# Complementary Dense Datasets Acquired in a Low-to-Moderate Seismicity Area for Characterizing Site Effects: Application in the French Rhône Valley

Berenice Froment, Andrés Olivar-Castaño, Matthias Ohrnberger, Loic Gisselbrecht, Katrin Hannemann, Edward Marc Cushing, Pierre Boué, Céline Gélis, Annabel Haendel, Marco Pilz, et al.

# Complementary dense datasets acquired in a low-to-moderate seismicity area for characterizing site effects: application in the French Rhône valley.

B. Froment[1], A. Olivar-Castaño[2], M. Ohrnberger[2], L. Gisselbrecht[1], K. Hannemann[2*], E.M. Cushing[1], P. Boué[3], C. Gélis[1], A. Haendel[4], M. Pilz[4], L. Hillmann[4], O. Barbaux[1], S. Beauprêtre[5], G. Bouzat[6], E. Chaljub[3], F. Cotton[4#], F. Lavoué[1§], L. Stehly[3], C. Zhu[4], O. Magnin[7], L. Métral[3], A. Mordret[3], Y. Richet[1], A. Tourette[7]

[1] *Institut de Radioprotection et Sûreté Nucléaire (IRSN), PSE-ENV, SCAN, BERSSIN, Fontenay-aux-Roses, France.*

[2] *Univ. of Potsdam, Institute of Geosciences, Potsdam, Germany*

[3] *Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, UGE, ISTerre, Grenoble, France*

[4] *GFZ, German Research Center for Geosciences, Potsdam, Germany*

[5] *Sisprobe by EGIS, Grenoble, France*

[6] *ORANO, Chimie-Enrichissement Tricastin, Pierrelatte, France*

[7] *EGIS, Seyssins, France*

** Now at Institute of Geophysics, University of Muenster, Muenster, Germany*

*# Also at Univ. of Potsdam, Institute of Geosciences, University of Potsdam, Potsdam, Germany*

[§] *Also at Univ. Grenoble Alpes, Univ. Savoie Mont Blanc, CNRS, IRD, UGE, ISTerre, Grenoble, France*

**Corresponding author:** Bérénice Froment

berenice.froment@irsn.fr;

Postal address:
PSE-ENV/SCAN
IRSN
BP 17
92262 Fontenay-aux-Roses cedex
France

ABSTRACT

Superficial geological layers can strongly modify the surface ground motion induced by an earthquake. These so-called site effects are highly variable from one site to another and still difficult to quantify for complex geological configurations. That is why site-specific studies can greatly contribute to improve the hazard prediction at a specific site. However, site-specific studies have historically been considered difficult to carry out in low-to-moderate seismicity regions. We present here seismological datasets acquired in the framework of the French-German DARE project in the heavily industrialized area surrounding the French Tricastin Nuclear Site (TNS). TNS is located above an ancient canyon dug by the Rhône River during the Messinian period. The strong lithological contrast between the sedimentary fill of the canyon and the substratum, as well as its expected confined geometry make this canyon a good candidate for generating site effects which are variable on short spatial scales. In order to investigate the impact of this geological structure on the seismic motion, we conducted complementary seismic campaigns in the area. The first main campaign consisted of deploying 400 nodes over a 10x10 km area for one month and aimed at recording the seismic ambient noise. A second seismic campaign involved the deployment of 49 broadband stations over the same area for more than eight months. This complementary campaign aimed at recording the seismicity (including local, regional and teleseismic events). These different designs allowed us to target a variety of seismic data, at different spatial and temporal scales. Beyond the interest for local operational seismic hazard applications,

these datasets may be valuable for studying seismic wave propagation within complex km-scale sedimentary structures. In this paper we present the deployment designs as well as initial analyses to provide information on the characteristics and the overall quality of the data acquired to future users.

## INTRODUCTION (EXPERIMENT MOTIVATION):

It is well-known that superficial geological layers can strongly modify the surface ground motion induced by an earthquake. Soil properties in the vicinity of the Earth's surface generally become softer leading to an amplification of the seismic motion. In the case of complex geological structures, such as sedimentary valleys, seismic waves can be trapped and the geometry of the soft deposits will further affect the ground motion by increasing both the duration and amplitude of the shaking (e.g. Bard and Bouchon, 1985; Kawase, 1996; Semblat et al., 2005). These so-called *site effects* are a source of particular concern for Seismic Hazard Assessment (SHA), as they can greatly increase the level of seismic hazard in critical zones located on sedimentary basins such as big cities (e.g. Mexico City, Mexico; Los Angeles, USA; Tokyo, Japan; Grenoble, France) or industrialized areas with critical infrastructure.

By being related to local conditions, site effects are highly variable from one site to another and still difficult to quantify for some geological configurations (e.g. deep valleys or canyons). Recent studies (e.g. Pilz and Cotton, 2019) have for example confirmed the limitation of 1D models to predict site amplifications. That is why site-specific studies can greatly contribute to improve the hazard prediction at a specific site in comparison to ergodic estimates based on data from global databases. However, they have historically been considered difficult to carry out in low-to-moderate seismicity regions where moderate to large earthquakes have long return periods.

The French-German project *Dense ARray for seismic site effect Estimation* - DARE (IRSN; Univ. of Potsdam; GFZ Potsdam; Univ. Grenoble Alpes), funded by the French and German Research Agencies, aims to propose new approaches based on the acquisition of dense in-situ datasets for the estimation of site effects (and the application of site-specific studies) in low-to-moderate seismicity regions. The contribution and interest of innovative methods, combining, in particular, dense array acquisition and the use of seismic ambient noise will be investigated within the framework of site effect studies. The DARE project targets the heavily industrialized area of the widespread Tricastin Nuclear Site (TNS) in the French Rhône valley. TNS is located on the deep and elongated Messinian Rhône Canyon. This canyon was dug about 6 million years ago during the Messinian Salinity Crisis (MSC) in the Mesozoic substratum (Cretaceous limestones and sandstones). This canyon is filled with Pliocene marine and continental sediments (sands and clays) nowadays covered by the Rhône Quaternary terrace (Holocene). Lithological information from boreholes reaching the bedrock and preliminary geophysical campaigns indicate that the canyon can reach locally >500 m and is deeply incised (Gélis et al., 2022). The strong material contrast between the sedimentary fill and the substratum (estimated Vs contrast around 3 from Gélis et al., 2022), as well as its expected confined geometry make this canyon a good candidate for generating site effects. Gélis et al. (2022) have reported local ground-motion amplification reaching a factor of 6 for some frequencies on top of the canyon relative to nearby Cretaceous limestone outcrops, using about one year of continuous recordings. This first study quantifies the seismic amplification associated with the presence of the canyon at two sites on top of the sediment canyon. One of the objectives of the DARE project is to extend this estimation to the scale of the sediment canyon in order to catch the spatial variability of the amplification caused by this complex geometric structure.

It is worth noting that the interest of studying this area has been brought to the forefront with the occurrence of the Mw4.9 Le Teil earthquake. This event took place on November 11, 2019 about 20 km north of TNS and severely damaged several villages in the vicinity of the rupture area (Ritz et al., 2020; Cornou et al., 2021). It corresponds to the most destructive and strongest earthquake in metropolitan France since 1967. This event highlights the issue of the seismic hazard in the region and brings a new dimension to the DARE project (launched and funded before the occurrence of this earthquake).

In the framework of the DARE project we conducted two complementary seismic campaigns. The first campaign, carried out by IRSN with the help of EGIS and SISPROBE companies, consisted of deploying more than 400 all-in-one seismic nodes over a 10x10 km area for one month (winter 2020). This campaign targeted the recording of seismic ambient noise generated by both global and local sources (Froment et al., 2023). To complement this first acquisition, a second seismic campaign was carried out by the 4 partners of the DARE project. This second campaign consisted of deploying about 50 broadband stations over the same area for more than eight months (September 2020 - May 2021) and aimed to record the seismicity (including teleseismic events, local and regional seismicity) (Pilz et al., 2021).

These two experiments provide complementary datasets with different temporal and spatial scales, targeting different observables (ambient noise; seismicity). Seismic ambient noise will be used as an alternative seismic data whose exploitation deserves to be encouraged in the estimation of seismic amplification due to site effects. This is particularly true in low-to-moderate seismic areas such as France and Germany where seismic campaigns may turn out to last long before catching enough seismicity to get statistically robust results; thereby limiting the widespread use of an empirical estimation of site effects in an operational context. These complementary datasets will make it possible to propose and compare alternative methods for site effect estimation; evaluate their interests, uncertainties and limitations. The density of instruments considered in these 2 experiments will help to 1) provide high-resolution imaging of the medium and 2) to capture the variability and multi-dimensional features of the site effects related to the expected complex geometry of the geological structure. It will therefore increase the resolution of the local site-specific study. The implication of such study in SHA will be investigated by comparing the site-specific site responses derived in the DARE project from extensive datasets and 3D medium characterization, with those from ergodic approaches (e.g. based on site proxies) or 1D modelling that are commonly used in SHA studies especially in low-to-moderate seismicity areas. This valley is also representative of deep valleys whose amplification cannot be correctly predicted from surface geophysical or geotechnical parameters (e.g. Vs30) alone. Taking into account the effects of valley thickness and configuration is currently an important topic of discussion in the groups in charge of seismic standards (e.g. Paolucci et al., 2021) and the new version of the European Seismic building code (2021-draft) introduces explicitly a further “F” category for deep soil deposits (H800 > 100 m). The data acquired in this experiment will contribute to a better understanding of the factors that control site effects. This will help to validate and improve, for such “deep valleys” site classes, building codes amplification factors and also identify the best parameters (proxies) for predicting them and reduce the variability of potential site response within site classes.

This extensive seismic campaign will provide a deep knowledge on the way the geological structure impacts the seismic motion in the area of Tricastin where some critical infrastructure is located. It is worth noting that studies are also underway to build a 3D accurate geological model of the area (Bagayoko, M.Sc. thesis, 2021). These various approaches and data will contribute to produce an extensive characterization of the medium and the seismic motion, of great interest for the study of site effects.

The direct contribution of such seismological acquisitions in terms of SHA as well as the occurrence of the Le Teil earthquake in the area enhance the interest of the 2 acquired datasets at a national scale beyond the initial framework. Moreover, to the best of our knowledge, such complementary dense arrays have not been deployed so far at such spatial scale in Metropolitan France and Germany, increasing their interest at the national scale.

Beyond the local or even national interest, these datasets provide extensive observations on a km-scale western European sedimentary basin. Continuous advances in seismic instrumentation, storage and computation capacities will favor similar campaigns in the future. Repeating the same kind of acquisition to other European structures (with a similar scale and context) will reveal to what extent they show common features. Results can then help to define how to consider the impact of such structures in seismic regulations or guides. They can also be confronted to what we know from worldwide sedimentary basins, in different contexts or at different scales that may dominate global databases. Again, this would help understand the conditions of applications and limitations of the use of ergodic approaches.

For all these reasons, these datasets (Pilz et al., 2021; Froment et al., 2023) will be made available to the scientific community at the end of the DARE Project (see Data and Resources section).

# INSTRUMENT DEPLOYMENT

## TARGET ZONE

In 2019, when the DARE project was initiated, the local geology of the Messinian canyon remained poorly documented in the region of Tricastin. Gélis et al (2022) provide some first insights about the canyon rims and the subsurface characteristics locally. Lithological information from boreholes in the area (BSS-Infoterre Underground database) combined with a thorough geological study allowed them to approximately locate the canyon rims in the area (Figure 1). Moreover, the same borehole data and 1D geophysical medium characterization provide local knowledge about the nature and characteristics of the sedimentary canyon fill and bedrock. In particular, Gélis et al. (2022) show that the base of the canyon deepens southward, consistently with the Rhône flow direction, reaching a depth of at least 500 m at a distance of 2-to-3 km to the south of TNS (in the vicinity of site BOLL in Figure 1(a)). At this location, the canyon bottom incises or at least, lies directly on top of Urgonian (lower cretaceous) limestones. Finally, in the same area, it has been deduced that the canyon is particularly narrow, with an E-W width that is not greater than 4 km.

From these first observations, we targeted a 10 km by 10 km area surrounding the imprint of Pliocene and Quaternary sediment deposits and TNS (Figure 1). This extension allows us to embed nearby outcrops of cretaceous series incised by the canyon and that constitutes the basement of the canyon sedimentary fill.

It is worth noting that most of this target zone is located in a heavily industrialized area including the widespread TNS, a hydroelectric dam and 5 towns (>45 000 inhabitants). It is also crossed from north to south by several railroads (including freight lines and the high-speed TGV train), the busy A7 highway and N7 national road. A map displaying this infrastructure is given in the electronic supplement (Figure S1). Many cultivated fields can also be found in the central part of the target zone. Quieter environments can be found at the eastern and western edges of the zone. It is worth noting that the spatial distribution of noisy and quiet environments matches approximately the geological setting: noisy environments are rather located in the valley, i.e. on top of the sedimentary canyon, whereas quiet environments are rather located on surrounding foothills, i.e. on cretaceous outcrops-.

# Node deployment

## Preliminary experiment: Noise Test

IRSN with the help of EGIS and SISPROBE companies, had the objective to deploy about 400 all-in-one seismic nodes over our target area to record the seismic ambient noise for one month. Before this massive deployment, a smaller scale campaign was carried out to investigate the feasibility, constraints, limitations of the planned dense ambient noise experiment. In particular, this so-called *noise test* aimed to investigate the quality of continuous measurements and the characteristics of seismic noise in the area in order to refine the design of the 400-instrument experiment. In this context, we deployed 30 3-component Geospace GSX nodes (with 5-Hz GSC-3C-LF geophones, sampling frequency of 250 Hz) following a spiral-shaped array over the 10 km x 10 km zone (Figure 1(a)). This design allowed us to sample a wide range of interstation distances and azimuths. The center (and denser) part of the spiral is located in the southeastern part of the target zone where the Messinian canyon is expected to be the deepest and the narrowest (Gélis et al., 2022) and where we planned to densify the 400-node deployment. These 30 nodes were deployed on November 5, 2019. This noise test was supposed to last for one week. However, on November 11, the Mw4.9 Le Teil earthquake occurred at about 20 km north of the 30-node array. The noise test array thus kept installed one more week, that is, until November 19. The selection of a high gain for the instrument response to measure background vibrations was not appropriate for motions as strong as the one generated by the Le Teil earthquake on the array. The recordings were thus clipped on most of the 30 nodes preventing usual ground motion analysis.

During this preliminary experiment, broadband Guralp CMG6-TD instruments were also co-located with nodes at 3 sites (see Figure 2(a) for a picture of co-located instruments). These 3 sites were previously instrumented during temporary campaigns since 2016 (Gélis et al., 2022). Originally, they were called BOLL, PAUL and ADHE in reference to the names of the localities where they were deployed. BOLL and PAUL are located on top of the Messinian Canyon while ADHE is located on nearby cretaceous outcrops (Figure 1(a)). ADHE has been considered as a local reference rock site for the estimation of seismic amplification associated with the presence of the canyon at BOLL and PAUL (Gélis et al., 2022). It is worth noting that these 3 historical sites have been instrumented during all the acquisitions carried out in the DARE project using different instrumentation. Table 1 summarizes information (naming and instrumentation) relative to these 3 historical sites for the different acquisitions. The co-location of nodes and CMG6-TD instruments allowed us to investigate the ability of node recordings to reproduce broadband station recordings especially at low frequencies (i.e. below the cut-off frequency of 5 Hz). This is detailed in the section discussing the quality of the nodal dataset.

Noise correlation functions computed over this whole test array revealed a clear propagation as well as the dispersion of surface waves over a large frequency range from 0.1 to a few Hertz (~8 Hz). This covers the frequency range of interest for our study, including the fundamental resonance frequency $f_0$ of the canyon (~0.5 Hz at BOLL, Gélis et al., 2022) and frequencies higher than 1 Hz for SHA and engineering applications. The design of the 400-node experiment has been refined following the analysis performed on the noise test array. The final design shown in Figure 1(b) is a compromise between 1) the need to cover the entire area of interest (array aperture), 2) the desired resolution (interstation distance) and 3) the number of instruments available. Note that a similar analysis has been conducted on the final dataset (400 nodes) and some results are shown in the section discussing the quality of the nodal dataset.

### Main Campaign

After the preliminary experiment, 409 nodes were deployed during the main campaign. The node array design for the main deployment is a combination of 5 sub-arrays:

- a loose grid covering the entire area composed of 164 nodes following East-West shifted lines of some 10 nodes. Node separation in this loose grid ranges from 400 to 1300 m and averages about 800 m.
- a denser grid located 2-to-3 km south of TNS, expected to cover the narrowest part of the Messinian Canyon as deduced by Gélis et al. (2022). This denser grid

is composed of 179 nodes spaced 200 to 250 m apart, deployed along roads and rural tracks.

- 2 dense East-West Lines following 2 roads designed to provide a denser coverage of the northern part of the study area, where the sedimentary fill is expected to be broader than in the southern part. The northern (resp. southern) line is composed of 29 (resp. 31) nodes separated by about 400 m.
- 6 more nodes were deployed out of our target area. One of these distant nodes was placed right on La Rouvière fault that broke during the Le Teil earthquake. The five others were deployed a few km away from our zone covering different azimuths. These sensors may be used as distant virtual sources (seismic interferometry applications) to illuminate the array with incoming wavefield from different directions.

It is worth noting that the first 2 digits of the station codes correspond to the codes of these sub-arrays. Further explanation about the station codes used for the 2 node deployments (preliminary noise test and the massive campaign) is given in electronic supplement.

We used the same 3-component Geospace GSX nodes as the ones deployed during the noise test. The nodes have been installed on public land, that is, mainly along roads. The deployment took place from February 17 to February 20, 2020. Instruments remained on field for one month and were de-installed between March 16 and 18. 402 nodes have been retrieved. 23 of them were found unburied. The data of the last days are therefore not exploitable for these nodes.

## BROADBAND STATION DEPLOYMENT

To complement the first dense and short-term campaign, a second campaign was carried out. This second campaign consisted in deploying 49 broadband stations over the same target area (Figure 1(c)) for at least six months and aimed at recording the seismicity, including local, regional and teleseismic events. 47 sites were instrumented with DATA-Cube[3] and 3-component Trillium compact 120s, and 2 sites were instrumented with Guralp CMG-6TD. Of the total 49 stations, 3 were deployed in sites that have been instrumented since 2016 (Table 1): BOLL (E01 in this survey) was instrumented with a DATA-Cube[3] and a Trillium Compact 120s, while PAUL and ADHE (E04 and G06 in this survey, respectively) were instrumented with Guralp CMG-6TD. All stations recorded continuously with a sampling frequency of 100 Hz. Installation sites were chosen following several criteria:

- Intent of catching the spatial ground motion variability expected from the overall geometry of the sedimentary valley (middle versus edges of the valley, small versus large sediment thickness);
- Instrumentation of different “rock” sites that could be considered as reference for the estimate of the amplification due to the sedimentary canyon. This implies the instrumentation of outcrops of various geological series, the canyon dug into. We finally instrumented 4 sites located on Urgonian hard limestones

(Lower Cretaceous formation in Figure 1) covering different azimuthal directions (A04 and A06 to the West, G06 to the East and D06 to the north). We also instrumented Miocene outcrop (G03) of the Saint-Restitut hill made of ten to twenty meters of calcareous sandstones. Note that this site is located on a high topography that could generate some topographic site effects. Other sites such as F02, and G01 have been settled on Cretaceous marly sands and sandstones (Upper Cretaceous formation in Figure 1). Near La Garde-Adhémar village, C06 was installed on Oligocene lacustrine limestones.

- The rest of the stations were deployed on the recent quaternary fluvial terrace (generally 10-20 m thick) overlying the Pliocene fill of the Messinian canyon or locally Upper Cretaceous marls, sands and sandstones (Lapalud town area);
- As for the node experiment, we instrumented the La Rouvière fault by installing 3 stations (RFN, RFC and RFS) along the rupture of the Le Teil earthquake;
- Sites as quiet as possible (by trying to get the station installed as far as possible from obvious noise sources);
- Satellite visibility for GPS-controlled clocks;
- Priority to free-field installations to limit the impact of the structure on the recorded motion. Only 2 sites were finally located inside buildings (A0 in a school and G4 in the city hall of Saint-Paul-3-châteaux).

- In total, 8 stations of the broadband deployment correspond to sites which have been also instrumented during the node experiment. In addition to the 3 “historical” sites listed in Table 1, The 5 other sites are listed in Table 2.

The array was installed between September 14 and September 18, 2020 and de-installed at the end of May 2021 (May 25-27). About half of the sites were located on private property. For most of the sites, sensors have been buried in free field and placed over a small concrete plate base. For a few sites, the sensor could not be buried, either because it was located inside buildings – sites A0 and G4 (Figure 2(g)) – or because the site was located on very hard limestone slab – site A4 (Figure 2(b))–. For the latter case, the sensor was placed at the surface and protected by a bucket filled with some foam thermal insulation. For each station, we used 2 pasture fence batteries (9V- 160 Ah) connected in series. This installation was designed to power the station for at least 6 months (i.e. the duration initially planned) but was expected to allow for a longer experiment. We finally decided to keep the installation for more than 8 months. Given the displacement restrictions due to the covid-19 pandemic during the experiment, we went on field for maintenance only once throughout these 8 months (end of January 2021). During this maintenance visit, we collected the data and checked the overall installation. We re-installed 2 stations that were found unburied and changed some batteries showing a low voltage.

# NODAL DATASET: OVERALL QUALITY AND INITIAL OBSERVATIONS

## GENERALITIES ABOUT THE DATASET

The data were resampled at 50 Hz leading to a total volume of about 1 TB for the main campaign and about 20 GB for the noise test. These data will be made publicly available in Fall 2023 on the French RESIF datacenter. A FDSN network code (XG) and a DOI (https://doi.org/10.15778/RESIF.XG2020) have been assigned (Froment et al., 2023).

## DATA COMPLETENESS

For the nodes that were found unburied, we visually checked the data to identify the day the sensor was dug up and we removed the files corresponding to days after that date from the dataset. Over the 400 nodes retrieved, 46 nodes (i.e. 11,5%) provide an incomplete dataset between the end of the deployment (February, 20) and the beginning of the deinstallation (March, 16). All these cases correspond to a premature stop in recording (no intermediate gaps were observed). Figure 3(a) shows the availability for these 46 nodes (the rest of the dataset is complete over the experiment duration). The overall data collection reaches more than 96% of completeness between February 20 and March 16.

## COMPARISON BETWEEN NODE AND BROADBAND RECORDINGS

This section focuses on the evaluation of the performance of the easy-to-deploy nodes, in particular below the instrument's natural frequency of 5Hz. Previous studies have discussed this aspect, for example within the SRL focus section on Geophone Array Seismology (e.g. Karplus and Schmandt, 2018). To address this issue in our context, we perform a comparison between signals recorded by co-located node and broadband CMG6-TD during the preliminary *noise test* experiment at ADHE. It is worth noting that BOLL was also instrumented with a Trillium compact sensor during the second campaign. Figure S2 in the electronic supplement shows also a comparison of PPSDs between the 2 broadband instrumentations involved in our experiment, although the recording period is different (November 2019 and 2020).

Using data from the noise test dataset allows us to compare waveforms of the Le Teil local earthquake. Figure 4 (a) and (b) shows a waveform comparison between the two instrumentations on November 11, 2019 (Le Teil earthquake, (a) ; 5-min noise window, (b)). Signals have been corrected from the respective instrument's response. Guralp CMG6-TD have been corrected using their own station calibration information. Regarding the nodes, a common correction has been applied for the whole pool deduced from instrument characteristics (frequency, gain). The earthquake waveforms filtered between 0.2 and 20 Hz recorded by the 2 instruments show a very good agreement both in phase and amplitude. A small discrepancy in amplitude is visible on the vertical component. This is explained by a slight clipping on this component for the

CMG6-TD. This comparison suggests that the GMG6-TD and the node recorded nearly identical waveforms down to frequencies much lower than 5Hz. As an example, a similar comparison for an aleatory 5-min noise window picked during nighttime is also shown (Figure 4(b)). This shows that the very good agreement between recordings is not limited to large-amplitude signals but is still observed for low-amplitude noise.

We also computed probabilistic power spectral densities (PPSD) using 1-hour windows with no overlap, over the 2 weeks of recording. Figures 4(c) and (d) show the PPSD for the two instruments at ADHE. Overall, the PPSD of the nodes for 1-hour windows match the broadband seismometer in shape. In detail, we can distinguish 3 ranges of frequency. For frequencies higher than 0.2 Hz, the comparison between the two instruments is very good. In this frequency range, the noise level is quite low at ADHE, due to its location on hard rock (no amplification due to geology) and in an isolated, very quiet area. Between 0.1 and 0.2 Hz, the PPSDs remain quite similar between the two instruments but show some slight differences. At low frequency (below 0.1 Hz), the noise level is getting higher than the New High Noise Model (NHNM, from Peterson, 1993) for the two instruments. On the horizontal components, one may see the influence of an imperfect protection from environmental changes and of the resulted tilt changes. Figure S2 shows that this effect is also observed on the Trillium Compact. On the vertical component, the noise level is slightly lower but the absence of variation suggests that the instrumental noise dominates in this band for the two instruments. Note that the overall noise level is significantly higher at BOLL (see Figure S2 in the electronic supplement) because of the location of this site (on top of the sedimentary fill whose local resonance frequency is about 0.5 Hz and within an industrialized environment).

Our different observations show that node signals reproduce CMG-6TD signals down to 0.2 Hz in terms of waveform comparison (local M4.9 earthquake and low-amplitude ambient noise) and statistics of 1-hr noise window amplitude. This analysis supports the possibility to exploit the node recordings at frequencies lower than 5 Hz. This result is of particular importance within the framework of the DARE project since it aims at characterizing the seismic site response (frequency range ~0.1-10 Hz), within a sedimentary canyon whose fundamental resonance frequency is significantly below 5 Hz (~0.5 Hz). This instrumentation comparison was a key aspect in the analysis of the preliminary noise test before the launch of the massive experiment.

## Dataset Control Quality

In order to get a rapid overview of the continuous recording at each node, we built a catalogue gathering different representations of the monthly seismic signal (temporal waveform, spectrogram and spectral density on 10-minute segments). This catalogue provides an easy way to explore basic features of the dataset. This makes it possible to identify signals and/or nodes presenting obvious issues (Figure 5 shows the catalogue sheets for ADHE, BOLL and a node presenting major issues). By doing this, we identified less than 1% of the 1-month recordings as unusable (as the example shown in Figure 5 -top row-). For the rest of the dataset, we consider that the 1-month recording may be analyzed at least for part of the 10-minute segments and/or in a limited frequency range. Labelling the quality of seismic signals is not trivial since it is strongly application-dependent. Therefore, we do not go further here quantifying the quality of the data since this needs to be addressed in relation to specific applications and will come with associated studies based on these data. The complete catalogue (i.e. for the 400 nodes) for the North, East, and vertical component is available respectively in Files S1, S2 and S3 in the electronic supplement.

Note that the analysis of this catalogue was the basis to investigate the impact of numerous cultural noise sources in this industrialized area on the continuous data. This is the scope of ongoing work within the DARE project (Gisselbrecht et al., submitted to Geophysical Journal International).

## PROPAGATION RECONSTRUCTED FROM NOISE CORRELATION FUNCTIONS

Figure 6 presents stacked sections of noise correlation functions (NCFs) computed for all the station pairs of the 400-node array. This representation is useful to give an estimate of distance ranges over which one can expect to extract coherent wavefields at different frequencies. To compute NCFs, continuous data were first split into 30 min segments. Each segment was then spectrally whitened. NCFs were computed for each 30-min segment and then stacked over the entire recording time (1 month). Averaged seismic sections shown in Figure 6 are constructed by binning NCFs in fixed distance intervals (every 100 m). Note that symmetrized NCFs are plotted, that is, the mean of the negative and positive lag-times. Figure 6 reveals the wave propagation reconstructed from NCFs in 2 frequency bands (0.1-1 Hz; 1-10 Hz) and at 2 spatial scales (the entire array and the denser part in the southeastern zone of the array).

At low frequency (i.e. <1Hz, Figure 6 - top panel), one can observe a clear propagation over regional distance (25 km; i.e. between the node located on La Rouvière fault and all the other nodes) with a frequency content dominated by the secondary microseismic peak. On the TT component, higher frequencies (0.5-1 Hz) are visible revealing the dispersion of Love waves, as well as more complex patterns associated with the propagation over the first 10-12 km (i.e. the core of our target zone). The middle panel in Figure 6 shows that the propagation of waves at frequencies higher than 1 Hz can be tracked on the stacked NCFs over about the same distance (10-12 km) but is clearer over a distance of about 5 to 6 km. NCFs computed only on the densest part of the array allows us to zoom in on shorter distances (Figure 6 – bottom panel). One can clearly see the dispersion of both Rayleigh (ZZ, RR components) and Love (TT component) waves. Multiple branches (multiple modes), ruptures in slopes (rapid changes in velocities, see for example around 3-3.5 km) and differences on the different components reveal a complex medium. It is worth noting that the interpretation of these sections in terms of structure is limited since this spatially averaged representation mitigates propagation patterns due to lateral heterogeneities.

# BROADBAND DATASET: OVERALL QUALITY AND INITIAL OBSERVATIONS

## Generalities about the dataset

The broadband dataset acquired in this study has already been uploaded into the GEOFON data archive under network code Y7 (Pilz et al., 2021). Free access to this dataset will be available at the end of the DARE project (end of 2023). The complete dataset has a size of 463 GB and includes the data for all usable broadband stations (one station has been tagged as faulty, see discussion in the next section), with a sampling rate of 100 Hz.

## Data completeness

The temporal availability of the broadband data is shown in Figure 3(b). It is good in general terms, taking into account the originally planned duration of 6 months for the deployment. During the field maintenance trip in January 2021, we used the available spare equipment to replace the batteries of the stations that were reporting the lowest voltages (B00, D01, D03, E02 and G03). The new batteries allowed these stations to keep recording for up to 2 months longer than originally planned. Some 20 stations continued to operate for the entire period from September 2020 to the end of May 2021. Considering this 8-month time period, data availability varies significantly from 99.18% (D00) to 19.73% (A06), with an average of 77.62%. The most important data gaps not related to battery failure belong to stations A06, B00 and B01 and B02. In the case of A06, an important part of the records was lost probably due to a faulty SD card. Station B00 was disconnected after the initial deployment and the sensor was tilted, which was fixed during the field maintenance trip at the end of January, 2021. Station B01 lost power in November and was reconnected during field maintenance. B02 stopped recording prematurely in early February 2021 after a flood in the Rhône river drowned the station. The remaining data gaps are much smaller and can mainly be attributed to short, temporary losses of GPS signal or problems when attempting to read the SD cards retrieved from the field. For one station none of the recordings are usable, i.e. A05, and therefore has not been included in the Y7 network in the Geofon data archive.

# DATASET QUALITY AND INITIAL OBSERVATIONS

## Reorientation of the broadband seismic sensors

Many seismological studies are sensitive to the correct orientation of the horizontal axes of the seismic sensors. However, the equipment that is required to perform a precise determination of the orientation of the horizontal components during field work is often costly and difficult to operate (e.g., Ringler et al., 2013). Therefore, this task is usually accomplished using a magnetic compass, which might introduce non-negligible errors (Wang et al., 2016). It is worth noting that the local declination is 1°54’. To account for the orientation errors in this survey, we have analyzed the arrival angles of teleseismic Rayleigh waves following the approach described by Doran & Laske (2017). Orientation angles are obtained through a grid-search procedure at seven discrete frequencies between 0.01 and 0.04 Hz in order to minimize any possible bias caused by the local laterally heterogeneous earth structure.

The angles obtained for the North component of the broadband stations, measured clockwise from true north, are shown in Figure 7 and listed in Table S1 in the electronic supplement. Most sensors were correctly oriented during installation as shown by the average deviation of 9.3°, with only few stations showing deviations higher than 15° (A01, B00, C01, C04, G06, E04).

## Noise levels across the broadband array

The main purpose for the broadband array deployment was to survey the seismicity. In general, the quality and utility of seismic data is strongly dependent on the background ambient noise levels at each site. This is particularly true for industrialized regions such as our study area. To characterize the ambient noise levels at each of the broadband array sites, we divided one month of continuous records into 1-hour segments with half-hour overlap and computed the power spectral density (PSD) for each segment. Then, we created spectrogram-like plots showing the temporal variation of the PSDs for the three components (examples are shown in Figure 8). We used these spectrogram-like plots as the features for a k-Means clustering algorithm (e.g. Lloyd, 1982) with the aim of classifying each site based on the overall noise levels. This analysis allowed us to identify three types of sites: 1) overall low noise levels, 2) high noise levels at short periods (< 1 s), and 3) high noise levels both at short (< 1 s) and long periods (> 30 s). Note that the spectral content of the broadband recordings will be discussed in terms of periods (i.e. in seconds) in the following. The top left panel in Figure 8 shows the variation of the PSD over time for an example of each type of site: A04, E03 and D04 for sites of type 1), 2) and 3), respectively. At short periods (< 1 s) the PSDs show a clear daily and weekly pattern, related to human activity (e.g., Groos and Ritter, 2009). At periods ranging from approximately 2 to 8 seconds, the microseismic frequency band can clearly be identified in all stations. The intensity and frequency range of the microseismic noise varies with time and increases towards the winter months. In the long period range (> 30 s) the characteristics of the noise are site-dependent and do not show any clear temporal patterns. The highest levels of noise at long periods appear predominantly in the horizontal components (HHE and HHN), with practically all stations showing higher noise levels than the NHNM (Peterson, 1993), and increase steadily with increasing period. Long period noise with similar characteristics has often been interpreted as the result of seismometer tilting, i.e. tilting of the sensors from the level position by a certain angle (e.g., Rodgers, 1968; Wielandt and Forbriger, 1999; Rhode et al., 2017). Tilt sources can be varied, ranging from changes in atmospheric pressure to moving vehicles and buildings under wind load in urban environments (e.g., Rhode et al. 2017; Forbriger, 2007).

The results of the ambient noise-based clustering are summarized in Figure 8. The map in the top right corner shows the broadband station sites colored by site type. The plots in the bottom row of Figure 8 show the mean of the probabilistic power spectral density function (PPSD) estimated from all the available PSDs for each station. Visual inspection of the mean of the PPSDs also supports the classification of the sites in three different clusters or categories. The correlation between the geographical location of the stations and their noise levels is not completely straightforward and suggests that the noise level is significantly variable at the local scale, probably strongly related to very local environment but also to geology, thereby explaining the observed broadband trends. This is particularly pronounced for the lowest noise levels (e.g. cluster 1, blue color in Figure 8- that shows a very good agreement with areas with no Pliocene sedimentary fill) despite being located in very different environments (i.e. urban environment for D06; isolated clearing in a forest for A04).

### On the recorded seismicity

After reviewing the ambient noise levels across the broadband array we selected station A04, deployed on a rock site outside of the valley and one of the quietest stations in the array, to elaborate a seismicity catalogue. The starting point was the ISC (International Seismological Centre) catalogue. A criterion based on lower-bound magnitude thresholds (relative to the epicentral distance) was used as a preselection, followed by a signal-to-noise-based selection and finally a visual inspection. The derived catalogue of 424 events is given in File S4 in the electronic supplement and illustrated in Figure S3. Note that this catalogue covers the lifetime span of A04 (i.e. until April 2021, see Figure 3(b)).

Figure 9 shows two examples of the varied seismicity recorded by the broadband network. The first example is a regional earthquake (Mw5.0 from September 30, 2020 in the Pyrenees). The top row in Figure 9 contains two plots showing the event for the complete duration and a close-up of the P-wave onset (Figures 9(a) and 9(b), respectively). The second example belongs to a teleseismic earthquake (Mw6.3 from September 18, 2020 in the Central Mid Atlantic Ridge). The waveforms for the complete duration of this event and a close-up on the P-wave onset are shown in the bottom row in Figure 9 (9(c) and 9(d), respectively). The signal-to-noise ratio of most stations is good for these kinds of events. The noisiest stations are often the ones located in the vicinity of the TNS area, and to the busy A7 highway (e.g. E03, E04, F03).

# SUMMARY

The datasets presented in this paper provide complementary seismic data in terms of spatial and temporal scales as well as instrumentation (a dense 1-month nodal experiment versus a few-month campaign of broadband stations). These different designs aimed at targeting a variety of seismic data and signals, including the recordings of the ambient noise, a local moderate earthquake (the 2019 Mw4.9 Le Teil earthquake) as well as regional and teleseismic seismicity. The first analysis made on these data and gathered in this paper provides information on the characteristics and the overall quality of these data that would be helpful for future users. These complementary data will be used in the framework of the DARE project to characterize the complex local sedimentary structure and its impact on the seismic motion. They will be of great interest to provide an extensive site-specific seismic study related to a deep valley in an industrialized area hosting critical infrastructure. In particular the idea is to consider different approaches based, on one hand, on numerical simulations of the ground motion in a model of the Earth's sub-soil (i.e. numerical approach), and on the other hand, on the direct analysis of recordings of seismic motions to estimate the site effects (i.e. empirical approach); both methods requiring to be constrained by seismic data. We will also benefit from ongoing studies to establish an accurate geological model in the area. More generally, this project has the objective to provide an example of the interest of acquiring and exploiting seismic data for seismic hazard operational applications in low-to-moderate seismicity areas and deep valleys. These datasets will be made available at the end of 2023.

## DATA AND RESOURCES

The nodal dataset (Froment et al., 2023; doi: https://doi.org/10.15778/RESIF.XG2020) will be in free access on the French RESIF datacenter (https://www.resif.fr/en/) at the end of the DARE project (end of 2023). The broadband dataset (Pilz et al., 2021; doi: http://doi.org/10.14470/L27575187372) will be in free access on the German GEOFON datacenter (GEOFON Y7 Seismic Network (https://geofon.gfz-potsdam.de/)) also at the end of 2023.

The python Toolbox ObsPy was used for processing the seismological data (Beyreuther et al., 2010). The ArcGis Software was used for map representations.

Supplemental Material for this article includes:

- A map showing the main anthropogenic elements of the area

- A description of the station codes used for the nodal deployment
- A comparison of PPSDs between the broad-band instrumentations used in these seismic campaigns (i.e. Guralp CMG6-TD and Trillium Compact+DATA-Cube3)
- The Quality Control catalog built for the nodal dataset
- A table listing the estimated error in the orientation of the N-component
- The catalogue of seismicity recorded at A04 station during the broad-band campaign

# DECLARATION OF COMPETING INTERESTS

The authors declare no competing interests

# ACKNOWLEDGEMENTS

This work was funded by a public grant overseen by the French National Research Agency (Grant # ANR19-CE31-0029) and the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project number 431362334).

We thank the private land owners and communities that have hosted our seismic stations and/or nodes. We also thank colleagues who come on field for station deployment and/or maintenance visits (Nick Arndt, Charles Beard, Florent Brenguier, Christophe Clément, Roméo Courbis, Sébastien Hok, Aurore Laurendeau, Anaïs Lavoué, Illdut Pondaven, Ludmila Provost, Flomin Tchawe Nziaha).

We also thank RESIF and GFZ technical staff for the support in formatting the database.

Finally, we thank the associate editor and two anonymous reviewers for their constructive comments and their help to improve the article.

# REFERENCES

Bagayoko, N. (2021), Intégration de données de sismique réflexion et de données géologiques dans un modèle structural 3D du canyon messinien du Rhône : Etude du site de Tricastin, *M.Sc. Thesis*, *Sorbonne Université*.

Bard, P. Y., and M. Bouchon (1985). The two-dimensional resonance of sediment filled valleys. *Bull. Seism. Soc. Am.* **75** 519–541.

Cornou, C., J.-P. Ampuero, C. Aubert, et al. (2021). Rapid response to the Mw 4.9 earthquake of November 11, 2019 in Le Teil, Lower Rhône Valley, France. *Compt. Rendus. Geosci.* **353**(S1), 441-463, https://doi.org/10.5802/crgeos.30.

Doran, A. K. and G. Laske (2017). Ocean-bottom seismometer instrument orientations via automated rayleigh-wave arrival angle measurements. *Bull. Seism. Soc. Am.* **107**(2) 691-708.

Forbriger, T. (2006). Low-frequency limit for H/V studies due to tilt, *32nd Meeting of the Working Group Seismology of the FKPE, 4-6.10.2006, Haidhof,* Germany, 10.5445/IR/1000007819.

Froment, B., E. M. Cushing, C. Gélis, L. Gisselbrecht, S. Beauprêtre, A. Tourette, and RESIF (2023). France 2020, Dense nodal seismic array in the Rhône Valley, DARE project [Data set]. RESIF - Réseau Sismologique et géodésique Français. https://doi.org/10.15778/RESIF.XG2020.

Gélis, C., L. Cauchie, E.M. Cushing, B. Froment, S. Franco, H. Jomard, D. Moiriat, L. Provost, B. Sariguzel, and H. Tebib (2022). Estimation of the local seismic amplification on an industrialized site in the French Rhône Valley, *Pure Appl. Geophys.*, in press.

Groos, J. C. and J. R. R. Ritter (2009). Time domain classification and quantification of seismic noise in an urban environment. *Geophys. J. Int.*, **179**(2), 1213-1231. doi:10.1111/j.1365-246X.2009.04343.

Karplus, M. and B. Schmandt (2018). Preface to the Focus Section on Geophone Array Seismology. *Seism. Res. Lett.* **89**(5) 1597–1600. doi: https://doi.org/10.1785/0220180212

Kawase, H. (1996). The Cause of the Damage Belt in Kobe: 'The Basin-Edge Effect,' Constructive Interference of the Direct S-Wave with the Basin-Induced Diffracted/Rayleigh Waves. *Seism. Res. Lett.* **67**(5) 25–34. https://doi.org/10.1785/gssrl.67.5.25.

McNamara, D. E. and R. P. Buland (2004). Ambient noise levels in the continental united states. *Bull. Seism. Soc. Am.* **94**(4) 1517–1527.

Paolucci, R., M Aimar, A Ciancimino, M Dotti, S Foti, G Lanzano, P Mattevi, et al. (2021) Checking the site categorization criteria and amplification factors of the 2021 draft of Eurocode 8 Part 1–1 Bulletin of Earthquake Engineering 19 (11), 4199-4234

Peterson, J. (1993), Observations and Modeling of Seismic Background Noise, U.S. Geological Survey open-file report 93-322, Albuquerque, N.M.

Pilz, M., & Cotton, F. (2019). Does the one-dimensional assumption hold for site response analysis? A study of seismic site responses and implication for ground motion assessment using KiK-Net strong-motion data. Earthquake Spectra, 35(2), 883-905.

Pilz, M., Cotton, F., Ohrnberger, M. Froment, B. (2021): DARE: Dense Array for seismic site effect Estimation. GFZ Data Services. Other/Seismic Network. Doi:10.14470/L27575187372.

Rhode, M. D., A. T. Ringler, C. R. Hutt, D. C. Wilson, A. A. Holland, L. D. Sandoval, and T. Storm, (2017). Characterizing Local Variability in Long-Period Horizontal Tilt Noise. *Seism. Res. Lett.*, **88**(3), 822-830.

Ringler, A. T., C. R. Hutt, K. Persefield, and L. S. Gee (2013). Seismic station installation orientation errors at anss and iris/usgs stations. *Seism. Res. Lett.* **84** 926–931.

Ritz, J.-F., S. Baize, M. Ferry, C. Larroque, L. Audin, B. Delouis, and E. Mathot (2020). Surface rupture and shallow fault reactivation during the 2019 Mw 4.9 Le Teil earthquake, *Nat. Comm. Earth Env.* **1**(10) https://doi.org/10.1038/s43247-020-0012-z.

Rodgers P. W. (1968). The response of the horizontal pendulum seismometer to Rayleigh and Love waves, tilt, and free oscillations of the Earth, *Bull. Seismol. Soc. Am*. **58**(5), 1385–1406.

Semblat, J. F., M. Kham, E. Parara, P. Y. Bard, K. Pitilakis, K. Makra, D. Raptakis (2005). Seismic wave amplification: Basin geometry vs soil layering. *Soil Dyn. Earthqu. Eng.*, **25**(7–10), 529–538. https://doi.org/10.1016/j.soildyn.2004.11.003.

Wang, X., Q. F. Chen, J. Li, and S. Wei (2016). Seismic sensor misorientation measurement using p-wave particle motion: An application to the necsaids array. *Seism. Res. Lett.*, **87**(4).

Wielandt E., and T. Forbriger (1999). Near-field seismic displacement and tilt associated with the explosive activity of Stromboli, *Ann. Geofis.* **42** (3), 407–416.

Bérénice Froment: berenice.froment@irsn.fr

*IRSN, BP 17, 92262 Fontenay-aux-Roses cedex, France*

Andrés Olivar-Castaño: andres.olivar-castano@uni-potsdam.de

*Institute of Geosciences, Campus Golm, Building 27, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany*

Matthias Ohrnberger: Matthias.Ohrnberger@geo.uni-potsdam.de

*Institute of Geosciences, Campus Golm, Building 27, Karl-Liebknecht-Str. 24-25, 14476 Potsdam-Golm, Germany*

Loic Gisselbrecht: loic.gisselbrecht@univ-grenoble-alpes.fr

*Université Grenoble Alpes, ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France*

Katrin Hannemann: katrin.hannemann@uni-muenster.de

*University of Münster, Institut für Geophysik, Corrensstr. 24, 48149 Münster, Germany*

Edward Marc Cushing: edward.cushing@irsn.fr

*IRSN, BP 17, 92262 Fontenay-aux-Roses cedex, France*

Pierre Boue: pierre.boue@univ-grenoble-alpes.fr

*Université Grenoble Alpes, ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France*

Céline Gélis: celine.gelis@irsn.fr

*IRSN, BP 17, 92262 Fontenay-aux-Roses cedex, France*

Annabel Haendel: ahaendel@gfz-potsdam.de

*Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, Building A 70, Telegrafenberg, 14473 Potsdam, Germany*

Marco Pilz: pilz@gfz-potsdam.de

*Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, Building A 70, Telegrafenberg, 14473 Potsdam, Germany*

Laura Hillmann: laura.hillmann@gfz-potsdam.de

*Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, Building A 70, Telegrafenberg, 14473 Potsdam, Germany*

Occitane Barbaux: occitane.barbaux-manpower@irsn.fr

*IRSN, BP 17, 92262 Fontenay-aux-Roses cedex, France*

Sophie Beauprêtre: Sophie.BEAUPRETRE@egis-group.com

*Sisprobe - EGIS geotechnique, 3 Rue du Dr Schweitzer, 38180 Seyssins, FRANCE*

Gilbert Bouzat: gilbert.bouzat@orano.group

*Orano Chimie Enrichissement, Site du Tricastin, 26700 Pierrelatte, France*

Emmanuel Chaljub: Emmanuel.Chaljub@univ-grenoble-alpes.fr

*Université Grenoble Alpes, ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France*

Fabrice Cotton: fabrice.cotton@gfz-potsdam.de

*Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, Building A 70, Telegrafenberg, 14473 Potsdam, Germany*

François Lavoué: francois.lavoue@univ-grenoble-alpes.fr

*Université Grenoble Alpes, ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France*

Laurent Stehly: laurent.stehly@univ-grenoble-alpes.fr

*Université Grenoble Alpes, ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France*

Chuanbin Zhu: chuanbin@gfz-potsdam.de

*Helmholtz-Zentrum Potsdam Deutsches GeoForschungsZentrum GFZ, Building A 70, Telegrafenberg, 14473 Potsdam, Germany*

Olivier Magnin: Olivier.MAGNIN@egis.fr

*EGIS geotechnique, 3 Rue du Dr Schweitzer, 38180 Seyssins, FRANCE*

Laurent Metral: laurent.metral@univ-grenoble-alpes.fr

*Université Savoie Mont Blanc, Campus Scientifique, 73376 Le Bourget-du-Lac Cedex, France*

Aurélien Mordret: aurelien.mordret@univ-grenoble-alpes.fr

*Université Grenoble Alpes, ISTerre, CS 40700, 38058 GRENOBLE Cedex 9, France*

Yann Richet: yann.richet@irsn.fr

*IRSN, BP 17, 92262 Fontenay-aux-Roses cedex, France*

Alexandre Tourette: Alexandre.TOURETTE@egis.fr

*EGIS geotechnique, 3 Rue du Dr Schweitzer, 38180 Seyssins, FRANCE*

Table 1: Information (naming and instrumentation) for the 3 « historical » sites instrumented during all the acquisitions mentioned in this study.

| | **Site 1** | **Site 2** | **Site 3** |
|---|---|---|---|
| Coordinates | 44.3743°N ; 4.7697°E | 44.3467°N ; 4.7357°E | 44.3010°N ; 4.7199°E |
| Historical acquisitions (2016-2019)<br>**Naming** [*instruments*] | **ADHE** [*Guralp CMG6-TD*] | **PAUL** [*Guralp CMG6-TD*] | **BOLL** [*Guralp CMG6-TD*] |
| DARE Noise Test | **ADHE** [*Guralp CMG6-TD*]<br>+<br>**60026** [*Geospace GSX node*] | **PAUL** [*Guralp CMG6-TD*]<br>+<br>**60029** [*Geospace GSX node*] | **BOLL** [*Guralp CMG6-TD*]<br>+<br>**60012** [*Geospace GSX node*] |
| DARE Acquisition 1 (Nodes) | **10011** [*Geospace GSX node*] | **16042** [*Geospace GSX node*] | **03085** [*Geospace GSX node*] |
| DARE Acquisition 2 (Broadband) | **G06** [*Guralp CMG6-TD*] | **E04** [*Guralp CMG6-TD*] | **E01** [*DATA-Cube + Trillium*] |

Table 2: Information regarding the common sites instrumented during the 2 main acquisitions of the DARE project. These 5 sites are in addition to the 3 historical sites listed in Table 1. The bottom row gives the approximate distance between the instruments deployed during the 2 acquisitions.

| | Site 1 | Site 2 | Site 3 | Site 4 | Site 5 |
|---|---|---|---|---|---|
| Coordinates of Broadband Site | 44.5240°N; 4.6574°E | 44.3016°N; 4.6706°E | 44.3029°N; 4.6941°E | 44.3778°N; 4.6981°E | 44.2889°N; 4.7334°E |
| Naming in DARE Acquisition 1 (Nodes) | **50006** | **25073** | **03041** | **10006** | **03162** |
| Naming in DARE Acquisition 2 (Broadband) | **RFC** | **C01** | **D01** | **D06** | **F01** |
| Approximate Distance (m) | < 5 | 20 | < 5 | 10 | 30 |

# LIST OF FIGURE CAPTIONS

**Figure 1:** Map of the 3 seismic deployments carried out. The left figure shows the design of the preliminary noise test (orange markers) as well as the three historical sites (triangles) instrumented by IRSN since 2016. The figure in the middle shows the design of the massive 400-node deployment with a very dense area contoured by the dashed black line. The right figure shows the design of the 49-broadband deployment. Grey markers represent deployed stations at which data turned out to be unusable (faulty stations). Note that the location of the Mw4.9 Le Teil earthquake (November 11, 2019) on La Rouvière Fault is also displayed in the northern part in each plot.

**Figure 2:** Co-located Geospace GSX node and Guralp GMG6-TD (a) and different conditions of installation during the broadband campaign (b-g): (b) on hard limestones outcrop preventing burial in the ground (A04); (c) on a residential neighborhood (G00); (d) in the remote garrigue (C06); (e) in the town of Pierrelatte (D06); (f) in a farm (B06); (g) in the city Hall of Saint Paul Troix Châteaux (G04).

**Figure 3:** (a) Temporal availability for the 46 nodes that stopped before the deinstallation of the massive node-deployment (the rest of the array operating correctly for the duration of the experiment). The dashed vertical lines (and thick grey line on top of the figure) show the beginning and end of common recording period for the complete array. (b) Temporal availability for the seismic stations deployed during the broadband campaign. Light grey lines represent unusable data from faulty stations. For clarity, only

the vertical components have been displayed.

**Figure 4:** Comparison between recordings of a GSX node and a co-located Guralp CMG6-TD at site ADHE. (a-b) Waveform comparison for the Le Teil local earthquake signal (a) and a 5-minute noise window (b). Signals have been filtered between 0.2 and 20 Hz. (c-d) PPSD comparison between the node (c) and CMG6-TD (d) recordings.

**Figure 5:** Example of a quality catalogue sheet (North component) for 3 nodes: a node presenting major issued (top), node 03085 located at BOLL site (middle) and node 10011 located at ADHE site (bottom). A description of this catalogue is available in the electronic supplement.

**Figure 6:** Stacked sections of noise correlation functions (NCFs) computed for all the station pairs of the 400-node array filtered between 0.1 and 1 Hz (top figure) and between 1 and 10 Hz (middle figure). The bottom figure shows the same figure (1-10 Hz filtering) using only pairs of stations located within the densest zone of the deployment (see Figure 1(b)). For clarity only the diagonal components of the NCF tensor are displayed. As indicated in the top right figure, the lines represent velocity lines of 0.25, 0.5, 1, 2.5 and 5 km/s.

**Figure 7:** Quiver plot showing the orientation of the nominal North component of the broadband sensors deployed in this work. The legend for the geological map is the same as for Figure 1.

**Figure 8:** Top left: Example of spectrogram-like plots used as features for the k-Means clustering for three stations (A04, E03 and D04) representing each cluster. Top right: broadband station locations color-coded showing the clustering of the overall ambient noise levels as described in the text. The legend for the geological map is the same as for Figure 1. Bottom row: mean of the overall PDF for all stations for each channel, again colored by cluster. Note that the analysis has not been done for E02 that presents only a few days of usable data (see Figure 3(b)).

**Figure 9:** Waveforms recorded by the broadband network for a Mw5.0 regional earthquake (Pyrenees; top row) and a Mw6.3 teleseismic earthquake (Central Mid Atlantic Ridge; bottom row). A bandpass filter with corner frequencies 0.05 and 0.5 Hz was applied to the data. The stations are sorted by epicentral distance.