays compared with weekends. This is a typical signature of anthropogenic sources and in particular of traffic footprint (Riahi and Gertstoft, 2015). Above 60 Hz, the temporal evolution of the noise level is more complex with daily variations which could be due to human activity but also bursts of high noise that may be due to natural phenomena such as meteorological ones (Dean, 2017). To better characterize the main noise sources, we apply beamforming techniques to determine the time evolution of the

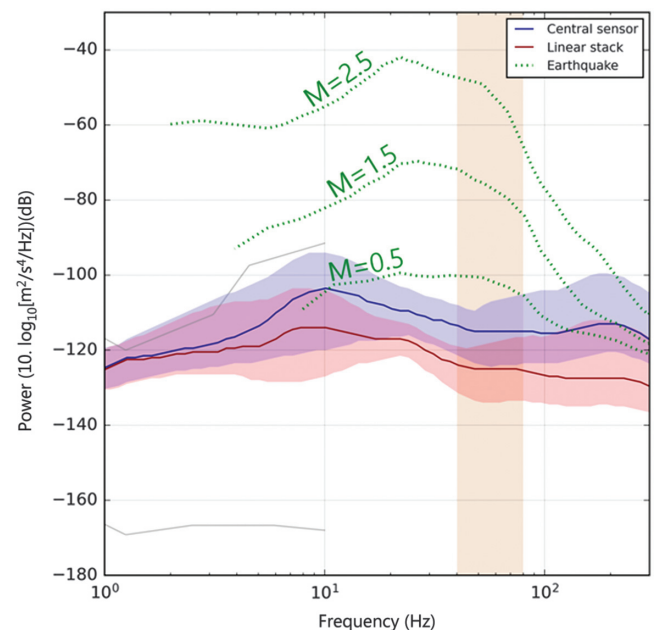


Figure 3. Comparison of the noise level from data at the central station (blue) and from linearly stacked data (red) of the northeast array. The solid line stands for the median of the set of computed PSD whereas the transparent areas are delimited by the 10th and 90th percentile. The dashed green lines are the PSD of local earthquakes by magnitude. The gray solid lines correspond to the new low-noise model (NLNM) and NHNM (Peterson, 1993; McNamara and Buland, 2004). The light brown rectangle corresponds to the 40–80 Hz frequency range.

azimuth and apparent velocity of the most coherent wavefields (e.g., Rost and Thomas, 2009). We use a classical Bartlett beamformer algorithm on 1 s time windows sliding along the continuous data. For each 1 s time window, we apply a Butterworth band-pass filter between 5 and 80 Hz. Then, we determine the back azimuth/apparent slowness couple for which the beamformer finds the maximum energy, with a code based on the Obspy module (obspy.signal.array\_analysis.array\_processing). We modify the code to look for the best couple by a grid search for back azimuth from  $0^\circ$  to  $360^\circ$  (step of  $1^\circ$ ) and apparent slowness from 0 to 6 s/km (step of 0.03 s/km).

Figure 5 shows some typical examples for different arrays and during different periods of the day. During the daytime (Figure 5a), we

observe clear patterns: the evolution of the back azimuth of the main noise sources with time is continuous on portions, and the corresponding apparent slowness is between 2 and 4 s/km, which corresponds to low-velocity surface waves. Those patterns can be attributed to moving vehicles on nearby roads around each of the three arrays. It should be noticed here that, according to Figure 2, the investigation of such high slowness can be affected by aliasing effects for the highest frequencies of the 5–80 Hz band, highlighting the limit of these arrays for noise characterization. Figure 5b shows another clear high-energy pattern in the back azimuth, also coupled with an apparent slowness between 2 and 3 s/km. The back azimuth is pointing to a railway and if we consider the back azimuth variation corresponding to a source moving along this railway, we estimate

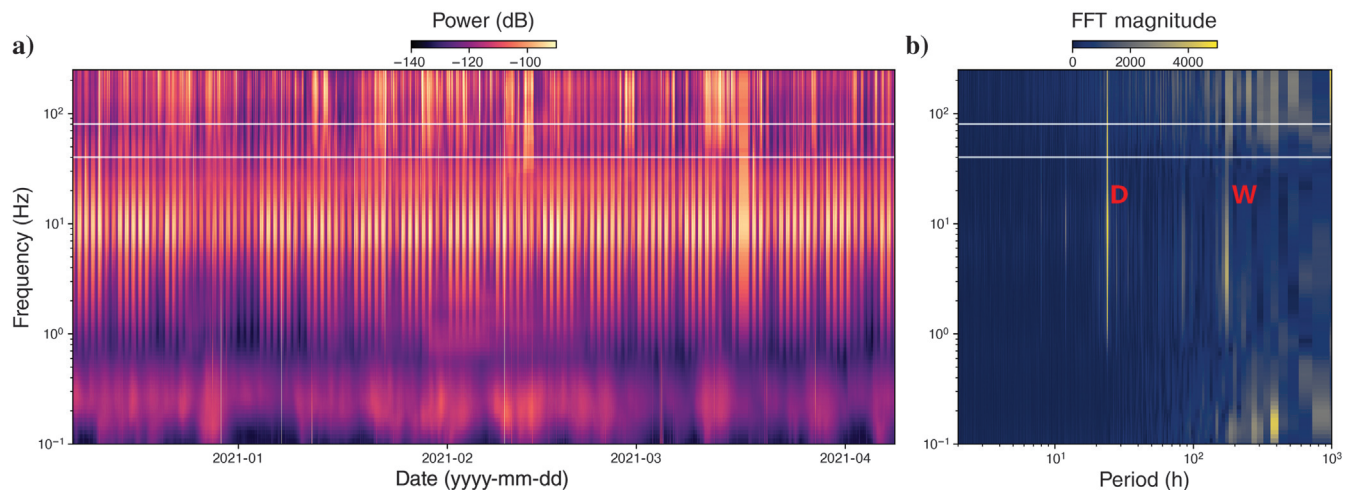


Figure 4. (a) Spectrograms of the PSD computed on vertical traces of the central station of array 1 over the four-month period of acquisition and (b) module of the FFT of the PSD time series for each frequency. The white lines delimit the 40–80 Hz frequency range. Here, D and W labels refer to the peaks at 24 and 168 h, respectively.

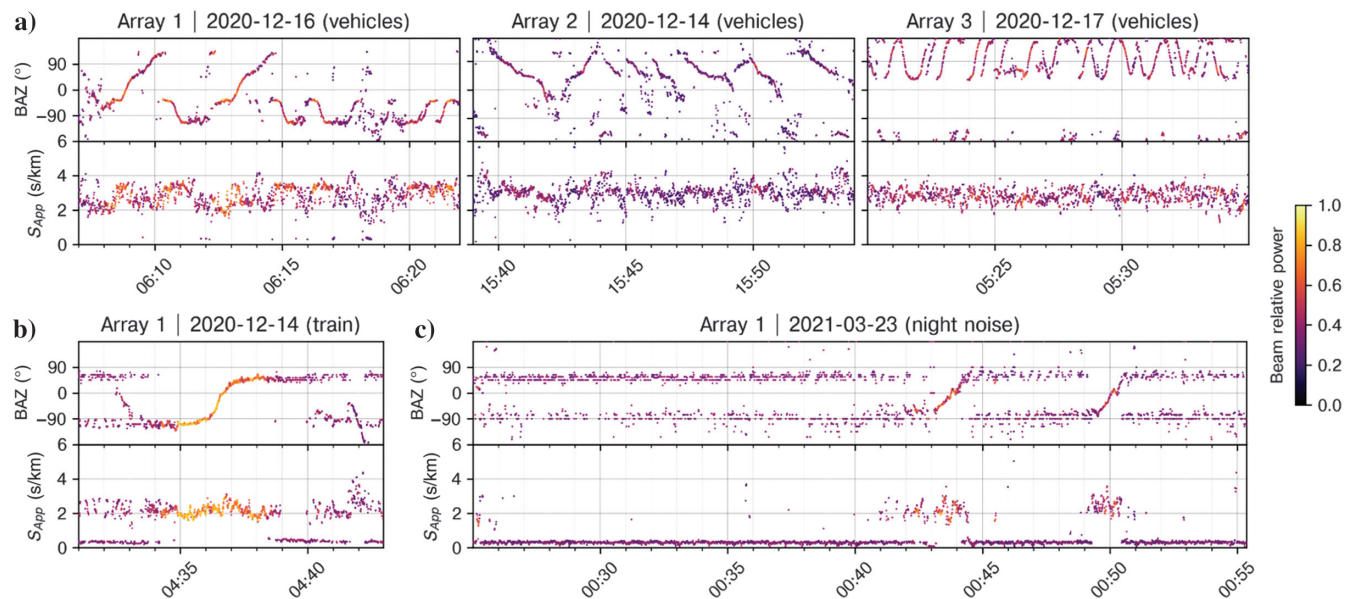


Figure 5. Evolution of the dominant back azimuth and apparent slowness with time, obtained by a beamforming technique on sliding time windows of 1 s after filtering between 5 and 80 Hz (the first line). Typical signature of (a) road traffic during the daytime, (b) a train passage, and (c) noise recorded at night.

the speed of the moving source to be approximately 90 km/h, which is the speed regulation of trains on this line. Moreover, the date and hours of the occurrence of these patterns match the schedule of the train company, confirming our capability to track such trains with our arrays. We also notice that the apparent slowness drops significantly from 2 to 3 s/km during the passage of vehicles (cars, trucks, and trains) to values of approximately 0.3 s/km for periods without the passage of vehicles, therefore approaching typical values measured when an earthquake occurs (supplemental material S3a). This high apparent velocity noise also can be observed more often during the night times (Figure 5c) and appears to originate from two distinct back azimuths at approximately 50° and -90°. By computing beamforming on the same time windows for frequency ranges 5–20 and 40–80 Hz, we found that this high apparent velocity noise is mostly observed for high frequencies (see supplemental material S3b and S3c for the beamforming in those frequency bands). Even if we consider that such noise is propagating at the highest mechanical wave velocity in this medium (e.g., P-wave velocity), such values of apparent slowness correspond to waves coming with a subvertical incidence. Several phenomena such as sources further away or reflected energy reaching the sensors with a nonhorizontal incidence could potentially explain those characteristics. Otherwise, this could be the marker of an aliased signal, as mentioned previously. However, we have not investigated this further as the determination of the precise origin of this noise is outside the scope of this paper.

## CATALOG ENHANCEMENT METHOD

### Detection and association of small earthquakes

#### Linear stacking

As shown in Figure 5, seismic signals generated by anthropogenic sources (such as road traffic) are mostly composed of a surface wave with a small apparent velocity, at least during the daytime. Conversely, induced earthquakes occurring almost beneath an array will have a very high apparent velocity, and seismic signals will reach each sensor of an array more on phase. Therefore, a linear stacking of the seismograms — even with no relative time shifts of the traces — for all the nodes of a small array will enhance waveforms of induced seismic events and reduce the amplitude of surface “noise” sources. We quantify the impact of linear stacking on the array’s sensitivity by computing PSD on a composite trace obtained after stacking all the traces of an array (the red line in Figure 3). It appears that such stacking reduces the noise level by more than 10 dB, especially for frequencies above 20 Hz. As small induced earthquakes still radiate much their energy in these high frequencies, this illustrates that the use of such arrays should ease the detection of very small induced earthquakes at the local scale. Also taking into account that the higher the frequencies investigated, the higher the resolution limit, we chose to work into the 40–80 Hz frequency band during the next processing steps.

#### Phase-weighted stacking

Other stacking techniques allow us to further increase the S/N of coherent signals such as earthquakes. In particular, we used PWS (Schimmel and Paulssen, 1997; Thurber et al., 2014), which consists of weighting the traces with a dynamic factor that evaluates the phase coherency of all the nodes for each time sample  $s_j(t)$  of the

$j = 1N(N = 21)$  individual traces. The summation consists of combining all the traces  $s_j(t)$  into

$$g(t) = \frac{1}{N} \sum_{j=1}^{N=21} s_j(t) \left| \frac{1}{N} \sum_{k=1}^{N=21} \exp[i\Phi_k(t)] \right|^\nu = \frac{1}{N} \sum_{j=1}^{N=21} s_j(t) \cdot C(t), \quad (4)$$

where  $\Phi_k(t)$  is the instantaneous phase (Bracewell, 1965) of the analytical signal of the trace  $s_k(t)$  evaluated at time  $t$ . Compared with a linear stack, PWS puts more weight on the phase coherence  $C(t)$ , quantified by the module of the averaged sum of the normalized analytical signals (between zero and one). The sharpness of this weighting is controlled by the power  $\nu$ ,  $\nu = 0$  corresponding to the linear stack case. This stacking method is especially appropriate to our case study where the arrays are deployed quasivertically above the sources. In the first approximation, because of a very high apparent velocity, the incoming wavefront is very coherent in phase at the scale of one array, without the need for prior trace delaying based on source location considerations.

Practically, we first filter the traces between 40 and 80 Hz as discussed previously, with a zero-phase Butterworth band-pass filter of order four. Then, we apply the PWS process to the continuous data with  $\nu = 3$ . Following a series of tests, we found that this value is sufficiently high to enhance the signal without distorting the waveforms too much, allowing us to use classical detection algorithms and visually check the earthquake’s signal.

Figure 6 illustrates the improvement brought by the phase-weighted stack compared with a standard linear stack from the example of a magnitude -0.4 earthquake, not present in the BCSF-Renass catalog. With linear stacking, the P-wave’s signals have an S/N of 5, 1.5, and 4 on, respectively, arrays 1, 2, and 3. In comparison, PWS improves the S/N to 102, 45, and 95, easing the detections of the event. Nevertheless, we notice that, for array 2, the signal considered as the P wave for this event is surrounded by other coherent signals, sometimes with a similar shape. These other phases may correspond either to other seismic events — but not observed on the other two arrays — or, more probably, to coherent noise with a high apparent velocity as previously discussed. Those artifacts can sometimes be difficult to distinguish from real induced seismic events.

#### Detection-association of induced earthquakes

For the extraction of induced events from the continuous traces obtained by PWS at each array, we apply an STA/LTA detection algorithm (0.1 s/15 s) considering a detection threshold of 15. To reduce the detection of artifacts due to coherent noise and the very low STA used, we keep a detection only if its PWS peak value is higher than 10, which has been determined empirically by looking at the typical PWS amplitude of signals considered ambient noise. The association of detections between the arrays is then performed by keeping all the detections with an interarray delay of fewer than 2 s, which is the maximum delay time for a horizontal P-wave propagation between two arrays considering the local velocity model. In the case where several detections occur in the same 2 s window at several arrays, we only keep the detection with the biggest amplitude. For this reason, this basic association approach is not adapted for very intense swarms.



## Magnitude estimation

We estimate the  $M_L$  of the detected earthquakes following the method implemented in SeisComp3 and used by BCSF-Renns:

$$M_L = \log_{10}(A) - \log_{10}(A_0), \quad (5)$$

where  $A$  is the peak amplitude of the displacement in millimeters measured on a simulated Wood-Anderson seismometer and  $A_0$  is

an empirical calibration function, which depends on the epicentral distance (for details, see [Helmholtz-Centre Potsdam–GFZ German Research Centre for Geosciences and gempa GmbH, 2008](#)). Instead of using the peak amplitude of individual stations, we use the peak values measured on the PWS traces. Figure 7a shows the relation between the PWS peak values and the linear stack peak values. Despite the theoretical nonlinearity of the PWS process, the maximum peak amplitude follows quite a linear relationship with linear stacking peak amplitudes especially for large

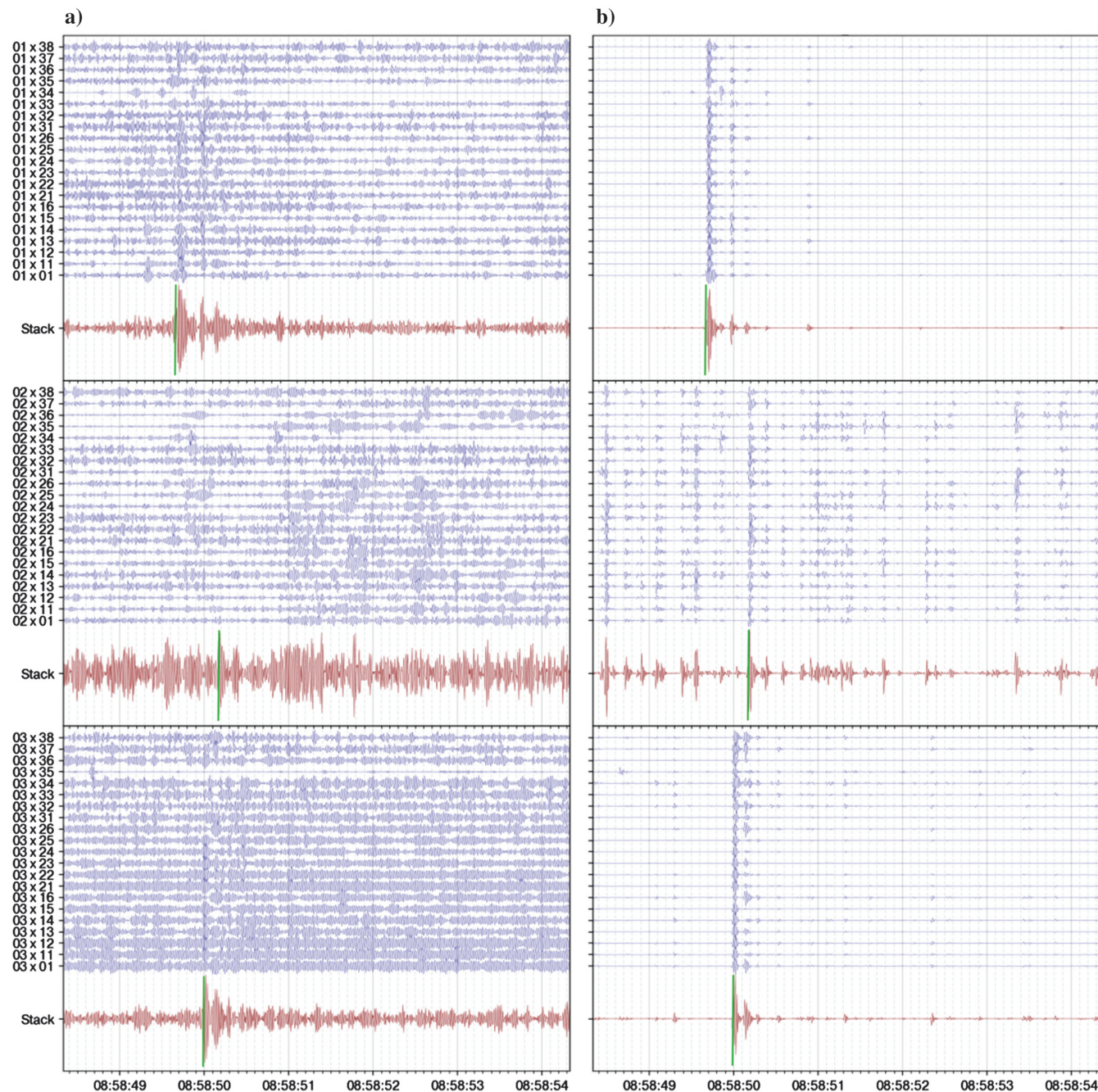


Figure 6. Illustration of the effect of the linear and phase-weighted stack, on a magnitude  $-0.4$  earthquake (13 December 2020), at the three arrays (1, 2, and 3 from top to bottom). The blue lines correspond to individual station traces and the red lines correspond to the summation of them; (a) traces band-pass filtered between 40 and 80 Hz and (b) traces filtered and multiplied by the coherency term  $C(t)$  defined in equation 4. Green solid vertical lines indicate the detection time.

amplitudes, which justifies our approach to use PWS amplitudes to compute the magnitudes. This is explained by the fact that all the earthquakes recorded at the same site have a similar value of  $C(t)$  at the peak time. Figure 7b shows a comparison of the magnitudes estimated by BCSF-Reness, from individual regional stations, and magnitudes estimated from the PWS peak amplitudes with the three arrays. The relation is linear for magnitudes above 0.7, except for the biggest event whose magnitude is underestimated by our approach. Magnitude estimations from the BCSF-Reness catalog below 0.7 seem to be overestimated. This is probably due to the fact that, for the smallest events, the measured peak amplitude on individual records is close to the noise level. Figure 7c shows the difference in magnitude estimation depending on whether we use the peak amplitude read on individual sensors or from the PWS traces, for each of the three arrays. Accordingly, we observed that magnitudes below approximately  $-0.4$  are gradually overestimated when using individual sensors. The slope of this overestimation is approximately  $-1$ , which means that all the earthquakes below  $-0.4$  have the same magnitude according to an estimation with only one sensor, whereas an estimation using PWS gives smaller magnitudes. This is in accordance with the hypothesis that the noise level is saturating the magnitude estimation. We note that saturation appears at slightly lower magnitudes than for the BCSF-Reness catalog (Figure 7b). This may be related to the fact that we are working in a higher frequency band than the one used by the BCSF-Reness (limited by the lower sampling rate of the regional stations they use), and according to Figure 4, the difference between the noise level and the typical spectrum of weak earthquakes is smaller at lower frequencies. For those reasons, we believe that our magnitude estimate based on the PWS amplitudes is likely to be less biased than the ones from the BCSF-Reness catalog.

## Events location

With our network geometry, a classical location algorithm based on phase picking would not provide precise estimations of hypocenter locations. Indeed, at the scale of the network composed of the three arrays, using all the stations is nearly equivalent to only using one for each array, which gives us a maximum of six phases (3 P and 3 S). Moreover, the picking is adding implicit uncertainty to the location, and the identification and picking of the S waves on phase-weighted or linear stacked traces are often challenging in the frequency band used (see the “Discussion” section about the identification of S waves at high frequencies).

Rather, we implement an MFP approach (Baggeroer et al., 1993; Corciulo et al., 2012; Umlauf and Korn, 2019; Umlauf et al., 2021). It consists of comparing the observed internodes’ arrival delays with the theoretical ones considering a given source location. On the one hand, we compute for each frequency a cross-spectral density matrix  $K$  from a time window around the P phase. This matrix basically contains the interstation delays based on cross-correlations of the seismograms. On the other hand, we compute the theoretical traveltimes at every node using a given local velocity model and create synthetic data  $E_i(x, y, z, \omega)$  related to the  $i$ th considered source at the location

$(x, y, z)$  for the angular frequency  $\omega$ . Note that the synthetic data  $E$  are actually computed in the frequency domain for simplicity and to save computation time. The match between the recorded ( $K$ ) and synthetic wavefield ( $E$ ) is then computed in the frequency domain through a traditional beamformer:

$$P_i(x, y, z, \omega) = [E_i^t(x, y, z, \omega)] \cdot [K] \cdot [E_i(x, y, z, \omega)], \quad (6)$$

which is proportional to the probability that the event occurred at that location. For each event, we compute  $K$  and  $E_i$  at each array independently based on a code proposed by Gouedard (2008) and modified by Zigone (2012). Within each array, the network is split into subnetworks, defined by determining a minimum coherency threshold for the locally recorded noise/event. This coherency selection is done in practice by keeping only the stations for which the maximum crosscorrelation value is higher than 0.3 with the central node of the subnetwork. The final location probability corresponds to the arithmetical average of the probabilities computed for each of the subnetworks in each of the three arrays and is weighted by the number of stations used per array. To find the most probable location, we explore a 3D grid of potential sources in the vicinity of the known seismic clusters.

We use a simple 1D local two-layer velocity model. Knowing that the stations are on a sedimentary basin lying on a granitic basement, we define the first layer as a linear gradient from 2.4 to 5.8 km/s down to the basement at a 4 km depth (according to the public database of the regional geologic model GeORG Projektteam, 2013), and the bottom one as a half-space with a constant velocity at 5.8 km/s. In such a two-layer model, the incidence angle needed to compute the travel-time from source to station cannot be found analytically, so we determine the travel path by a dichotomy search (as described by Jiang et al., 2016).

The time window of the extracted waveforms is 0.4 s long starting 0.05 s before the P-detection time given by the STA/LTA procedure on each array. To find the maximum location probability, we make two successive grid searches: first, in a  $15 \text{ km} \times 15 \text{ km} \times 12 \text{ km}$  grid centered on the barycenter of the three arrays and with a horizontal grid step of 0.25 km and a vertical grid step of 0.5 km and second, in a  $0.5 \text{ km} \times 0.5 \text{ km} \times 1 \text{ km}$  grid centered on the maximum found during the first iteration, with a horizontal step of 10 m and a vertical step of 100 m. We reject all events that fall on the boundaries of the first grid.

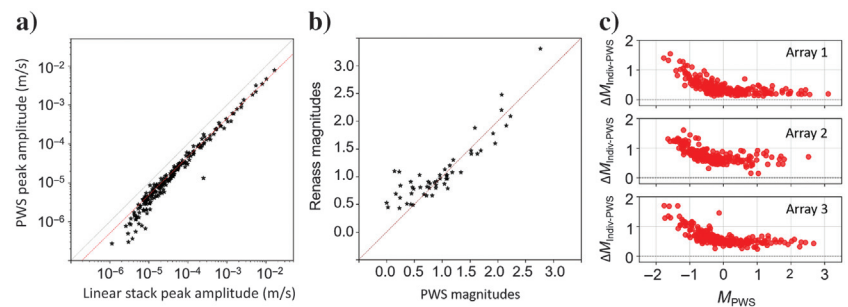


Figure 7. (a) PWS peak amplitudes versus linear stack peak amplitudes. The dashed red line is the best linear regression line; (b) BCSF-Reness magnitudes versus magnitude estimated with the PWS amplitude using equation 5; and (c) evolution of the difference between the magnitudes estimated from individual sensors and the magnitude from PWS peak amplitudes. The dashed black and red lines in, respectively, (a) and (b) represent a 1:1 relationship.

Looking at the individual probability maps obtained at each array, we see that the further away an array is from the source, the better it constrains the event depth. Conversely, having an array above the source helps to better constrain the horizontal location. However, with our network geometry, the depth of the earthquakes, expected at approximately 5 km based on the reference catalog, cannot be strongly constrained. Figure 8 shows examples of probability maps for two small earthquakes. The cross sections and horizontal view show a vertically elongated focal spot around the maximum location probability. The resolution is slightly improved for the larger earthquake of  $M = 1.1$  (Figure 8a) probably because, in a given array, the maximum value of the crosscorrelations of the traces between the most distant nodes is sufficiently high to be included in  $K$ . Our estimated hypocenter, using only the three arrays, is located at 100 m in the horizontal direction and 310 m in the vertical direction from the BCSF-Reness location. The second example (Figure 8b) is the smallest earthquake in the reference catalog ( $M = 0.3$ ). Its hypocenter is located at 760 m in the horizontal direction and 560 m in the vertical direction, with a lower location probability, and therefore lower confidence.

### Improving detections assuming the location of the events

In our first detection approach based on phase-weighted stacks, we assumed that the earthquakes were located vertically below the arrays such that the P wave is exactly in phase at all the nodes. If the hypocenter is actually shifted horizontally, this will generate a phase difference that will decrease the amplitude of the PWS, which will eventually be lower than our imposed STA/LTA threshold for the detection.

Assuming that our induced earthquakes form a relatively compact cluster, we can use the earthquake locations obtained by the approach presented to date to define the barycenter of this cluster. We then use this position to compute the relative traveltimes for all the nodes of an array and correct for the phase differences before applying the PWS. This will eventually allow us to detect and locate new events that occurred in a close area.

In our case, we located events mainly in the northern cluster and the southern cluster (see the “Results” section). Therefore, we apply this procedure twice using the barycenter of each of these clusters. A chart of the whole workflow is shown in Figure 9.

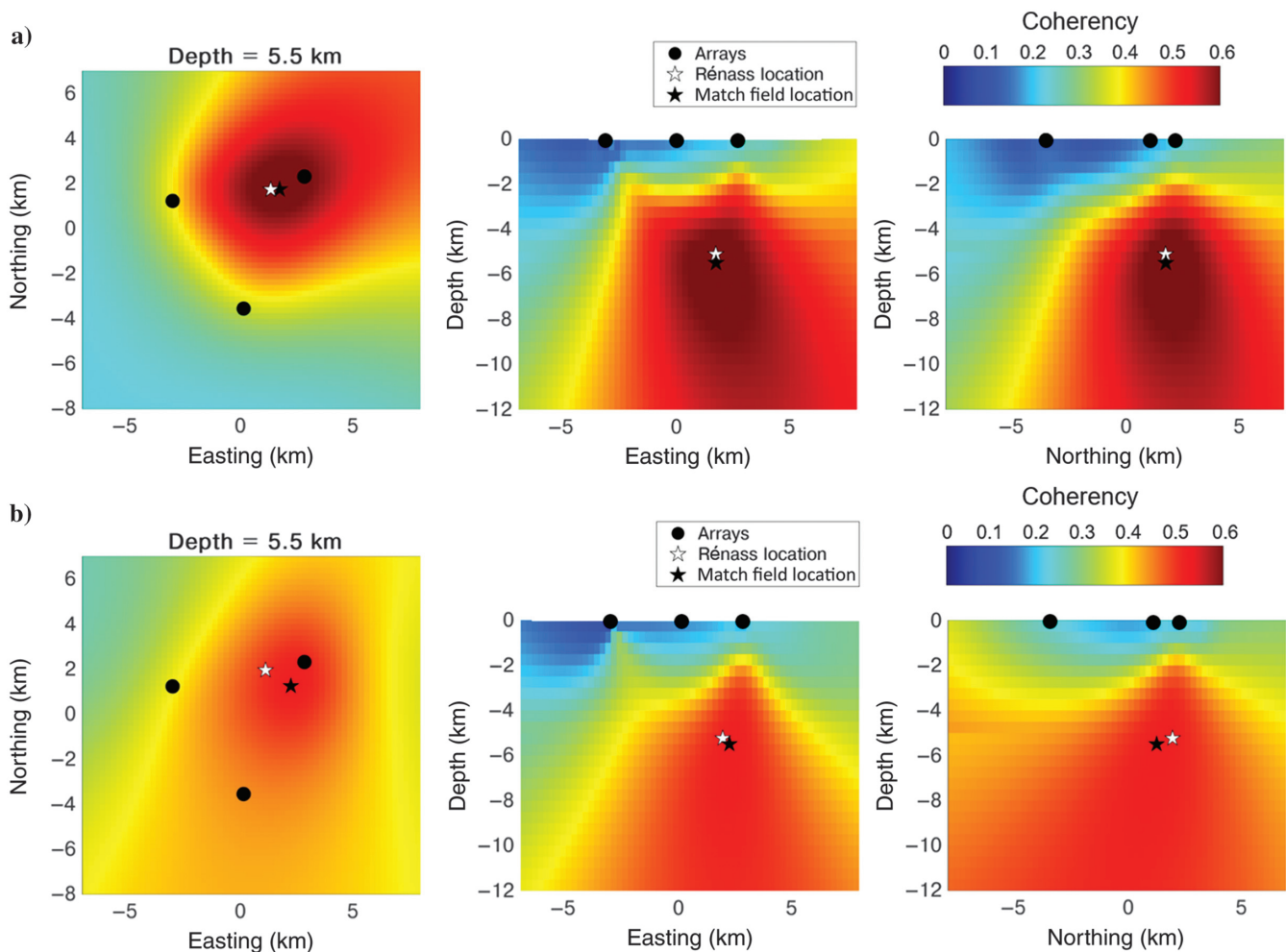


Figure 8. Location probability maps from the MFP for (a) a magnitude 1.1 earthquake (3 February 2021 02:01:16) and (b) the smallest earthquake recorded by the BCSF-Reness (24 December 2020 14:40:36  $M = 0.5$ ).

RESULTS

Detections and associations

We apply our detection procedure to the four-month period when the three arrays were in operation. During this period, the BCSF-Reness catalog contains 54 events labeled as induced and all are located in the northern cluster. With our approach based solely on the three arrays, we find 1793 events detected by any combination of two arrays and 252 events detected simultaneously by the three arrays. We manually checked most of the waveforms corresponding to these detections and found that many detections made at only two arrays have unusual waveforms compared with the known earthquakes of the reference catalog. Moreover, we observe that the number of detections at only two arrays (Figure 10a) is higher during the daytime, although the higher noise level should make it difficult to detect small earthquakes. This suggests that part of these detections are actually false positives. As already discussed (Figure 5), an explanation could be that some local noise sources around each array (even if they are independent) may generate high apparent velocity wavefronts enhanced by the PWS and generate false detections, which are associated if they occur within the same 2 s time window at two arrays.

In comparison, the number of detections concomitant to three arrays is not correlated to the hourly noise level (Figure 10b). Therefore, we decide to keep and locate only events detected at the three arrays to limit the misinterpretation of false detections. Note that this is a conservative approach because we observe several detections common to only two arrays, although their waveform actually resembles real earthquakes.

All 54 events of the reference BCSF-Reness catalog have been detected concomitantly by the three arrays and are therefore part of our final catalog of 252 events. Among these events, 14 have been obtained based on the delay and stack process, assuming location in

the northern cluster (and zero when we consider the southern cluster). This indicates that, with a network roughly located in the area where induced seismicity is expected, it is not necessary to have precise a priori information on the location of these earthquakes to detect a majority of them.

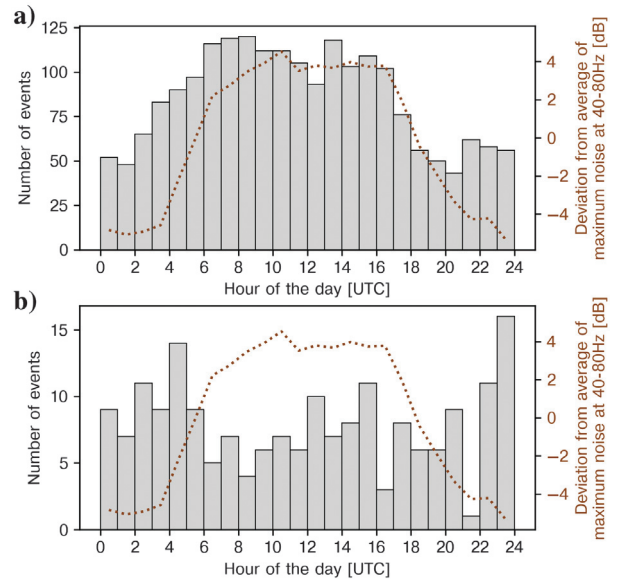


Figure 10. Histogram of the number of events per hour of the day: (a) detected at two arrays and (b) detected by three arrays (the left y-axis). The dotted brown curve represents the difference between the values of the hourly noise level (taken as the PSD power at the frequency where it reaches the maximum value between 40 and 80 Hz) and its average over the whole day.

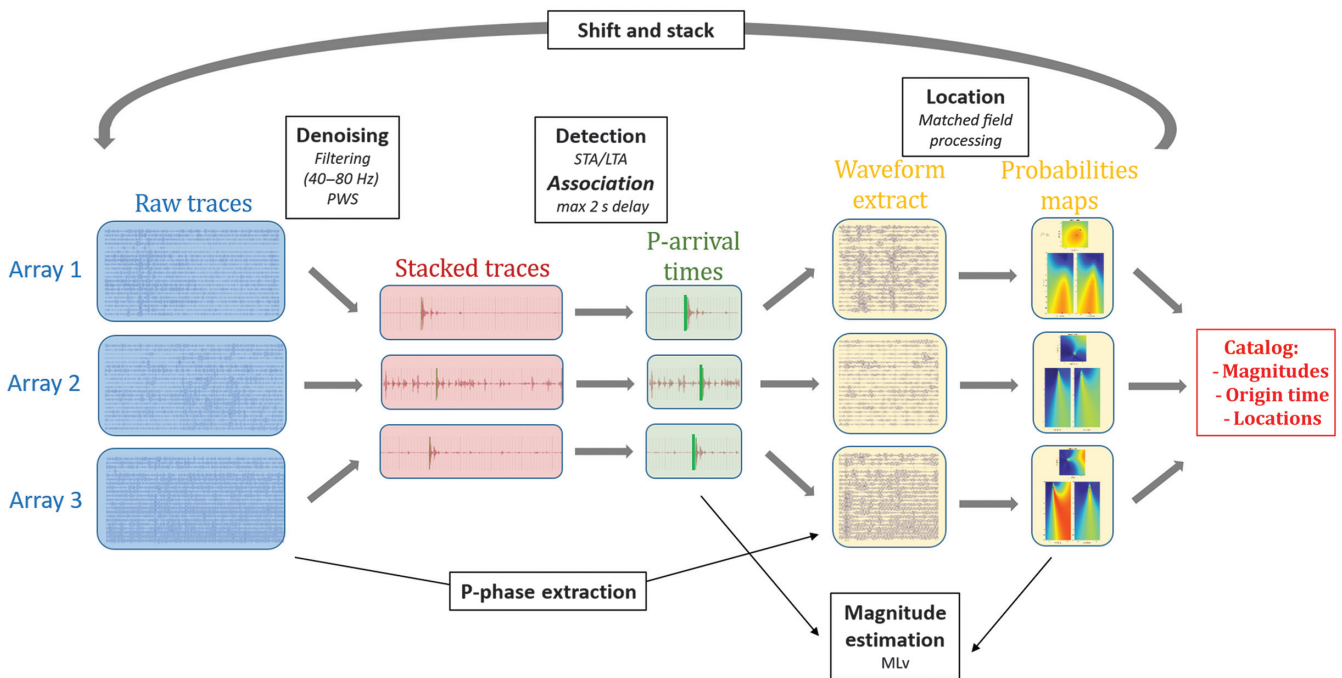


Figure 9. Chart of the whole data process, from raw data acquisition to seismic catalog.

## Time history of the present catalog

Figure 11b shows the magnitude-time distribution of the events located by MFP. According to our magnitude estimation, most of the newly detected events are of lower magnitude than the ones in the reference catalog. The distribution of the number of events per magnitude bins follows a Gutenberg-Richter law (Gutenberg and Richter, 1944) (Figure 11c), which gives credit to the magnitude estimation from PWS stacks. We estimate a  $b$ -value of  $0.62 \pm 0.03$  based on the algorithm proposed by Ogata and Katsura (1993), which is consistent with the  $b$ -value of  $0.60 \pm 0.44$  computed for the reference catalog for the same period, also using array peak amplitudes. However, the  $b$ -value estimated on our catalog has a much lower uncertainty due to the higher number of low-magnitude events. We estimated a lowering of the magnitude of completeness from 0.3 to  $-0.5$  with our approach.

The overall seismicity rate (shown in Figure 11a) is decreasing with time during the four months, especially after the shut-in. However, we notice periods of higher density of events in our catalog, which are not observed in the reference catalog. This is particularly the case around 12 December 2020 and 25 December 2020 and around 3 January 2021. These bursts of seismic activity are not associated with observable changes in the fluid pressure at the GT2 wellhead, which is smoothly decreasing with time following the shut-in (Figure 11). The fact that these periods of more intense earthquake activity are not observed in the reference catalog, which has a higher magnitude of completeness, means that the  $b$ -value evolves temporally, and is higher (in absolute value) during these bursts.

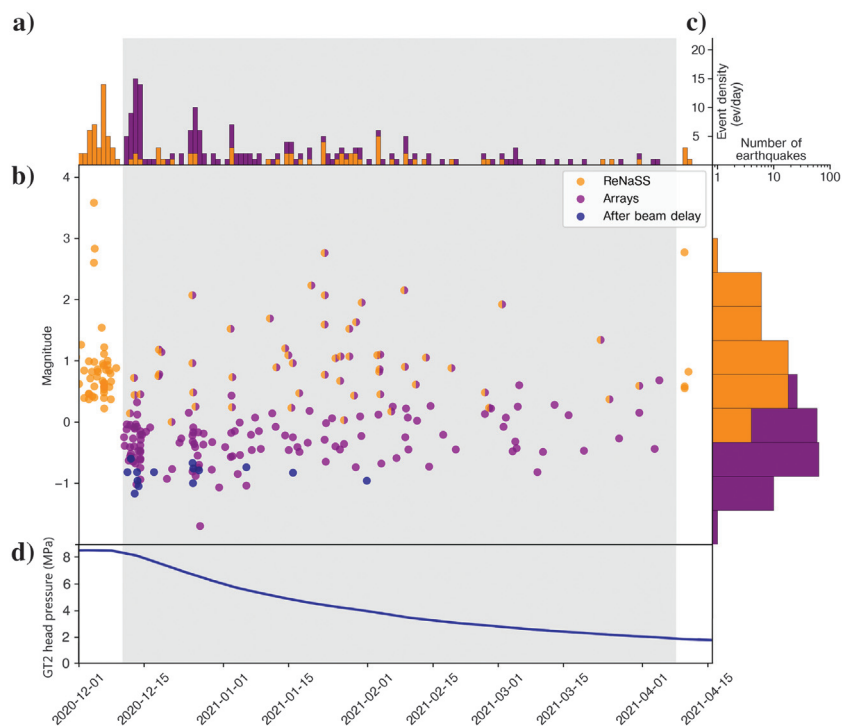


Figure 11. (a) Histogram of the event density on the whole period; (b) magnitude-time distribution of events of the reference and the present catalog (orange and purple, respectively) from 1 December 2020 to 15 April 2021, where the gray area stands for the four-month period of acquisition (from 12 December 2020 to 8 April 2021); (c) histogram of the number of earthquakes per magnitude bin for the period in gray, with the related Gutenberg-Richter laws (the dashed lines); and (d) GT2 wellhead pressure evolution during the bleed-off and shut-in of the well (modified after Terrier et al., 2022).

The first burst occurs at the very beginning of the installation period of the three arrays (10 December 2020) following the strong MLv3.6 induced event of 4 December 2020. The observed increase in the small seismicity rate from 10 December to 12 December and its stop on 15 December indicate that the temporal distribution of seismicity does not follow a traditional Omori law. The second and quite intense burst, which begins on 25 December, follows an MLv2.1 event and exhibits a time evolution of the seismic rate that is more typical of an aftershock sequence.

On the contrary, the MLv2.8 earthquake of 22 January, the largest event during our experiment, has significantly fewer aftershocks. Moreover, during the few days before and after this event — approximately between 12 January and 3 February — we observe a deficit of small-magnitude earthquakes.

As the reference catalog has been enriched with template matching, earthquakes with a similar waveform should be already referenced, so our array processing approach might be revealing new kinds of sources and aftershocks. To analyze more precisely the temporal organization of clusters of events, we compute the crosscorrelation-based similarity matrix on 1.5 s of a linearly stacked signal around the P phases arrival filtered between 20 and 80 Hz. We choose this frequency band to take into account the complexity of the signals and use a linear stack because PWS tends to distort the signal too much for a waveform similarity approach.

We build groups of events at each array independently by gathering events that have a similarity of more than 0.7 with all the others of a group. We obtain different groups at every array, but we consider as a family the combination of all the groups that share at least a common event. The threshold of 0.7 has been determined in comparison to the similarity of a signal with itself returned and inverted in time, which we consider here to be a good indicator of the minimum similarity we can observe between dissimilar signals.

To make the comparison, we apply this clustering approach to our catalog and the reference catalog on the data from arrays. We obtain 36 families of two events or more in our catalog, versus 12 with the BCSF-Reness catalog (see Figure 12 and supplemental material S5 for an example of a family). With the present catalog, the family formation rate is higher at the beginning and then decreases after the shut-in. It seems higher when the density of events is higher. Most of the families contain no event previously found by the permanent network. We note that major events are sometimes correlated with a high family creation rate, for example, the MLv = 2.1 and the MLv = 2.8, 2.2, and the 2.3 trio. However, the main events are sometimes not particularly at the origin of creating new families, even if they seem to sometimes come along with the reactivation of previous families, as suggested by the vertical alignment of earthquakes on the red lines. All this interpretation is not possible if only considering the events of the reference catalog (the orange diamonds).

## Spatial repartition

MFP allows us to successfully locate 216 events, corresponding to 86% of the events left after the association at three arrays (Figure 13). This represents more than four times more events than in the reference catalog. Different reasons can explain the nonlocation of 36 of the detected events. First, our catalog contains events not occurring around the geothermal wells, although we are filtering the signals at high frequencies. Also, there is a low but existing probability to detect non-earthquake PWS artifacts at the three arrays simultaneously.

Most of the events of the present catalog are located within or close to the northern cluster (the blue-filled ellipse). Despite not using the same velocity model as the BCSF-Reness, we observe a similar depth distribution of the events centered at approximately a 5 km depth, which is close to the depth of the wells (4.5 km). To estimate the error made on the horizontal epicenter locations, we compute the dispersion of the location probability (or coherence) maps. We approximate the coherence function as a 3D Gaussian function around the maximum value, of which we determine the standard deviation. This standard deviation is approximately 600 m on average. In addition, we use as a reference the relative relocation obtained by double differences by Schmittbuhl et al. (2021), whose relative location error has been estimated to be a few tens of meters. All the relocated hypocenters are gathered close to one another. By assuming no shift between the barycenter of the relocated events and the barycenter of the events located by MFP, the average distance between the hypocenters gives an approximation of the error made with the array processing location. Figure 14 shows such distance, computed for all the earthquakes common to both catalogs (see supplemental material S6 for the vertical distances distribution). We find that the distance distribution follows a Gaussian curve whose average is approximately 650 m. Considering these values, the localization precision of the MFP to locate events in the targeted area is similar to the precision of the BCSF-Reness (approximately 1 km according to Schmittbuhl et al., 2021). Another way to quantify the precision of the method could be by testing on synthetic data, allowing us to compare the location found by MFP with the known exact location.

Our catalog reveals three events in the southern cluster whereas the reference BCSF-Reness catalog does not show any event in this area during this four-month period. We manually checked the waveforms and relative delays between the three arrays for these events to confirm that they were actually located in this southern cluster. For the largest one, we were able to observe a faint signal at some stations from the permanent network and locate it roughly in the same area by a traditional phase-picking approach, thanks to stations used by the BCSF-Reness close to the southern cluster (see Figure 1). This observation could mean that there is still an active connection between the northern cluster and the southern cluster, even after the injection is stopped.

## DISCUSSION

We saw that using three small-aperture arrays of 21 sensors each in combination with our processing workflow can give access to a much richer earthquake catalog than a classical surface network composed of 76 local and regional stations. All the previously referenced earthquakes are well detected and located, and four times more new ones are added. Their location is coherent with the known area of activity, and the  $b$ -value of our catalog is coherent with the reference catalog in the same period of time, even with significantly better precision. The sensibility of the network allows for revealing seismic activity in the southern cluster, considered inactive for two months. This remote triggering could be due to either poroelastic pressures changes (Goebel et al., 2017) or aseismic slip (Eyre et al., 2020), but the study of those mechanisms is outside the scope of this paper, as the time period covered proved to be too short for that purpose.

Some limitations to these array processing techniques must be highlighted. We notice in Figure 13 that some events are located outside the two known clusters. Looking at those events with a low location probability, their estimated magnitude is approximately zero, and it is difficult to tell if they are real earthquakes based on their waveforms on individual traces. For the two events outside the cluster areas but with a high location probability, their waveforms suggest that they are induced earthquakes. By checking them, we notice that their P-wave arrival times are not coherent with the obtained MFP location. This shows a limit of the proposed location approach: by computing the location probability at each array separately and then combining them, we are only

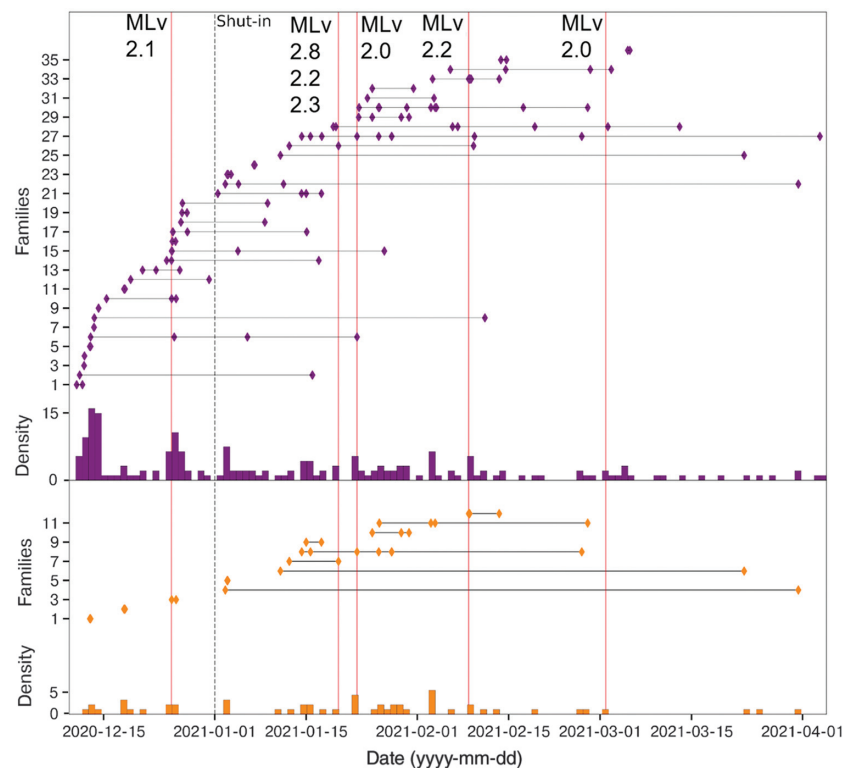


Figure 12. Formation and evolution of earthquake families with time, in parallel with event density evolution in ev/day (the bottom). Purple stands for the new catalog and dark orange for the reference catalog.

considering the intra-array time delays but not the interarray time delays, which can lead to bad locations like this. This usually does not happen when locating events of large magnitude or with a higher S/N. A way to avoid this problem is to perform MFP using longer time windows encompassing the arrival time at the three arrays while having a common starting time. Doing so, a single probability map is computed using all the nodes from the three arrays at once considering the interarray delays. We tested this approach but it did not perform well in our case with low S/N events as we are including too much noise into the signal. To overpass this difficulty, an alternative could be to compute the interstation delays in each array and then build a single cross-spectral density matrix by incorporating the time shift between the beginnings of the time windows used at each array. However, the latter should be computed accurately as it would probably have significant weight during the inversion, compared with the intra-array station delays.

Another interesting alternative would be to also work with S waves, using the horizontal components of the stations. The benefit of this could be multiple. The MFP could be performed with the P and S waves independently, making the inversion theoretically more robust. However, the S waves of small earthquakes do not show up clearly in our high-frequency band of interest, as illustrated with an example of a magnitude 0.57 event shown in

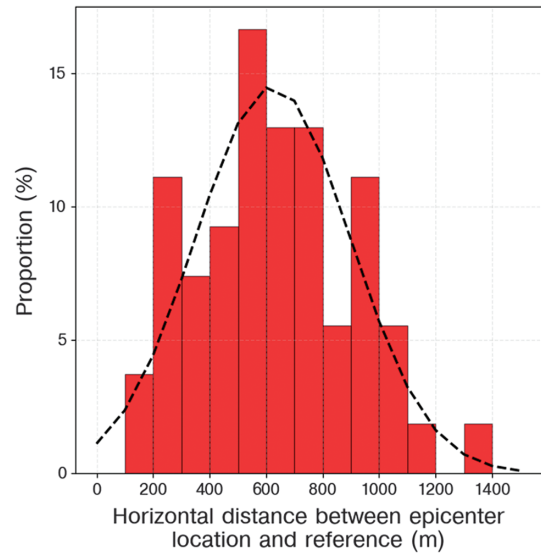


Figure 14. Distribution of the absolute horizontal distance between the differential relocation obtained by Schmittbuhl et al. (2021) and MFP hypocenters on seismic arrays for events common to both catalogs. The dotted line is the best-fitting Gaussian curve to this distribution.

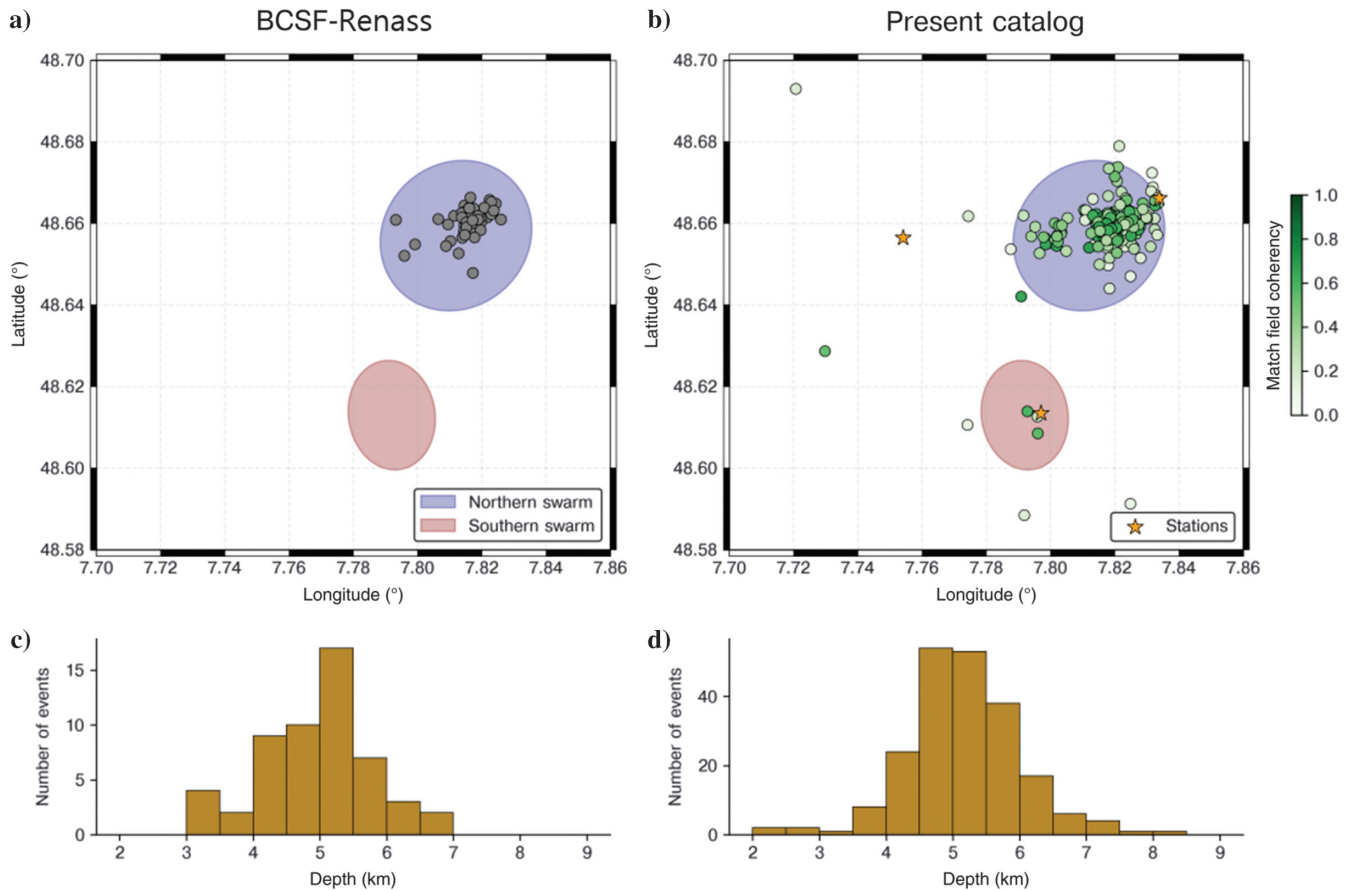


Figure 13. Comparison of the induced earthquake locations for the four-month period from (a and c) the BCSF-Reness French institute catalog and (b and d) the present catalog built using arrays for the four-month period. (a and b) Maps of the epicenters. The areas where the seismicity occurred during the whole induced seismic sequence are roughly drawn in blue and red color. The stars show the array's location. The orange color scale of scatter markers refers to the maximum probability found with the MFP. (c and d) Depth distribution.

Figure 15. According to the vertical component, the P wave radiates an important amount of energy between 10 and 80 Hz. In comparison, the S wave seen on the horizontal components radiates energy at lower frequencies, mainly between 5 and 30 Hz. As shown previously, this corresponds to a frequency range where the noise level is significantly higher. Of the 216 located events in our catalog, we were able to confidently isolate the S waves on only 92 of them in the 5–30 Hz frequency band. Moreover, even if the selected window is well centered on the S wave, there is still a risk for the crosscorrelation performed during the MFP to be biased by the important anthropogenic noise in such a frequency range, especially for low S/N earthquakes.

We also investigate the influence of the geometry and number of stations on the detection performances by evaluating the detection performances of our arrays if we only use a portion of them. We computed PWS traces using a portion of the available stations, whether by removing one or two of the circles or by decimating by half the number of sensors of each circle. The number of sen-

sors used seems to have an impact on the computed S/N, but not the relative placement of the station (supplemental material S4). The gain seems to approximately follow a  $\sqrt{N}$  law, as estimated by the theory. However, the S/N alone does not provide all the information about the detectability of a signal, as the latter highly depends on the way the S/N is computed, the context, and the kind of detector used (Desimoni and Brunetti, 2015). Therefore, to better evaluate the detection performances in practice, we applied all the detection and association processes described in this paper on the PWS traces obtained for those different arrays' configuration and compared the number of events finally obtained for each 0.5 magnitude bin (Figure 16). Globally, using more stations allows us to obtain more events. However, it can be seen that using the smaller circle is the best choice when limiting the array to a central sensor and a single circle. Also, using the largest circle together with the smallest is better than using any other combination of two circles. A quantitative comparison of location performances for different array geometries would be complementary to these observations and would help to find the best trade-off between the number of stations/repartition and performances for each case study.

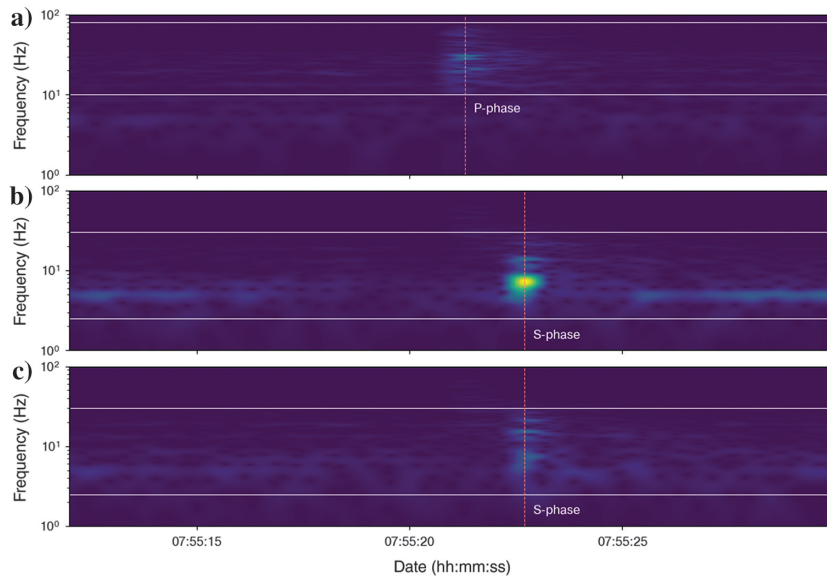


Figure 15. The spectrogram on a 20 s long linear stack on array 1 during a magnitude 0.57 earthquake (17 January 2021) on the (a) vertical (DLZ), (b) east (DLE) and (c) north (DLN) components (normalized over the three components).

CONCLUSION

We propose an array processing method with small seismic arrays to monitor induced micro-seismicity in urban contexts where the strong anthropogenic noise level makes the detection of small events difficult with standard seismic networks. The design of our arrays allows us to investigate up to high frequencies (40–80 Hz) without meeting aliasing problems and to enhance the S/N using a PWS technique. After an STA/LTA detection and association, we automatically locate earthquakes with an MFP approach.

We apply this method to the four-month period following the shut-in of the injection in the Georhin wells (Vendenheim, France). We significantly enriched the reference catalog made with the local surface network with five times more earthquakes, lowering the magnitude of completeness by more than half an order of magnitude. Some of these earthquakes have been manually checked by analysts using the permanent network. The higher number of referenced events and the enhanced S/N allow us to have more insight into the sub-cluster behaviors of low-magnitude events. Our enriched catalog highlights an intermittence of seismic activity in the northern cluster during the period preceding the largest event, although the fluid pressure is decreasing progressively. Also, some activity is revealed in the southern cluster, located 5 km south of the injection well, previously considered as quiescent during this period. Similarly to what the reference catalog shows for the whole sequence, we do not find any evidence of earthquakes in between the two clusters, which leaves open the question of the mechanisms for stress transfer.

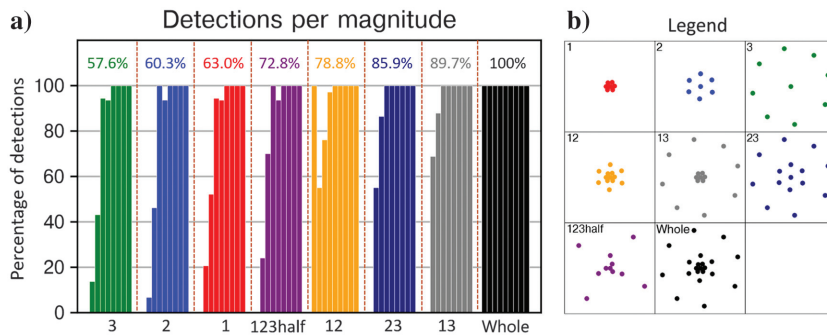


Figure 16. (a) Histogram of the number of detected earthquakes per 0.5 wide magnitude bins (from  $M_L = -1.5$  to  $M_L = 3$ ) for each array configuration, as a proportion of the total number of events detected using all the available sensors. (b) A legend describing the different configurations tested.



The proposed method could potentially be implemented in an automatic process, and in real-time if the instrument allows a continuous transfer of data. It is particularly suitable to refine the monitoring of the time evolution of the induced microseismicity, while being self-sustaining in locating the new events. However, the current implementation of the location algorithm sometimes produces for the smallest events a location that is not consistent with the relative arrival times between the arrays. To address this issue, we plan to investigate the horizontal component and use the S-wave arrivals in addition to the P-wave arrivals in the location process, which is challenging because of the lower S/N at frequencies where the S waves contain more energy.

## ACKNOWLEDGMENTS

The authors are grateful to Maxime Bès de Berc, who managed the SmartSolo nodes setting and data recovery, also to the sismo-ITES team for their precious help with the fieldwork, and to the owners of the fields for allowing us to deploy the instruments on their properties. The paper benefitted from relevant comments by Kyosuke Okamoto, two anonymous reviewers, peer-review advisor Barbara Cartwright, and associate editor Erika Gasperikova. This work was performed under the framework of the Laboratory of Excellence LABEX ANR-11-LABX-0050- G-EAU-THERMIE-PROFONDE and the Interdisciplinary Thematic Institute GeoT, as part of the ITI 2021-2028 program of the University of Strasbourg, CNRS, and Inserm, supported by IdEx Unistra (ANR 10 IDEX 0002) and by SFRI STRAT'US project (ANR 20 SFRI 0012) under the framework of the French Investments for the Future Program. It was part of the GEOTHERMICA DEEP project (Innovation for De-Risking Enhanced Geothermal Energy Projects) (<http://deepgeothermal.org>). R. Fiori was partly funded by ADEME grant #TEZ20-020 and LABEX ANR-11-LABX-0050- G-EAU-THERMIE-PROFONDE.

## DATA AND MATERIALS AVAILABILITY

Data associated with this research are available and can be obtained by contacting the corresponding author.

## REFERENCES

- Aki, K., and P. G. Richards, 1980, Quantitative seismology — Theory and methods: Freeman and Company, Vol II, 609–625.
- Baggeroer, A., W. Kuperman, and P. Mikhalevsky, 1993, An overview of matched field methods in ocean acoustics: *IEEE Journal of Oceanic Engineering*, **18**, 401–424, doi: [10.1109/48.262292](https://doi.org/10.1109/48.262292).
- Bracewell, R. N., 1965, The Fourier transform and its applications: McGraw-Hill.
- Bullen, K. E., and B. A. Bolt, 1985, An introduction to the theory of seismology, 4th ed.: Cambridge University Press, 451–454.
- Buttkus, B., 2000, Spectral analysis and filter theory in applied geophysics: Springer Verlag.
- Capon, J., 1967, High-resolution frequency-wavenumber spectrum analysis: *Proceedings of the IEEE*, **57**, 1408–1418, doi: [10.1109/PROC.1969.7278](https://doi.org/10.1109/PROC.1969.7278).
- Capon, J., 1973, Signal processing and frequency-wavenumber spectrum analysis for a large aperture seismic array, in B. A. Bolt, ed., *Methods in computational physics*: Academic Press, 1–59.
- Chmiel, M., P. Roux, and T. Bardainne, 2019, High-sensitivity microseismic monitoring: Automatic detection and localization of subsurface noise sources using matched-field processing and dense patch arrays: *Geophysics*, **84**, no. 6, KS211–KS223, doi: [10.1190/geo2018-0537.1](https://doi.org/10.1190/geo2018-0537.1).
- Corciulo, M., P. Roux, M. Campillo, D. Dubucq, and W. A. Kuperman, 2012, Multiscale matched-field processing for noise-source localization in exploration geophysics: *Geophysics*, **77**, no. 5, KS33–KS41, doi: [10.1190/geo2011-0438.1](https://doi.org/10.1190/geo2011-0438.1).
- Cuenot, N., J.-P. Faucher, D. Fritsch, A. Genter, and D. Szablinski, 2008, The European EGS project at Soultz-sous-Forêts: From extensive exploration to power production: *IEEE Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, 1–8.
- Dean, T., 2017, The seismic signature of rain: *Geophysics*, **82**, no. 5, P53–P60, doi: [10.1190/geo2016-0421.1](https://doi.org/10.1190/geo2016-0421.1).
- Desimoni, E., and B. Brunetti, 2015, About estimating the limit of detection by the signal to noise approach: *Pharmaceutica Analytica Acta*, **6**, 355, doi: [10.4172/2153-2435.1000355](https://doi.org/10.4172/2153-2435.1000355).
- Eisner, L., B. J. Hulsey, P. Duncan, D. Jurick, H. Werner, and W. Keller, 2010, Comparison of surface and borehole locations of induced seismicity: *Geophysical Prospecting*, **58**, 809–820, doi: [10.1111/j.1365-2478.2010.00867.x](https://doi.org/10.1111/j.1365-2478.2010.00867.x).
- Ellsworth, W. L., D. Giardini, J. Townend, S. Ge, and T. Shimamoto, 2019, Triggering of the Pohang, Korea, earthquake ( $M_w$  5.5) by enhanced geothermal system stimulation: *Seismological Research Letters*, **90**, 1844–1858, doi: [10.1785/0220190102](https://doi.org/10.1785/0220190102).
- Eyre, T. S., M. Zecevic, R. O. Salvage, and D. W. Eaton, 2020, A long-lived swarm of hydraulic fracturing-induced seismicity provides evidence for aseismic slip: *Bulletin of the Seismological Society of America*, **110**, 2205–2215.
- Foulger, G. R., M. P. Wilson, J. G. Gluyas, B. R. Julian, and R. J. Davies, 2018, Global review of human-induced earthquakes: *Earth-Science Reviews*, **178**, 438–514, doi: [10.1016/j.earscirev.2017.07.008](https://doi.org/10.1016/j.earscirev.2017.07.008).
- GeORG Projektteam, 2013, Geopotenziale des tieferen Untergrundes im Oberheingraben: Fachlich-Technischer Abschlussbericht des INTER-REG-Projekts GeORG, <http://www.geopotenziale.org>, accessed 07 July 2022.
- Goebel, T. H. W., M. Weingarten, X. Chen, J. Haffener, and E. E. Brodsky, 2017, The 2016 Mw5.1 Fairview, Oklahoma earthquakes: Evidence for long-range poroelastic triggering at >40 km from fluid disposal wells: *Earth and Planetary Science Letters*, **472**, 50–61.
- Gouedard, P., 2008, Techniques de corrélation: Aspects méthodologiques et applications à la sub-surface: Doctoral dissertation, Université Joseph-Fourier - Grenoble I, Français.
- Gutenberg, B., and C. F. Richter, 1944, Frequency of earthquakes in California: *Bulletin of the Seismological Society of America*, **34**, 185–188, doi: [10.1785/BSSA0340040185](https://doi.org/10.1785/BSSA0340040185).
- Halldórsson, B., R. Sigbjörnsson, and J. Schweitzer, 2009, ICEARRAY: The first small-aperture, strong-motion array in Iceland: *Journal of Seismology*, **13**, 173–178, doi: [10.1007/s10950-008-9133-z](https://doi.org/10.1007/s10950-008-9133-z).
- Helmholtz-Centre Potsdam–GFZ German Research Centre for Geosciences and gempa GmbH, 2008, The SeisCompP seismological software package (GFZ Data Services), <https://www.seiscomp.de/doc/toc/extensions/global.html>, accessed 20 November 2022.
- Jiang, H. Y., Z. B. Chen, X. X. Zeng, H. Lv, and X. Liu, 2016, Velocity calibration for microseismic event location using surface data: *Petroleum Science*, **13**, 225–236, doi: [10.1007/s12182-016-0092-7](https://doi.org/10.1007/s12182-016-0092-7).
- Kang, J. Q., J. Zhu, and J. Zhao, 2019, A review of mechanisms of induced earthquakes: From a view of rock mechanics: *Geomechanics and Geophysics for Geo-Energy and Geo-Resources*, **5**, 171–196, doi: [10.1007/s40948-018-00102-z](https://doi.org/10.1007/s40948-018-00102-z).
- Kennett, B. L., 2002, The seismic wavefield, Volume II: Interpretation of seismograms on regional and global scale: Cambridge University Press, 163–170.
- Kwiatk, G. S., P. Martinez-Garzon, K. Passmore, P. Passmore, S. Valenzuela, and C. Wollin, 2019, Controlling fluid-induced seismicity during a 6.1-km-deep geothermal stimulation in Finland: *Science Advances*, **5**, 1–11, doi: [10.1126/sciadv.aav7224](https://doi.org/10.1126/sciadv.aav7224).
- McNamara, D. E., and R. P. Buland, 2004, Ambient noise levels in the continental United States: *Bulletin of the Seismological Society of America*, **94**, 1517–1527, doi: [10.1785/012003001](https://doi.org/10.1785/012003001).
- McNeff, J. G., 2002, The global positioning system: *IEEE Transactions on Microwave Theory and Techniques*, **50**, 645–652, doi: [10.1109/22.989949](https://doi.org/10.1109/22.989949).
- 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: *Scientific Reports*, **7**, 2045–2322, doi: [10.1038/s41598-017-13585-9](https://doi.org/10.1038/s41598-017-13585-9).
- Ogata, Y., and K. Katsura, 1993, Analysis of temporal and spatial heterogeneity of magnitude-frequency distribution inferred from earthquake catalogues: *Geophysical Journal International*, **113**, 727–738, doi: [10.1111/j.1365-246X.1993.tb04663.x](https://doi.org/10.1111/j.1365-246X.1993.tb04663.x).
- Peterson, J., 1993, Observation and modeling of seismic background noise: *U.S. Geol. Surv. Tech. Rept.* 93-322, 1–95.
- Riahi, N., and P. Gertstoft, 2015, The seismic traffic footprint: Tracking trains, aircraft, and cars seismically: *Geophysical Research Letters*, **42**, 2674–2681, doi: [10.1002/2015GL063558](https://doi.org/10.1002/2015GL063558).
- Rost, S., and C. Thomas, 2009, Improving seismic resolution through array processing techniques: *Surveys in Geophysics*, **30**, 271–299, doi: [10.1007/s10712-009-9070-6](https://doi.org/10.1007/s10712-009-9070-6).

- Rothe, J., and E. Peterschmitt, 1950, Etudes sismiques des explosions d'Haslach (Forêt noire): *Journal de Physique et Le Radium*, **11**, 30–31.
- Schimmel, M., and H. Paulssen, 1997, Noise reduction and detection of weak, coherent signals through phase-weighted stacks: *Geophysical Journal International*, **130**, 497–505, doi: [10.1111/j.1365-246X.1997.tb05664.x](https://doi.org/10.1111/j.1365-246X.1997.tb05664.x).
- Schmittbuhl, J., S. Lambotte, O. Lengliné, M. Grunberg, H. Jund, J. Vergne, and F. Masson, 2021, Induced and triggered seismicity below the city of Strasbourg, France from November 2019 to January 2021: *Comptes Rendus Geosciences*, **353**, 561–584, doi: [10.5802/crgeos.71](https://doi.org/10.5802/crgeos.71).
- Schweitzer, J., J. Fyen, S. Mykkeltveit, S. J. Gibbons, M. Pirli, D. Kühn, and T. Kværna, 2011, New manual of seismological observatory practice (NMSOP-II), Chap. 9 seismic arrays, 28–30.
- Schweitzer, J., J. Fyen, S. Mykkeltveit, S. J. Gibbons, M. Pirli, D. Kühn, and T. Kværna, 2012, New manual of seismological observatory practice 2 (MNSOP-2).
- Sick, B., and M. Joswig, 2017, Combining network and array waveform coherence for automatic location: examples from induced seismicity monitoring: *Geophysical Journal International*, **208**, 1373–1388, doi: <https://doi.org/10.1093/gji/ggw468>.
- Terrier, M., F. De Santis, R. Soliva, B. Valley, D. Bruel, Y. Géraud, and J. Schmittbuhl, 2022, Rapport Phase 1 du CE créé en appui à l'administration sur la boucle géothermique GEOVEN: Préfecture du Bas-Rhin, Strasbourg, France, 26 avril 2022.
- Thurber, C. H., X. Zeng, A. M. Thomas, and P. Audet, 2014, Phase-weighted stacking applied to low-frequency earthquakes: *Bulletin of the Seismological Society of America*, **104**, 2567–2572, doi: [10.1785/0120140077](https://doi.org/10.1785/0120140077).
- Umlauf, J., and M. Korn, 2019, 3-D fluid channel location from noise tremors using matched field processing: *Geophysical Journal International*, **219**, 1550–1561, doi: [10.1093/gji/ggz385](https://doi.org/10.1093/gji/ggz385).
- Umlauf, J., F. Lindner, P. Roux, T. D. Mikesell, M. M. Haney, M. Korn, and F. T. Walter, 2021, Stick-slip tremor beneath an Alpine glacier: *Geophysical Research Letters*, **48**, e2020GL090528, doi: [10.1029/2020GL090528](https://doi.org/10.1029/2020GL090528).
- Wathelet, M., D. Jongmans, and M. Ohrnberger, 2005, Direct inversion of spatial autocorrelation curves with the neighborhood algorithm: *Bulletin of the Seismological Society of America*, **95**, 1787–1800, doi: [10.1785/0120040220](https://doi.org/10.1785/0120040220).
- Wathelet, M., D. Jongmans, M. Ohrnberger, and S. Bonnefoy-Claudet, 2008, Array performances for ambient vibrations on a shallow structure and consequences over  $V_S$  inversion: *Journal of Seismology*, **12**, 1–19, doi: [10.1007/s10950-007-9067-x](https://doi.org/10.1007/s10950-007-9067-x).
- Wiemer, S., D. Giardini, D. Fäh, N. Deichmann, and S. Sellami, 2009, Probabilistic seismic hazard assessment of Switzerland: Best estimates and uncertainties: *Journal of Seismology*, **13**, 449–478, doi: [10.1007/s10950-008-9138-7](https://doi.org/10.1007/s10950-008-9138-7).
- Yeo, I. W., M. R. Brown, S. Ge, and K. K. Lee, 2020, Causal mechanism of injection-induced earthquakes through the  $M_w$  5.5 Pohang earthquake case study: *Nature Communications*, **11**, 2614, doi: [10.1038/s41467-020-16408-0](https://doi.org/10.1038/s41467-020-16408-0).
- Zang, A., V. Oye, P. Jousset, N. Deichmann, R. Gritto, A. McGarr, and D. Bruhn, 2014, Analysis of induced seismicity in geothermal reservoirs — An overview: *Geothermics*, **52**, 6–21, doi: [10.1016/j.geothermics.2014.06.005](https://doi.org/10.1016/j.geothermics.2014.06.005).
- Zigone, D., 2012, Les tremors non volcaniques: Observations et modélisations: Doctoral dissertation, University of Grenoble, Français.

Biographies and photographs of the authors are not available.