Insights into the conditions of application of noise-based spectral ratios in a highly industrialized area: a case study in the French Rhone Valley

Loic Gisselbrecht, Bérénice Froment, Pierre Boué, Celine Gelis

Geophysical Journal International

Geophys. J. Int. (2023) **234,** 985–997
Advance Access publication 2023 March 11
GJI Seismology

https://doi.org/10.1093/gji/ggad108

Insights into the conditions of application of noise-based spectral ratios in a highly industrialized area: a case study in the French Rhone Valley

Loïc Gisselbrecht [1,2] Bérénice Froment,[1] Pierre Boué[2] and Céline Gélis[1]

[1] *Institut de Radioprotection et Sûreté Nucléaire (IRSN), PSE-ENV, SCAN, BERSSIN,* 31 *Avenue de la Division Leclerc, Fontenay-aux-Roses* 92260, *France. E-mail: loic.gisselbrecht@irsn.fr*
[2] *Université Grenoble Alpes, Université Savoie Mont Blanc, CNRS, IRD, UGE, ISTerre,* 1381 *Rue de la Piscine, Gières* 38610, *France*

Accepted 2023 February 27. Received 2022 December 20; in original form 2022 August 1

SUMMARY

The local ground motion amplification related to the geology at a specific site (i.e. the so-called site effects) may be classically quantified through the SSR (standard spectral ratio) technique applied on earthquake recordings. However, such a quantification might be challenging to carry out in low-to-moderate seismicity regions. Methods based on background ambient noise, such as noise-based standard spectral ratio (SSRn), might be of great interest in these areas. But noise-derived amplification is particularly sensitive to local anthropogenic sources, which may introduce biases in the evaluation of site effects by dominating the geological effects, especially for frequencies higher than 1 Hz. A hybrid approach (SSRh), developed to reduce biases in noise-based spectral ratios by combining classical earthquake-based spectral ratio (SSR) and SSRn, was recently introduced and relies on a site reference. We here investigate the applicability of SSRn and SSRh in a heavily industrialized environment in the Tricastin Valley (south-east France), where critical facilities are located. We continuously recorded ambient noise from 2020 February to March on a 400-sensor seismic array covering an area of about 10 km by 10 km. We demonstrate that SSRn and SSRh computed below 1 Hz are able to reproduce amplification factors provided by SSR. By contrast, at frequencies higher than 1 Hz, SSRn strongly deviates from SSR. SSRh shows closer results to SSR but presents a dependence to the choice of the local site reference, thereby questioning the possibility to use SSRh blindly to estimate the local amplification in our context. These discrepancies reflect the impact of local anthropogenic sources. We therefore introduced a two-step workflow to mitigate the influence of local sources. The first step is to define a characteristic time window to optimally isolate significant transient signals. The second step consists in selecting the time segments that do not contain these transients with a clustering-based approach. By applying this workflow, we were able to remove some strong anthropogenic transient signals likely to be generated by local sources at some sites and therefore to locally improve the amplification assessment through noise-based spectral ratios. However, stationary sources, whose impact cannot be removed through our procedure, remain a major issue. This study provides some insights into the application of SSRn and SSRh in noisy industrialized areas, especially regarding the impact of local noise sources. It illustrates the difficulty of having a procedure for mitigating the impact of these sources that is efficient everywhere inside such a complex anthropized environment, where different types of sources (including stationary sources) cohabit.

Key words: Fourier analysis; Earthquake ground motions; Earthquake hazards; Seismic noise; Site effects; Spectral ratio.

1 INTRODUCTION

It is now well known that shallow complex geological structures can modify seismic motion. From a seismic hazard point of view, this modification of ground motion caused by local geological structures is usually referred to as site effects. In the case of sedimentary basins (or valleys), these site effects are mainly caused by the impedance contrast at the interface between soft soil layers filling the

sedimentary basin and the bedrock. Several phenomena can thus take place within the basin, including body wave resonance, edge-generated surface waves and focusing effects. These phenomena can lead to a dramatic ground motion amplification and an increase in the duration of shaking when an earthquake occurs (e.g. Bard & Bouchon 1985; Celebi *et al.* 1987; Kawase 1996; Semblat *et al.* 2005). As cities and critical infrastructures may be located on top of such geological structures, it is of great importance, for the purpose of seismic hazard assessment at these specific sites, to have a reliable method in predicting the local amplification that would occur during an earthquake.

1.1 Earthquake and noise-based spectral ratios for local amplification assessment

Several approaches have been introduced since the 1970s to assess the local amplification associated with site effects. The well-known SSR (standard spectral ratio, Borcherdt 1970), has been widely used since its introduction as it provides reliable estimates of amplification factors (e.g. Field & Jacob 1995; Raptakis *et al.* 1998; Huang & Teng 1999). This method consists in computing the ratio of the Fourier amplitude spectra (FAS) recorded at two stations for a given earthquake. Local amplification is estimated relatively to a reference station. The numerator FAS corresponds to the ground motion observed at a station located at the site under study (i.e. on top of the sedimentary filling) and the denominator FAS corresponds to a rock-reference station. The choice of the reference station is sometimes not straightforward, as it must ensure that site effects are negligible at this station, and has been discussed in numerous papers (e.g. Steidl *et al.* 1996). The inter-station distance must be small with respect to the earthquake distance, so that both stations are supposed to experience the same source and path effects on the seismic motion. As a result, the source and path effects cancel out in the SSR, thereby isolating the site effects. This can be summarized as follows:

$$\mathrm{SSR}_{s/r,k}(f) = \frac{\mathrm{FAS}_{s,k}}{\mathrm{FAS}_{r,k}} = \frac{S_k(f) \times P_k(f) \times H_{s,k}(f)}{S_k(f) \times P_k(f) \times H_{r,k}(f)} = H_{s,k}(f), \quad (1)$$

where f is the frequency. $\mathrm{SSR}_{s/r,k}(f)$ is the SSR between station s located at the site under study and the rock-reference station r, for an earthquake k. $S_k(f)$ and $P_k(f)$ are, respectively, the source and path terms. $H_{s,k}(f)$ is the amplitude of the site transfer function at s. By assuming that site effects are negligible at the rock-reference station, $H_{r,k}(f)$ is considered to be equal to 1. Therefore $H_{s,k}(f)$ estimates the local ground motion amplification factor between s and r.

SSR technique may be however difficult to apply in low-to-moderate seismicity areas as it requires the recordings of many earthquakes with a good signal-to-noise ratio over a wide frequency range to get a statistically robust estimate (e.g. Perron *et al.* 2020). Ambient noise on the other hand appears to be an appealing source of data in low-to-moderate seismicity zones. This weak seismic wavefield, generated by natural or anthropogenic sources, has the characteristics of being permanent and exploitable everywhere on the surface of the Earth. Therefore, ambient noise-based approaches may be of great interest where good quality earthquake recordings are challenging to get. In this context, Irikura & Kawanaka (1980) introduced the noise-based standard spectral ratio (SSRn), which is equivalent to the SSR (eq. 1) but relies on ambient noise recordings. The first studies using this method for the purpose of site responses assessment were these of Kagami *et al.* (1982, 1986). They applied the SSRn using long-period microtremor recordings (1–5 s) in the Niigita Plain (Japan) and in Los Angeles (California, USA), where the geological properties at the surface may have played an important role in the strong ground motion observed during the M_w 7.6 Niigita and M_w 6.6 San Fernando earthquakes, which respectively occurred in 1964 and 1971. The amplification estimated through SSRn was found to be qualitatively consistent with the local sedimentary deposit, and the authors concluded that this easy-to-use method should be applied to other surveys at low frequencies. Later, several studies qualitatively compared SSRn results, either with the geology of the site under study (e.g. Ferritto 1995), or with the SSR amplification factor (for an extended review, see for instance Bard 1999 or Perron 2017). Most of these studies have shown that SSRn can provide comparable measurements to SSR at low frequency but largely overestimates the amplitude of the site transfer function obtained through SSR approach at higher frequencies if the site to reference distance is greater than several hundreds of metres (e.g. Zhao *et al.* 1998). This discrepancy is generally explained by local anthropogenic noise sources that are close to the sensors at the site, but far from the rock-reference sensor (e.g. Horike *et al.* 2001). Hence, the source signature at both locations is different and the method is failing in isolating the amplitude of the site transfer function. In other words, the background noise wavefield is not necessarily fulfilling the required hypothesis underlying the simplifications of eq. (1) which occur for a distant earthquake and are at the basis of the SSR approach.

Perron *et al.* (2018) introduced a new approach combining SSR and SSRn techniques for site effects estimation in sedimentary basins and that may mitigate such limitations. The bottom line of the so-called SSRh (hybrid standard spectral ratio, eq. 2) is first to use earthquake-based SSR to measure the amplification at an intermediate station $sref$ (located within the sedimentary basin) relative to a rock-reference station. Then, the amplification variability between $sref$ and a station sx in the basin is estimated through site-to-site noise-based SSRn computation. $sref$ is therefore used as a local site reference. This might be expressed as:

$$\mathrm{SSRh}_{sx/r} = \mathrm{SSR}_{sref/r} \times \mathrm{SSRn}_{sx/sref} = \frac{\mathrm{FAS}_{sref}}{\mathrm{FAS}_{r}} \times \frac{\mathrm{FASn}_{sx}}{\mathrm{FASn}_{sref}}, \quad (2)$$

where $\mathrm{SSRh}_{sx/r}$ is the SSRh at a sensor sx located at the site under study. $sref$ and r are respectively the local intermediate reference station located within the basin and the rock-reference station outside the basin. FAS and FASn are the Fourier amplitude spectra computed, respectively, with earthquakes and noise recordings. Perron *et al.* (2018) have compared both SSRh and classical SSRn to SSR for two basins, in Provence (France) and in Argostoli (Greece). Their results showed that SSRh was able to reproduce SSR for a wide frequency range (up to 12 Hz in Provence and 20 Hz in Argostoli). They also observed unrealistic amplification factors with SSRn for frequencies higher than the minimal resonance frequency f_0 of the basins. However, as the rock-reference station was located on a noisier area than the stations within the basin in Argostoli, and as no clear relation with site-to-reference distance was found in the SSRn results for both basins, they did not assign this discrepancy to the influence of local anthropogenic noise sources. They rather proposed that the sedimentary basin acts as a natural barrier, so that part of the ambient noise wavefield is trapped within it at frequencies higher than f_0. As a result, only a limited part of the wave energy generated within (or crossing) the basin would reach the rock-reference station above f_0. Such behaviour is not observed in SSR because body waves generated by a distant earthquake reach the Earth's surface with a quasi-vertical incidence. This interpretation explains not only the discrepancy between the rock-referring

SSRn and the SSR above f_0, but also the agreement observed in Perron *et al.* (2018) between SSRh (that uses SSRn between two stations under the same geological conditions) and SSR over a wide frequency band. The authors concluded that SSRh should be tested to other sites in order to better characterize its conditions of applicability. More recently, Perron *et al.* (2022) proposed a site-specific hazard map derived from an empirical amplification model based on SSRh for the Sion area (Swiss Rhone Valley).

1.2 Applicability of noise-based spectral ratios

Perron *et al.* (2018) show that SSRh is a promising method to mitigate the impact of the difference in the geological conditions between sensors at the site and the reference location in noise-based spectral ratios. The authors also explained the good results of SSRh by showing that site-to-site SSRn is in good agreement with site-to-site SSR up to 12 Hz in Provence basin for both horizontal and vertical components, whatever the phase of the signal that is considered for SSR computation. This result tends to validate an important assumption for noise-based spectral ratios applicability to assess site response, which is that the seismic response of a site to an ambient noise wavefield, usually considered as dominated by surface waves (e.g. Bonnefoy-Claudet *et al.* 2006), must be comparable to that of the incident seismic (body and surface) waves generated by earthquakes (Lermo & Chávez-García 1994; Mori *et al.* 2016; Kawase *et al.* 2019). However, a second assumption is necessary, in which the only resulting factor when computing the spectral ratio corresponds to the amplitude of the site transfer function, meaning that this ratio makes it possible to isolate the site influence in the ambient noise wavefield. This assumption is challenged by the presence of strong local sources of seismic energy. Ignoring the influence of such strong local sources would mean relying on some kind of 'local ergodicity' assumption, that is, any point of the array experiences a fully equipartitioned noise field, even if not excited by the same sources. In other words, all oscillating modes are locally excited at any point of the array meaning that a spectral ratio between the wavefield amplitude at two locations would mainly capture the local geological structure imprint, whatever the sources. Although not straightforward to verify, it is most likely not a valid hypothesis for most highly anthropized sites, especially at high frequency where localized and dominant sources like a railway or a persistently vibrating dam could itself break such assumption.

1.3 Purpose of our study

In this study, we investigate the potential of SSRn and SSRh in a highly industrialized region in the area of the Tricastin Nuclear Site (TNS, south-east France). This will allow us to complement the work by Perron *et al.* (2018, 2022) by investigating SSRn and SSRh in a very anthropized area. Our objective is to give some further insights into the conditions of application of SSRn and SSRh for them to reproduce SSR, especially regarding the impact of local noise sources. In the following, we first briefly introduce the target area and the data acquisition before identifying the main anthropized features of our site of study. Such analysis is important to better understand amplification results that reflect here, the influence of strong sources in local ground motion. Hence, we propose a two-step workflow to attenuate source effects in local amplification assessment through noise-based approaches. We then show results of noise-based spectral ratios, considering our workflow or not. We discuss the impact of our workflow for the utilization of SSRn and SSRh in noisy industrialized environments. We here speculate that by selecting quiet periods within acquisition, and thus muting strong transient signals emanating from local sources, we are getting closer to the assumption we refer to as 'local ergodicity', at least in some parts of the array.

2 CONTEXT AND DATA ACQUISITION

2.1 Context of the study

The Tricastin region (French Rhone Valley, Fig. 1a) is located on top of an elongated valley that was dug during the Messinian Salinity Crisis in Cretaceous sandstones and limestones and that is now filled with Pliocene and Quaternary sediments (sands and clays, Gélis *et al.* 2022). The site is interesting to study for several reasons. First, in terms of seismic hazard, since (1) the expected strong lithological contrast between the sedimentary filling and the bedrock, as well as the valley's incised geometry, is prone to generate strong and complicated site effects; (2) the area hosts critical infrastructures including the large TNS and the hydroelectric dam of Donzère–Mondragon, as well as five towns with a total population over 45 000. Second, in terms of experimenting SSRh method in a different context than in Perron *et al.* (2018), as the Tricastin Valley is larger than the studied basins in Provence and Argostoli and as the targeted area is heavily industrialized. Indeed, in addition of hosting the TNS and the hydroelectric dam, the site under study is crossed from north to south by two freight lines and one high-speed railway line as well as the busy A7 highway and N7 National road. These main expected anthropogenic sources are displayed in Fig. 1(b). In the context of investigating noise-based spectral ratios computations, it is worth noting the unequal repartition between noisy environments, mostly located on top of the sedimentary filling (i.e. mostly the centre area of the valley), and quiet environments, rather located on top of the cretaceous outcrops (on both sides of the valley).

In 2019 November, an M_w 4.9 earthquake occurred on La Rouvière Fault (LRF), located about 20 km away from the targeted area (Fig. 1a). This event, which is the strongest in metropolitan France since 1967, took place near the Le Teil village. Strong ground motion and severe damage were observed in the villages located in the vicinity of the rupture zone (Ritz *et al.* 2020; Causse *et al.* 2021; Cornou *et al.* 2021). Schlupp *et al.* (2021) estimated, on the basis of field observations and communal surveys, a maximum intensity (EMS98) of about VII–VIII in Le Teil and at least VI in 30 nearby localities. This recent earthquake further highlights the importance of conducting seismic hazard studies in the area.

2.2 Data acquisition

For the purpose of conducting seismic hazard studies in the area of Tricastin, we continuously recorded ambient noise from 2020 February to March on a 400-sensor seismic array (all-in-one three-components Geospace GSX nodes) in the framework of the French-German DARE (Dense ARray for seismic site effect Estimation) project. The array was designed on the basis of five subarrays (Fig. 1a): a loose grid over the entire zone; a very dense grid located 2–3 km south of the TNS and covering the narrowest part of the valley with an inter-station distance of about 200—250 m; two dense east–west lines that follow two departmental roads in the north of the targeted area and six distant nodes covering different azimuths. One of these distant nodes was located on LRF. The total size of the combination of the first four subarrays, that is the core of our target

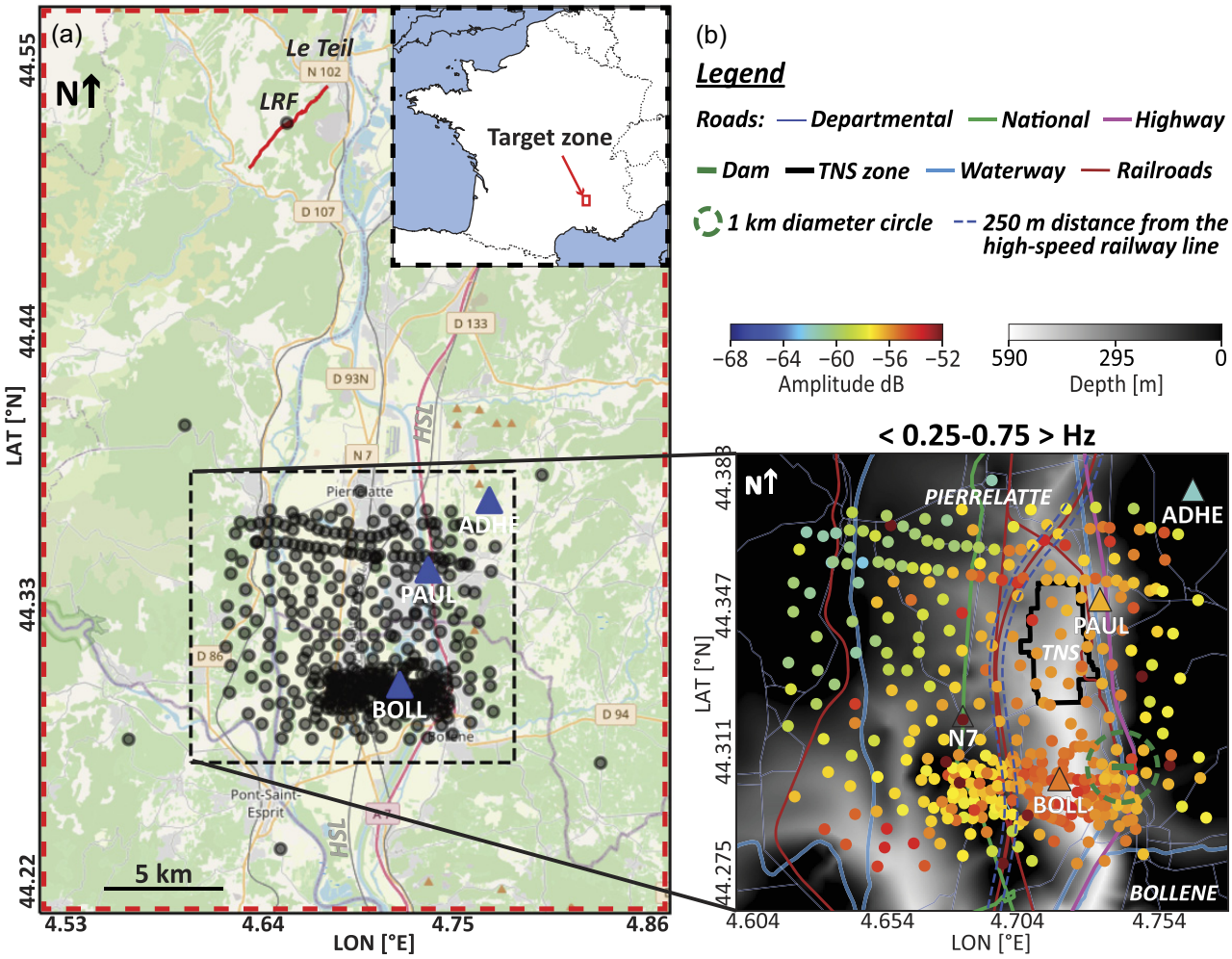

Figure 1. (a) Location of the study area. The thin red line represents the LRF. The nodes are represented in grey dots. BOLL, PAUL and ADHE stations are represented by blue triangles. The dashed black rectangle delimits the core of our target zone. (b) Median FAS at each sensor, averaged between 0.25 and 0.75 Hz (coloured dots), superimposed with the depth of the bedrock (grey scale; Bagayoko 2021, M.Sc. thesis). The FAS are computed from 10-min long signal windows over the whole recordings. The main anthropized features are displayed and referenced in the legend. The results in (b) are shown for the E-W component.

zone, is about 10 km by 10 km (Fig. 1). For more information on the experimental design and the global data set, we invite the reader to refer to Froment *et al.* (2022a). Three nodes were co-located with three seismic stations (called ADHE, BOLL and PAUL, Fig. 1) that have been deployed intermittently since 2016–2017 and have therefore recorded the seismicity for several months (Gélis *et al.* 2022). BOLL and PAUL sites are located on top of the sedimentary filling whereas ADHE is placed on top of the cretaceous outcrop. Gélis *et al.* (2022) showed almost flat H/V ratio with a value around 1 between 0.2 and 10 Hz at ADHE and therefore used this station as rock-reference for SSR computation.

In order not to blindly use continuous data in the spectral ratios, we evaluated at each of the 400 nodes the data quality for the east–west, north–south and vertical components. For this purpose, we constructed a catalogue in which the continuous time-series, the FAS based on a 10-min long segmentation as well as the spectrogram are computed for each sensor. Based on a qualitative examination of this catalogue, two sensors for the east–west and north–south components, as well as one sensor for the vertical component were considered as not suitable for the purpose of this study (and therefore not considered) as no time window and frequency range were considered exploitable for these nodes. For more information about this survey, see Froment *et al.* (2022a) since the complete catalogue is attached as electronic supplement.

With the aim of investigating the spatial variability of the low frequency noise amplitude, we computed the median over all the 10-min long windows FAS (used in the data quality survey) at each sensor and averaged within the [0.25–0.75 Hz] frequency band (Fig. 1b). Note that in our computations, the FAS correspond to the Fourier transform of non-overlapping sliding windows and that no smoothing filter was applied here, neither to the FAS nor to the median of the FAS (called median FAS through the paper). Fig. 1(b) reveals a good agreement between the noise amplitude and the geology revealed by the topography of the bottom of the canyon deduced from the work of Bagayoko (2021, M.Sc. thesis). Indeed, higher amplitudes are observed on top of the sedimentary filling where the valley is the deepest (lightest area on the background map). By contrast, lower noise amplitudes are observed for nodes located on top of outcropping bedrock (darkest areas). It is worth noting that, according to Froment *et al.* (2022b), the minimal

resonant frequency f_0 of the valley is about 0.5 Hz (hence situated in the middle of the frequency band considered). This suggests that the ambient noise captures, at least qualitatively, the effect of local amplification around the resonant frequency of the valley, and that we may relate to the effect of the overall geological structure. Note that the results are shown for the E-W component.

3 CHARACTERIZATION OF THE LOCAL SOURCES

In order to better understand amplification results obtained through noise-based methods at a specific site and to investigate the impact of local noise sources, we start our analysis by examining the different noise sources in the study zone. For this purpose, we averaged the median FAS within higher frequency bands than in Section 2.2 (Fig. 2a). We separated daytime (from 6 a.m. to 8 p.m.) and night time (8 p.m.–6 a.m.) recordings, to better characterize the impact of human activities on the seismic noise wavefield. Indeed, it is usually assumed that above 1 Hz, ambient noise is mostly dominated by anthropogenic sources near inhabited areas (Bonnefoy-Claudet *et al.* 2006). First, Fig. 2(a) shows that for all frequency bands, we tend to observe smaller noise amplitudes at night, especially in the central part of the area. This is particularly clear within 8.5–9 and 11.5–12 Hz, for which much higher amplitudes are observed along the high-speed railway line during the day. Between 2.5 and 3 Hz, for both daytime and night-time recordings, higher amplitudes are observed at some sensors located to the south of the TNS imprint (Fig. 1b). In the [5.5–6 Hz], [8.5–9 Hz] and [11.5-12 Hz] frequency bands, again for both daytime and night-time recordings, very high noise amplitude level is found for sensors placed all along the Western side of the A7 highway and around the TNS imprint. Finally, strong amplitudes are also seen nearby the hydroelectric dam in the [8.5–9 Hz] and [11.5–12 Hz] frequency bands for both daytime and night-time records. The striking spatial consistency between the distribution of the main identified noise sources (Fig. 1b) and high noise amplitude patterns illustrates the impact of local anthropogenic sources for frequencies of a few hertz (Fig. 2a).

Let us now focus on the impact at specific sensors close to dominant sources. We plotted in Fig. 2(b), the median FAS (thin solid lines) for sensors located in the vicinity of the high-speed railway line (left-hand panel) and of the hydroelectric dam (right-hand panel). We consider to be in the vicinity, sensors located up to 250 m away from the high-speed railway line (i.e. dashed blue lines in Fig. 1b) and within a 1 km radius from the dam (dashed green circle in Fig. 1b). We smoothed each median FAS with a Konno-Ohmachi smoothing (Konno & Ohmachi 1998), taking b parameter equal 40 (eq. A1 in Appendix A). First of all, the ambient noise amplitude in the vicinity of the high-speed railway line is significantly higher during the day than at night for frequencies higher than 1 Hz, which is consistent with the associated daily activity. By contrast, almost no difference is seen between daytime and night-time recordings for sensors close to the hydroelectric dam (right-hand panel), that is indeed not expected to stop during night-time. Looking more into details on the left panel, we note two peaks between 4 and 4.7 Hz and around 9 Hz for daytime. During the quality survey, these peaks have been found to be observed mainly for sensors located in the vicinity of the high-speed railway line at day. Hence, we consider these peaks as the spectral signature of high-speed trains traffic. Finally, regarding the right-hand panel, the spectral signature of the dam has been found to be characterized by a broad-band peak between 7 and 10 Hz. This analysis illustrates the variety of spectral signatures of the different anthropogenic sources we are dealing with in the area. Note that, as mentioned in Section 2.2, we simply consider the Fourier transform of non-overlapping sliding windows for the FAS computation. We can note that alternative processing, for instance the use of a multitaper approach (Prieto *et al.* 2009; Burjánek *et al.* 2018) to compute the FAS, could be explored to further discriminate sharp frequency peaks, as the ones discussed here and related to human activities.

Spectrograms of recordings at BOLL and PAUL stations are represented in Fig. 2(c) to have a closer look at the spectral content at these stations that will be used later as site reference in the SSRh. On both spectrograms, we can see that below 0.9 Hz, despite some slight fluctuations, the noise amplitude level remains quite stable over time which is consistent with the commonly assumed natural origins of the background noise at these frequencies (oceanic activity and large-scale meteorological conditions). At higher frequencies, we observe daily and weekly variations as expected for human activities. When paying more attention, we note two daily peaks around 1 and 3 Hz. As no clear spatial consistency was found to properly discriminate these peaks during the quality survey, we do not interpret their origin as being related to processes at the scale of the basin. These features are likely to reveal very local effects. For BOLL station, a clear energy peak is also present around 9 Hz. It is however probably not related to the railway activity (as the 9 Hz feature discussed above) given its stability through time. For PAUL station, the daily variations are not that pronounced, which is consistent with the quasi non-stop activity of the nuclear facilities in the vicinity.

Fig. 2 highlights the complexity of the noise signal above 1 Hz inside the valley. This complexity is related to the specific distribution of local sources. In such an anthropized context and for noise-based spectral ratios applications, it is crucial to be able to decouple the contribution of the geology from the influence of these local sources on the ground motion.

4 PROPOSED WORKFLOW TO MITIGATE LOCAL SOURCES IMPACT ON AMPLIFICATION ASSESSMENT

From the analysis of local sources realized in the previous section, comparing local ground motion without removing their effects might end up highlighting abnormally high and localized patterns of amplification which are not related to the underlying geological structure. For the computation of the so-called H/V ratios and noise correlation functions (NCFs), some strategies have been proposed to mitigate biases due to local sources in noise recordings by removing time-impulsive signals in the data (e.g. SESAME team 2004; Bensen *et al.* 2007; Denolle *et al.* 2013; Viens & Iwata 2020). For SSRn and SSRh computation, Perron *et al.* (2018) followed the processing proposed by Denolle *et al.* (2013). These transient signals are indeed likely to be generated by local transient disturbances like footsteps, cars, trains, etc. One commonly used strategy to mitigate transients, is the so-called STA/LTA anti-triggering filter (e.g. SESAME team 2004). However, some tests performed on our data set suggest the difficulty of choosing optimal time windows in the STA/LTA procedure to properly detect and discriminate the transients when their repeatability in time is high. In this study, we introduce an alternative procedure based on a two-step workflow. Note that in comparison to the STA/LTA procedure, the proposed workflow aims at isolating and removing transient signals based on

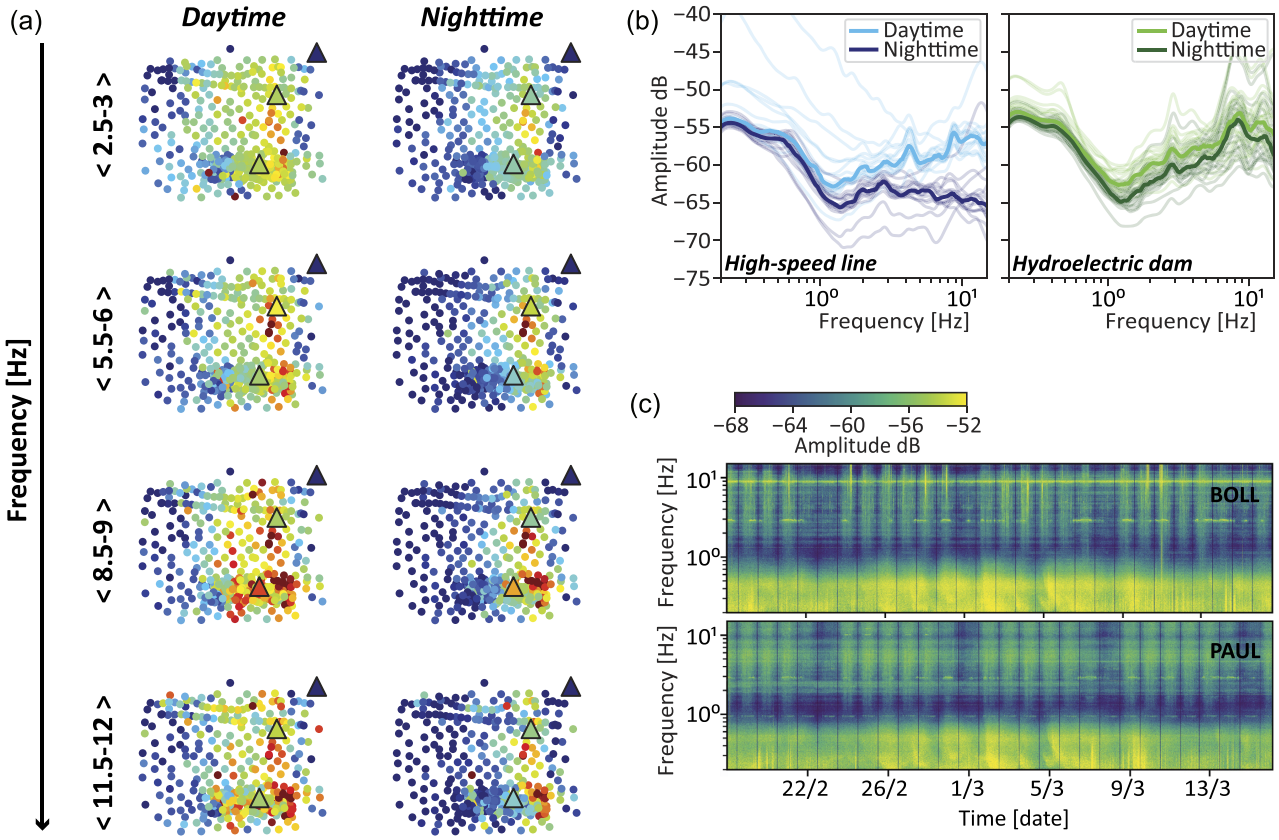

Figure 2. (a) Median FAS at each sensor for daytime and night-time recordings, averaged over different frequency bands. (b) Median FAS (thin solid lines) for sensors located in the vicinity of the high-speed railway line (blue lines in left-hand panel) and of the hydroelectric dam (green lines in right-hand panel) for daytime and night-time recordings. We consider to be in the vicinity, sensors located up to 250 m away from the high-speed railway line (i.e. dashed blue lines in Fig. 1b) and within a 1 km radius from the dam (dashed green circle in Fig. 1b). The blue and green thick solid lines show the median for daytime (light) and night-time (dark) curves. (c) Spectrogram at BOLL and PAUL stations for frequencies between 0.2 and 15 Hz. All results in this figure correspond to the E-W component.

a spectral representation of the wavefield. Our workflow is therefore consistent with the purpose of spectral ratios computation.

The first step of our workflow consists in defining an optimal duration of the time window before computing the FAS. The initial definition of 10-min long durations was ad hoc. Like classical trade-off in time–frequency analysis, the longer the window the better the frequency resolution and the representativity of the current FAS as compared with the continuous data, but the lower our ability to separate transient signals. Since our goal is to remove the contribution of significant transient signals, the optimal window would then be the largest window able to properly discriminate the different transient sources whatever their time and frequency characteristics. In order to better characterize the impact of the time window duration (Tw) on the data, we compared the FAS relative density for Tw = 10 min (Fig. 3a) and Tw = 1 min (Fig. 3b) at a specific sensor located nearby the N7 National road (the so-called N7 site in Fig. 1b). The FAS relative density simply gives the relative amount of spectra at a given amplitude and a specific frequency. When looking at the highest amplitudes over the whole frequency band, lower density is found for Tw = 1 min compared to Tw = 10 min. We also note two distinct behaviours whether looking below 4 Hz or above. Below 4 Hz, several spectra separate from the rest of the data reaching high amplitudes and causing a large dispersion around the median FAS for Tw = 10 min, while a very low amount of spectra reaches these amplitudes for Tw = 1 min. On the other hand, above 4 Hz, all the spectra are relatively concentrated around the median FAS for Tw = 10 min, while several separate from the rest of the data and reach very high amplitudes for Tw = 1 min. From these observations, we suggest that, depending on the frequency, a certain amount of 10-min long segments contain some strong time-impulsive signals of duration shorter than 10 min and that weigh the amplitude when computing the FAS. Below 4 Hz, some 10-min long segments are impacted. Above 4 Hz, the repetition in time of the characteristic transients may be high enough to impact the large majority of the segments. By doing very short time windows (i.e. Tw = 1 min, Fig. 3b), we concentrate these transients into some segments while the rest of the data is now free from these spurious arrivals and thus characterized by a lower amplitude.

To further understand how the transients are isolated according to the duration of the time window, we evaluated at each frequency, the median of the absolute deviations (MAD) from the median FAS (thin solid red line in Figs 3a and b) depending on Tw (Fig. 3c). The MAD is a measure of dispersion. The benefit of using the MAD instead of the standard deviation is that it might be more robust to outliers (e.g. Rousseeuw & Croux 1993; Leys *et al.* 2013). This analysis may be seen as a quantification of the distribution of transient signals over time. The two behaviours described above are clearly emphasized. Below 4 Hz (blueish curves), the MAD (i.e. the dispersion) is quite high for long segments between Tw = 8 and 60 min; this is because a significant portion of temporal

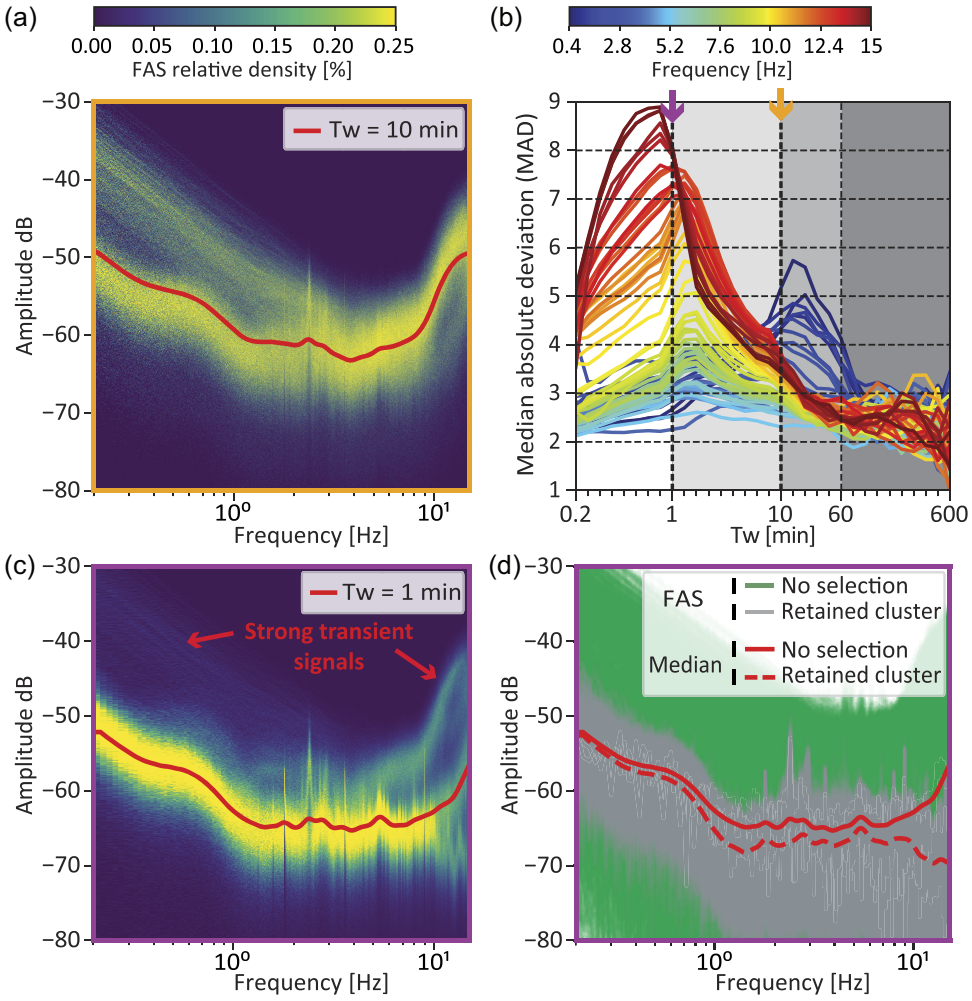

Figure 3. (a) and (b) FAS relative density for respectively 10 and 1 min long windows (Tw). The median FAS is in solid red line. (c) MAD as a function of Tw and frequency. Purple and orange arrows indicate, respectively, Tw = 1 and 10 min. Note that the horizontal axis is not linear and reflects different timescales highlighted by shaded areas. (d) Comparison between all the FAS (in green) and the FAS forming the cluster retained by DBSCAN (in grey) for Tw = 1 min. The median over all the FAS and over the retained FAS are, respectively, in solid and dashed red line. All results in Fig. 3 are derived for the sensor located at site N7 (Fig. 1b) for the E-W component. The colour of the frame in Figs 3(a), (b) and (d) stands for the Tw used in the respective panels (i.e. orange: Tw = 10 min and purple: Tw = 1 min).

segments contains strong transients, causing a large dispersion (like in Fig. 3a). Reducing Tw ends up densifying the data around the median FAS by decreasing the portion of segments impacted. Hence, the MAD decreases with Tw from Tw = 8 min. On the other hand, above 4 Hz (greenish to reddish curves), the MAD is low for long segments (i.e. Tw superior or equal to 10 min); this is because the large majority of the segments may contain strong transient signals. As a result, while isolating transients by decreasing Tw, it increases the dispersion produced by high amplitude transients as seen in Figs 3(a) and (b). This dispersion reaches its maximum when Tw is approximately equal to what could be the actual characteristic length of a transient signal (see the maximal deviation just around 1-min time window for high-frequency range, Fig. 3c). Below this peak, the probability of splitting a transient into two different segments (and therefore reducing its relative impact on each segment) increases, decreasing the MAD. The fact that this peak is shifted to lower Tw when the frequency is increasing is consistent with the assumed length of a transient depending on its frequency.

For Tw higher than 60 min, each segment becomes more and more representative of the average amplitude level of the whole recording, so that all the spectra will again converge towards the median FAS (i.e. the noise is stationary on a long-term basis), no matter the frequency we are looking at. This explains the very low MAD from Tw = 60 min.

Based on this analysis, we chose Tw = 1 min as the optimal time window in order to separate transients at high frequency (Fig. 3c) since we aim at mitigating source effects impacting the signal at high frequency. This time window also ensures a suitable frequency resolution of the FAS.

The second step of our proposed workflow is to further identify families of FAS reflecting different behaviours to isolate time segments cleaned from strong transients. We used the DBSCAN algorithm (Density-Based Spatial Clustering of Applications with Noise, e.g. Deng 2020), for this purpose. The underlying idea is that FAS associated with strong transients could form specific clusters of high amplitude that can then be easily removed. Based on this idea, we discarded the cluster tagged as −1 in the classification result of each sensor, which corresponds to non-consistent data according to DBSCAN, and retained the cluster for which the median FAS is the lowest. More information about DBSCAN algorithm can be found

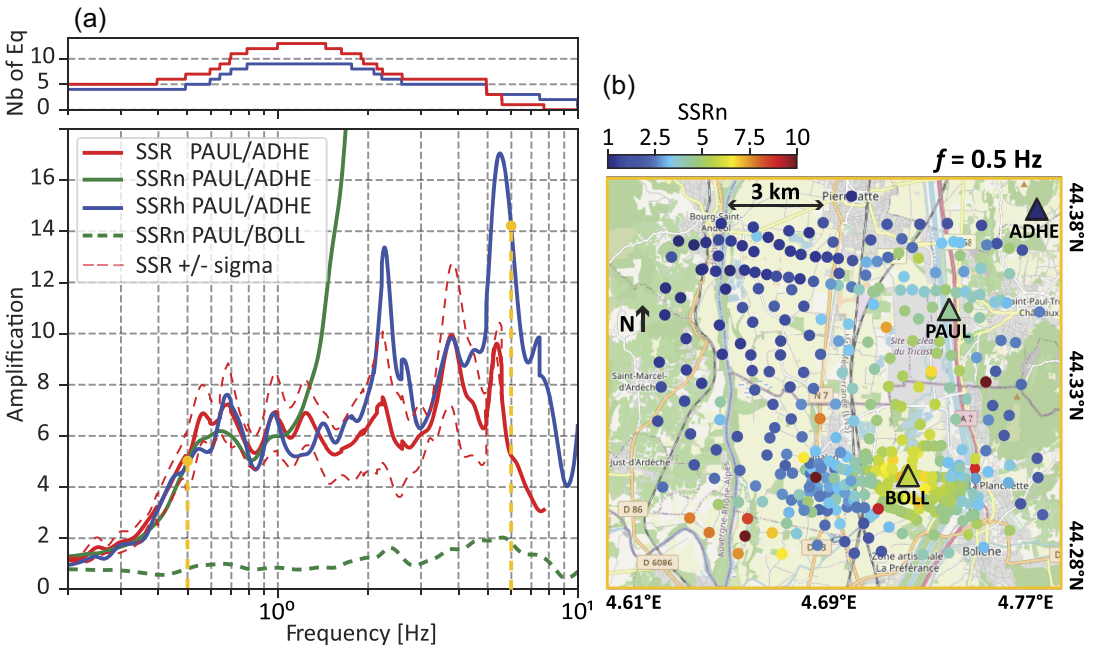

Figure 4. (a) Comparison between $SSR_{PAUL/ADHE}$, $SSRn_{PAUL/ADHE}$ and $SSRh_{PAUL/ADHE}$ with BOLL as local site-reference station, respectively, represented by the solid red, green and blue curves. The site-to-site $SSRn_{PAUL/BOLL}$ is shown in dashed green line. Plus and minus one standard deviation to SSR are plotted in dashed red lines. The number of earthquakes used to compute $SSR_{PAUL/ADHE}$ and $SSRh_{PAUL/ADHE}$ is represented on the top panel, respectively, in thin solid red and blue line. Orange dots indicate $SSRn_{PAUL/ADHE}$ value at 0.5 Hz and $SSRh_{PAUL/ADHE}$ value at 6 Hz. (b) Map of SSRn value at 0.5 Hz computed at each sensor of the array. All results in this figure correspond to the E-W component.

in Appendix B. Fig. 3(d) shows all the FAS computed using 1-min long segments (in green) and the FAS forming the retained cluster (in grey) for site N7. Retained FAS represent ∼9.6 per cent of all the FAS. By comparing to Fig. 3(b), we can assert that the resulting selection is able to put aside most of the transient signals of high amplitude, most probably due to traffic along the National road. This translates into a lower resulting median FAS (dashed red line). Computing SSRn and SSRh on the basis of this selection should then help reducing the impact of local sources, thereby reflecting better the actual amplification associated with the underlying geology.

5 SPECTRAL RATIOS COMPUTATION AND RESULTS

5.1 Spectral ratios computation

In the following sections, we present the results of SSRn and SSRh with or without considering our workflow in the computations.

First, Section 5.2 is dedicated to the results obtained without considering our workflow. We present in this section a comparison of SSRn and SSRh with classical earthquake-based SSR (Fig. 4), as well as an analysis of SSRh amplification factors across the entire array (Fig. 5). No pre-processing in the noise-based spectral ratios except a selection of night-time data segments to remove major, and obvious anthropogenic sources (like trains passing, Figs 2a and b) was applied. We recall that 10-min long segments are chosen to compute initial noise-based ratios. We thus computed the ratios between 10-min long windows FAS obtained at each sensor and the synchronous time windows FAS at rock-reference site ADHE between 0.2 and 15 Hz for the three components of motion. To attenuate the impact of outliers, we defined our SSRn at a sensor sx as the median over these ratios and smoothed the median using Konno–Ohmachi smoothing (Konno & Ohmachi 1998), taking b parameter equal 40 (eq. 3). Note that we do not smooth the FAS, as it may be commonly done, given the amount of data and the related difficulty due to the duration of such processing. However, some tests have been performed on a limited set of data revealing comparable results for the two processing. We used the Konno–Ohmachi smoothing because of its ability of achieving a uniform smoothing over the whole frequency band through an adaptative window size (eq. A1 in Appendix A). Hence, one can avoid oversmoothing low frequencies. 40 is commonly chosen as the value of parameter b: this has been found to be a good compromise to efficiently smooth the spectra without oversmoothing. We benefit from SSR computed by Gélis *et al.* (2022) at sites BOLL and PAUL relative to ADHE (following eq. 4) for comparison. We invite the reader to refer to Gélis *et al.* (2022) for details regarding the SSR computation. Using co-located nodes at BOLL and PAUL (see Section 2.2), we computed the SSRh at each sensor sx, either considering BOLL or PAUL as a local site reference (eq. 5). As the node sensor at site PAUL is considered unusable for the N-S component (see quality control in Froment *et al.* 2022a), we did not estimate the SSRn and the SSRh using this station for this component. The following are the mathematical expressions of our SSRn, SSR and SSRh computation:

$$SSRn_{sx/ADHE} = \left\{ \left\langle \frac{FASn_{sx,\,Tw(i)}}{FASn_{ADHE,\,Tw(i)}} \right\rangle \right\}, \tag{3}$$

$$SSR_{s/ADHE} = \left\{ \left| \frac{FAS_{s,\,k}}{FAS_{ADHE,\,k}} \right| \right\}, \tag{4}$$

$$SSRh_{sx/ADHE} = \left\{ \left| \frac{FAS_{s,\,k}}{FAS_{ADHE,\,k}} \right| \right\} \times \left\{ \left\langle \frac{FASn_{sx,\,Tw(i)}}{FASn_{s,\,Tw(i)}} \right\rangle \right\}, \tag{5}$$

where sx is a sensor of the array, s is either BOLL or PAUL station, $Tw(i)$ is the time window at time i, k is the considered earthquake, $\langle . \rangle$, $|.|$ and $\{.\}$ are, respectively, the median over the noise-based FAS

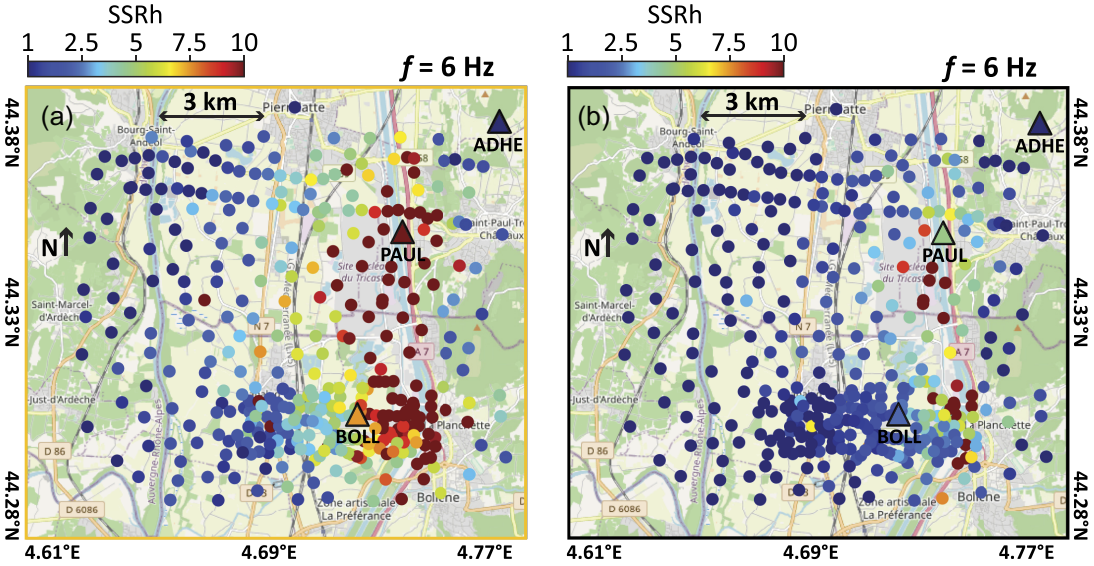

Figure 5. Maps of SSRh value at 6 Hz computed at each sensor, respectively, using BOLL (a) and PAUL (b) as local site-reference station. Given eq. (2), note that the SSRh value at BOLL in (a) and at PAUL in (b) corresponds to the SSR value. All results in this figure correspond to the E-W component.

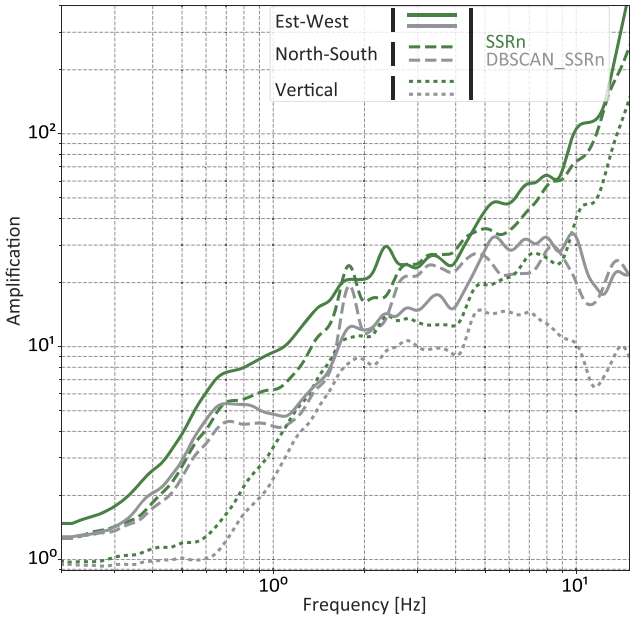

Figure 6. Comparison between SSRn (in green) and DBSCAN_SSRn (in grey) at site N7 for the E-W (solid lines), N-S (dashed lines) and vertical (dotted lines) components. The logarithmic scale has been used here to visualize properly the amplification that covers several orders of magnitude.

ratios, the geometric mean over the SSR and the Konno–Ohmachi smoothing.

Second, Sections 5.3 and 5.4 are dedicated to the results obtained following our workflow. In these sections, we evaluate its impact by comparing SSRn to our so-called DBSCAN_SSRn (Fig. 6) and SSRh to our so-called DBSCAN_SSRh (Fig. 7). In these figures, SSRn and SSRh are estimated considering all synchronous 1-min long windows FAS between each sensor and the reference station. DBSCAN_SSRn and DBSCAN_SSRh are estimated considering only the synchronous FAS between the retained cluster at each sensor and the reference. Still only night-time segments are used in the computations.

5.2 Results of SSRn and SSRh without considering our workflow

In Fig. 4(a), we compare $SSR_{PAUL/ADHE}$, $SSRn_{PAUL/ADHE}$ and $SSRh_{PAUL/ADHE}$ using BOLL as local site-reference station for the E-W component. Below 1 Hz, a good agreement is found between $SSRn_{PAUL/ADHE}$, $SSRh_{PAUL/ADHE}$ and $SSR_{PAUL/ADHE}$. Moreover, as expected from the spatial variability of low frequency noise amplitude discussed in Section 2.2 (Fig. 1b), SSRn value at 0.5 Hz across the area (Fig. 4b) is well correlated with the underlying geology, showing higher amplification on top of the sedimentary filling where the valley is expected the deepest (Fig. 1b). This confirms that, below 1 Hz, noise-based spectral ratios seem to be able to capture site effects. Above 1 Hz, $SSRn_{PAUL/ADHE}$ strongly deviates from $SSR_{PAUL/ADHE}$ (Fig. 4a), reaching values that are unlikely to be correct. As 1 Hz is usually described as a threshold in the origin of noise (i.e. between ambient noise generated by natural sources at low frequency and noise coming from human activities at high frequency), the observations above 1 Hz suggest the impact of local anthropogenic sources close to PAUL site on $SSRn_{PAUL/ADHE}$ (e.g. A7 highway and the TNS). $SSRh_{PAUL/ADHE}$ presents a better agreement to $SSR_{PAUL/ADHE}$, by giving close estimate of amplification until around 2 Hz and in between 3 and 4 Hz. For the rest of the frequency band, $SSRh_{PAUL/ADHE}$ still tends to give higher estimates than $SSR_{PAUL/ADHE}$, but it is able to retrieve its global shape with some peaks at specific frequencies. Note that we represent the amplification using a linear scale and not a logarithmic scale (that may be more commonly used for SSR representation) in order to unambiguously compare the different curves that cover a limited range of amplification. The difference between $SSRh_{PAUL/ADHE}$ and $SSR_{PAUL/ADHE}$ seems to be related to the intermediate $SSRn_{PAUL/BOLL}$ used in the SSRh computation (dashed green line). This site-to-site ratio is indeed significantly above 1 for the frequencies where $SSRh_{PAUL/ADHE}$ is overestimating $SSR_{PAUL/ADHE}$, reaching a factor of 2 at 6 Hz. Hence, we reported SSRh value at this frequency across the array, either considering BOLL (Fig. 5a) or PAUL (Fig. 5b) as local site-reference station. SSRh estimation is supposed to give similar results whatever the intermediate site-reference used. Yet, results show lower SSRh values across the array when PAUL is

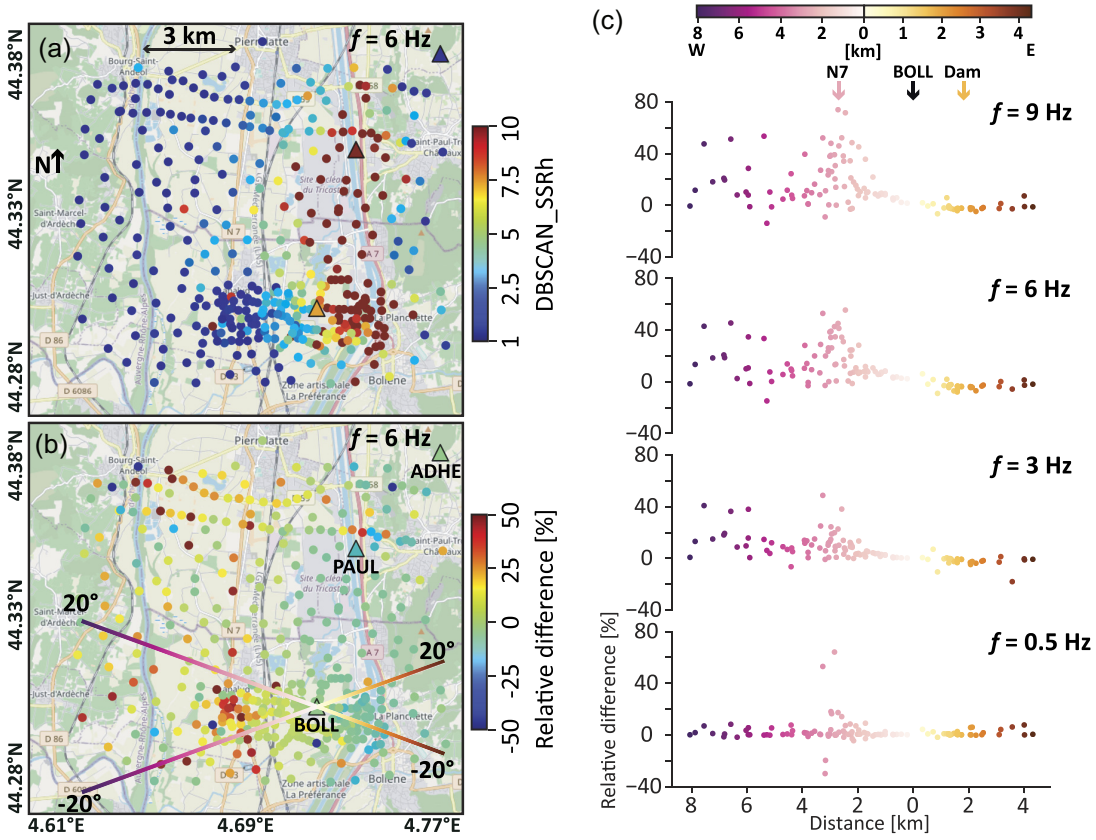

Figure 7. DBSCAN_SSRh results (a) and the relative difference (b) between DBSCAN_SSRh and SSRh at 6 Hz at each sensor. (c) Relative difference value as a function of distance from BOLL site, for sensors located inside an azimuth of ±20° east and west of BOLL, for $f = 0.5$, 3, 6 and 9 Hz. BOLL and ADHE (Fig. 1) are, respectively, the site- and rock-reference stations considered for the SSRh computation. All results in Fig. 7 are derived for the E-W component.

used as site reference. This is likely to be due to the difference in the source signature experienced at 6 Hz between PAUL, that is close to the A7 highway and the TNS, and BOLL that is further away from these strong anthropogenic sources. This result reveals that considering a local intermediate site reference is not sufficient to fully cancel source effects at high frequency. This illustrates the limitations of the application of SSRh in our highly industrialized area using simple criteria to select noise windows (10-min long windows, night-time recordings).

5.3 Impact of our workflow on the resulting SSRn at N7

A comparison between DBSCAN_SSRn and SSRn at site N7 is shown in Fig. 6 as grey and green lines, respectively. As expected, DBSCAN_SSRn exhibits lower amplification factors than SSRn over the whole frequency range for all components. The discrepancy between DBSCAN_SSRn and SSRn is generally increasing with the frequency, which highlights the importance of a careful data selection for high frequencies. It is worth noting that this observation holds for all components.

Fig. 6 shows that computing SSRn following our workflow ends up attenuating unrealistically high values. However, the resulting amplification factors remain very high, especially for the horizontal components. Amplification reaching values of 20–30 at frequencies of a few hertz are much higher than values observed at PAUL using earthquake-based SSR (see red curve in Fig. 4a). Even if the very local geology may influence the amplification within the valley between PAUL and N7 sites, N7 is not located on top of the axis of the canyon (by contrast to PAUL) and is not expected to be experiencing much larger amplification (i.e. as the ones estimated through DBSCAN_SSRn) at these frequencies. This suggests that DBSCAN_SSRn still presents some limitations to provide a reasonable estimate of the local amplification, at least at this site.

5.4 Impact of our workflow on the resulting SSRh over the entire array

Figs 7(a) and (b) show, respectively, DBSCAN_SSRh results and the relative difference $D_{f,\,sx}$ between DBSCAN_SSRh and SSRh at 6 Hz across the array for the E-W component. In addition to the broad-band vision at a specific sensor given in Fig. 6, Figs 7(a) and (b) allow us to quantify the impact of our workflow on SSRh amplification factors across the array at a given frequency. BOLL and ADHE (Fig. 1) are the site- and rock-reference stations, respectively. $D_{f,\,sx}$ at a sensor sx and frequency f is computed as follows:

$$D_{f,\,sx} = \frac{\mathrm{SSRh}_{f,\,sx} - \mathrm{DBSCAN\_SSRh}_{f,\,sx}}{\mathrm{SSRh}_{f,\,sx}} \times 100. \qquad (6)$$

Fig. 7(b) shows that a positive difference, which means that DBSCAN_SSRh is lower than SSRh, is seen nearby significant roads on the Western part of the array. It is particularly visible along the N7 National road (orange line crossing the centre of the array from north to south) where most of the sensors exhibit a positive difference of more than 25 per cent. These observations are consistent with the capacity of our workflow to attenuate unrealistically high

values in the spectral ratio by filtering the data from strong transient noise produced by the road traffic on this busy local road. Nevertheless, on the Eastern part of the array, almost no difference is found between DBSCAN_SSRh and SSRh. This means that the first step of our workflow fails to isolate transient signals among the collection of 1-min segments, suggesting that the main sources in this area, like the TNS, the hydroelectric dam and the A7 highway rather act as stationary sources where no transient signals can be isolated on a 1-min basis.

To extend this analysis on the proposed workflow impact to other frequencies, we represent $D_{f,\,sx}$ in Fig. 7(c) as a function of distance from BOLL site for sensors located inside an azimuth of $\pm 20^{\circ}$ east and west of BOLL, for $f = 0.5$, 3, 6 and 9 Hz. Almost no relative difference is seen for the sensors located east of BOLL, regardless of the distance and the frequency considered. On the other hand, as the frequency is increasing, more sensors exhibit high positive $D_{f,\,sx}$ when located west of BOLL, especially around the N7 National road location. This dependence to higher frequencies simply reflects the spectral characteristics of the signal produced by cars and trucks traffic (typically a few tens of hertz). These series of plots confirm that the proposed workflow performs well at all considered frequencies for sensors nearby the N7 National road, but does not help reducing the influence of the hydroelectric dam whatever the frequency. The contrast in the processing response between east and west of BOLL station emphasized in Fig. 7 reveals the difficulty of having a procedure that is efficient everywhere inside such a complex anthropized environment, where different types of sources cohabit.

6 DISCUSSION AND CONCLUSION

In this paper, we investigated the applicability of SSRn and SSRh in a heavily industrialized environment in the Tricastin area (French Rhône Valley, south-east France). In such anthropized conditions, comparing local ground motion without removing source effects might end up highlighting abnormally high and localized patterns of amplification which are not related to the underlying geological structure (Fig. 2). In order to tackle this problem and mitigate these source effects, we introduced a two-step workflow to remove strong impulsive signals (Fig. 3). These so-called transient signals are indeed likely to be generated by local disturbances like traffic (cars, trucks, trains, etc.), or any natural or human related sources. After selecting only night-time data segments, a first step is to define a characteristic time window to optimally isolate significant transients in the continuous time-series. This time window must ensure a suitable frequency resolution of the FAS. We here chose a time window of 1 min. The second step consists in selecting the time segments that do not contain these significant transients based on a clustering of the FAS according to their spectral characteristics.

For the purpose of testing SSRn and SSRh applicability in our case, we computed a first estimation across the array without applying our workflow or any other pre-processing than a selection of night-time data segments (Figs 4 and 5). Below 1 Hz, SSRn amplification factor seems to be consistent with the expected geology of the site and able to capture the site effects. This is quantitatively confirmed by a comparison with earthquake-based SSR amplification factor. However, as expected, SSRn strongly deviates from SSR above 1 Hz. This is likely to be due to anthropogenic sources of noise in the vicinity of some sensors, mainly within the valley. SSRh, which involves a local intermediate reference, gives closer estimate of amplification to SSR above 1 Hz. However, its amplification factor still shows some differences with SSR curve that we attribute to the impact of local sources. SSRh limitations at high frequency are emphasized by the fact that taking two different site-reference stations leads to differences in amplification estimates. It is worth noting that the differences revealed must be discussed depending on the applications of these estimates and the targeted degree of precision. However, these features need to be examined since they contribute to understand what controls noise-based spectral ratios, thereby gaining insight into their applicability as tools for estimating local site amplification. The results show that the amplification provided by the SSRh approach may be affected by local noise sources, especially at high frequencies and in highly industrialized areas. This can be interpreted as a need of having a fairly well equipartitioned wavefield within the site under consideration; this obviously does not hold in our case.

Figs 4 and 5 illustrate the motivation for applying careful pre-processing steps and data selection before computing SSRn and SSRh to address the problem of local source effects in the resulting amplification factors. Hence, we estimated SSRn and SSRh following our two-step procedure and evaluated the impact of our proposed workflow on these source effects. Comparison between SSRn computed with or without our data selection at site N7 (Fig. 6) shows that computing SSRn following our workflow ends up attenuating unrealistically high values. However, the resulting amplification factors remain abnormally high as the frequency is increasing, especially for the horizontal components. We quantitatively evaluated the impact of our approach on SSRh across the array for different frequencies (Fig. 7). We concluded that the proposed workflow is indeed effective to attenuate the influence of some transient sources (like some busy roads), rather located on the Western part of the array, but has almost no effect on stationary sources (i.e. the nuclear site, hydroelectric dam, A7 highway), mainly located on the Eastern part. It is worth noting that the traffic may be almost continuous on the highway, by contrast to the N7 National road for which we can find some quiet time windows. Our results reveal the difficulty of having a simple and automatic procedure that is efficient everywhere inside such a complex anthropized environment, where different types of sources cohabit, and in particular stationary sources.

Our procedure focused on mitigating source effects. It is directly related to the spectral content of the noise wavefield and therefore consistent with the purpose of spectral ratios computation. The time window is chosen by analysing the data in order to discard transient signals. Another key point is that it is a fully automatic procedure, which is decisive for the processing of dense array measurements. As an area for improvement though, we think that this approach should be tested to other sites with an adaptation of the characteristic time window and clustering parameters depending on the sensor location and a careful analysis of the main local sources, for the method to be fully data adaptative.

Another idea would be to model the remaining stationary sources to estimate and correct their impact. Based on these models, one could then be able to mitigate the remaining differences between noise-based spectral ratios and SSR in order to bring to the forefront the amplification associated with the site's geology. However, this would rely on strong assumptions (e.g. on the source radiation, attenuation within the propagation medium) whose related uncertainties could not be easily quantified in the resulting estimated amplification.

Finally, one could also include the phase information when computing the spectral ratio, which corresponds to a full noise-based

deconvolution (e.g. Prieto *et al.* 2011). Similar to NCFs, deconvolutions computed between different station pairs allow to spatially filter the ambient wavefield and mitigate remaining local sources to retrieve the relative amplitude through the array. Although not straightforward as it is strongly dependent on the spatial distribution of noise sources (e.g. Tsai 2011; Stehly & Boué 2017), some studies demonstrated that amplitude of NCFs can be exploited, with specific processing, to map local amplification and 3-D propagation effects associated with sedimentary basins (e.g. Prieto *et al.* 2011; Denolle *et al.* 2014; Boué *et al.* 2016; Viens *et al.* 2016).

ACKNOWLEDGMENTS

We gratefully acknowledge the editor and assistant editor, as well as Matsushima Shinichi and an anonymous reviewer for their useful comments and critics. This work was funded by a public grant overseen by the French National Research Agency (grant no. ANR19-CE31-0029). We would like to thank the French and German colleagues of ISTerre and GFZ Potsdam as well as the University of Potsdam for their collaboration. Special thanks must go to François Lavoué and Vincent Perron for interesting discussions on this work as well as Nanaba Bagayoko, Edward Marc Cushing and Damien Do Couto for their important work on the geology of the Tricastin Valley. We also would like to cite colleagues who helped installing the nodes array (Flomin Tchawe Nziaha, Christophe Clément, Charles Beard, Sophie Beauprêtre, Gilbert Bouzat, Florent Brenguier, Edward Marc Cushing, Aurore Laurendeau, Olivier Magnin, Laurent Stehly, Alexandre Tourette) and thank the private land owners and communities that have hosted this array.

DATA AVAILABILITY

Data of the nodes array (Froment *et al.* 2023, doi:10.15778/RESIF.XG2020) will be available on the French RESIF datacenter (https://www.resif.fr/en/) at the end of the DARE project (end of 2023).

REFERENCES

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

Bard, P.-Y., 1999. Microtremor measurements: a tool for site effect estimation?in *Proceedings of the 2th IASPEI/IAEE (ESG2),* Yokohama, Japan, 1998 December 1-3, pp. 1251–1279.

Bard, P.-Y. & Bouchon, M., 1985. The two-dimensional resonance of sediment-filled valleys, *Bull. seism. Soc. Am.,* **75**(2), 519–541.

Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.,* **169,** 1239–1260.

Bonnefoy-Claudet, S., Cotton, F. & Bard, P.-Y., 2006. The nature of noise wavefield and its applications for site effects studies, *Earth Sci. Rev.,* **79,** 205–227.

Borcherdt, R.D., 1970. Effects of local geology on ground motion near San Francisco Bay, *Bull. seism. Soc. Am.,* **60,** 29–61.

Boué, P., Denolle, M.A., Hirata, N., Nakagawa, S. & Beroza, G.C., 2016. Beyond basin resonance: characterizing wave propagation using a dense array and the ambient seismic field, *Geophys. J. Int.,* **206,** 1261–1272.

Burjánek, J., Gischig, V., Moore, J. R. & Fäh, D., 2018. Ambient vibration characterization and monitoring of a rock slope close to collapse, *Geophys. J. Int.,* **212**(1), 297–310.

Causse, M., Cornou, C., Maufroy, E., Grasso, J.R., Baillet, L. & Haber, El, E., 2021. Exceptional ground motion during the shallow M_w 4.9 2019 Le Teil earthquake, France, *Commun. Earth Environ*., **2,** 14, https://doi.org/10.1038/s43247-020-00089-0.

Celebi, M., Dietel, C., Prince, J., Onate, M. & Chavez, G., 1987. Site amplification in Mexico City (determined from 19 September 1985 strong motion records and from recordings of weak motions), *Dev. Geotech. Eng.,* **44,** 141–151.

Cornou, C., *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.

Deng, D., 2020. DBSCAN clustering algorithm based on density, in *7th International Forum on Electrical Engineering and Automation (IFEEA),* 2020, pp. 949–953, doi:10.1109/IFEEA51475.2020.00199.

Denolle, M.A., Dunham, E.M., Prieto, G.A. & Beroza, G.C., 2013. Ground motion prediction of realistic earthquake sources using the ambient seismic field, *J. geophys. Res.—Solid Earth,* **118,** 2102–2118.

Denolle, M.A., Dunham, E.M., Prieto, G.A. & Beroza, G.C., 2014. Strong ground motion prediction using virtual earthquakes, *Science,* **343**(6169), 399–403.

Ferritto, J., 1995. Ground motion amplification using microseisms, in *Int. Conf. Recent. Adv. Geotech. Earthq. Eng. Soil. Dyn., Paper 4,* pp. 561–566, St. Louis, Missouri, 1995 April 2–7.

Field, E.H. & Jacob, K.H., 1995. A comparison and test of various site-response estimation techniques, including three that are not reference-site dependent, *Bull. seism. Soc. Am.,* **85**(4), 1127–1143.

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

Froment, B., *et al.* 2022b. First 3D characterization of the Rhône Messinian Canyon in the Tricastin area from complementary geophysical approaches, in *Proceedings of the 17th European Conference on Earthquake Engineering (17ECEE),* Bucharest, Romania, 4–9 September 2022.

Froment, B., *et al.*, 2022a. Complementary dense seismic campaigns in a low-to-moderate seismicity area for characterizing site effects: application to the French Rhône Messinian canyon, *Seismol. Res. Lett.,* **94**(1), 531–547.

Gélis, C., *et al.*, 2022. Estimation of the local seismic amplification on an industrialized site in the French Rhône Valley, *Pure appl. Geophys.,* **179,** 2119–2145.

Horike, M., Zhao, B. & Kawase, H., 2001. Comparison of site response characteristics inferred from microtremors and earthquake shear waves, *Bull. seism. Soc. Am.,* **91**(6), 1526–1536.

Huang, H.C. & Teng, T.L., 1999. An evaluation on H/V ratio vs. Spectral ratio for site-response estimation using the 1994 Northridge Earthquake sequences, *Pure appl. Geophys.* **156,** 631–649.

Irikura, K. & Kawanaka, T., 1980. Characteristics of microtremors on ground with discontinuous underground structure, *Bull. Disaster Prev. Res. Inst.,* **30,** 81–96.

Kagami, H., Duke, C.M., Liang, G.C. & Ohta, Y., 1982. Observation of 1-to 5-second microtremors and their application to earthquake engineering. Part II. Evaluation of site effect upon seismic wave amplification due to extremely deep soil deposits, *Bull. seism. Soc. Am.,* **72**(3), 987–998.

Kagami, H., Okada, S., Shiono, K., Oner, M., Dravinski, M. & Mal, A.K., 1986. Observation of 1-to 5-second microtremors and their application to earthquake engineering. Part III. A two-dimensional study of site effects in the San Fernando Valley, *Bull. seism. Soc. Am.,* **76**(6), 1801–1812.

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, *Seismol. Res. Lett.,* **67**(5), 25–34.

Kawase, H., Nagashima, F., Nakano, K. & Mori, Y., 2019. Direct evaluation of S-wave amplification factors from microtremor H/V ratios: double empirical corrections to "Nakamura" method, *Soil Dyn. Earthqu. Eng.,* **126,** 105067, https://doi.org/10.1016/j.soildyn.2018.01.049.

Konno, K. & Ohmachi, T., 1998. Ground-motion characteristics estimated from spectral ratio between horizontal and vertical components of microtremor, *Bull. seism. Soc. Am.,* **88,** 228–241.

Lermo, J. & Chávez-García, F.J., 1994. Are microtremors useful in site response evaluation?, *Bull. seism. Soc. Am.,* **84**(5), 1350–1364.
Leys, C., Ley, C., Klein, O., Bernard, P. & Licata, L., 2013. Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median, *J. Exp. Soc. Psychol.,* **49**(4), 764–766.
Mori, Y., Matsushima, S., Kawase, H. & Nagashima, F., 2016. Comparison of observed earthquake and microtremor horizontal-to-vertical spectral ratios and inversion of velocity structures based on their empirical ratios, *J. Jpn. Assoc. Earthq. Eng.,* **16**(9), 13–32.
Perron, V., 2017. *Apport des enregistrements de séismes et de bruit de fond pour l'évaluation site-spécifique de l'aléa sismique en zone de sismicité faible a modérée,* PhD thesis, Université Grenoble Alpes, Grenoble.
Perron, V., Bergamo, P., Danciu, L. & Fäh, D., 2020. *Statistic of the earthquake's site response: Determination of the minimum number of earthquakes, Final Rept*., ETH Zurich, doi:10.1785/0220210289.
Perron, V., Bergamo, P. & Fäh, D., 2022. Site amplification at high spatial resolution from combined ambient noise and earthquake recordings in Sion, Switzerland, *Seismol. Res. Lett.,* **93**(4), 2281–2298.
Perron, V., Gélis, C., Froment, B., Hollender, F., Bard, P.-Y., Cultrera, G. & Cushing, E.M., 2018. Can broad-band earthquake site responses be predicted by the ambient noise spectral ratio? Insight from observations at two sedimentary basins, *Geophys. J. Int.,* **215,** 1442–1454.
Prieto, G. A., Parker, R. L. & Vernon, F. L., 2009. A Fortran 90 library for multitaper spectrum analysis, *Comput. Geosci.,* **35,** 1701–1710.
Prieto, G.A., Denolle, M.A., Lawrence, J.F. & Beroza, G.C., 2011. On amplitude information carried by the ambient seismic field, *Compt. Rendus. Geosci.,* **343,** 600–614.
Raptakis, D., Theodulidis, N. & Pitilakis, K., 1998. Data analysis of the euroseistest strong motion array in Volvi (Greece): standard and horizontal-to-vertical spectral ratio techniques, *Earthquake Spectra,* **14**(1), 203–224.
Ritz, JF., Baize, S., Ferry, M., Larroque, C., Audin, L., Delouis, B. & Mathot, E., 2020. Surface rupture and shallow fault reactivation during the 2019 Mw 4.9 Le Teil earthquake, France, *Commun. Earth Environ.,* **1,** 10, https://doi.org/10.1038/s43247-020-0012-z.
Rousseeuw, P. J. & Croux, C., 1993. Alternatives to the median absolute deviation, *J. Am. Stat. Assoc.,* **88**(424), 1273–1283.
Schlupp, A., Sira, C., Maufroy, E., Provost, L., Dretzen, R., Bertrand, E., Beck, E. & Schaming, M., 2021. EMS98 intensities distribution of the "Le Teil" earthquake, France, 11 November 2019 (Mw 4.9) based on macroseismic surveys and field investigations, *Compt. Rendus. Geosci.,* **353**(S1), 465–492.
Semblat, J.F., Kham, M., Parara, E., Bard, P.-Y., Pitilakis, K., Makra, K. & Raptakis, D., 2005. Seismic wave amplification: basin geometry vs soil layering, *Soil Dyn. Earthqu. Eng.,* **25**(7-10), 529–538.
SESAME team, 2004. Guidelines for the implementation of the H/V spectral ratio technique on ambient vibrations: measurements, processing and interpretation (Deliverable), 1–62, SESAME European research project.
Stehly, L. & Boué, P., 2017. On the interpretation of the amplitude decay of noise correlations computed along a line of receivers, *Geophys. J. Int.,* **209,** 358–372.
Steidl, J.H., Tumarkin, A.G. & Archuleta, R.J., 1996. What is a reference site?, *Bull. seism. Soc. Am.,* **86**(6), 1733–1748.
Tsai, V.C., 2011. Understanding the amplitudes of noise correlation measurements, *J. geophys. Res.—Solid Earth,* **116,** B09311, doi:10.1029/2011JB008483.
Viens, L. & Iwata, T., 2020. Improving the retrieval of offshore-onshore correlation functions with machine learning, *J. geophys. Res.—Solid Earth,* **125**(8), doi:10.1029/2020JB019730.
Viens, L., Koketsu, K., Miyake, H., Sakai, S. & Nakagawa, S., 2016. Basin-scale Green's functions from the ambient seismic field recorded by MeSO-net stations, *J. geophys. Res.—Solid Earth,* **121,** 2507–2520.
Zhao, B.M., Horike, M. & Takeuchi, Y., 1998. Reliability of estimation of seismic ground characteristic by microtremor observation, in *Proceedings of the 11th European Conference on Earthquake Engineering (11ECEE),* Paris, France, 1998 September 6–11, pp. 6–11.

APPENDIX A: KONNO-OHMACHI SMOOTHING

The following equation is the Konno–Ohmachi formula as introduced in Konno & Ohmachi (1998).

$$W\ (f,\ f_c) = \left[\frac{\sin\left(\log_{10}\left(\frac{f}{f_c}\right)^b\right)}{\log_{10}\left(\frac{f}{f_c}\right)^b} \right]^4, \qquad \text{(A1)}$$

where f and f_c are the frequency and a center frequency respectively and b a bandwidth coefficient.

APPENDIX B: DBSCAN CLUSTERING ALGORITHM

DBSCAN is a density-based data clustering algorithm. A key point of DBSCAN for big data applications is that it is an unsupervised clustering method. Therefore, we do not need a prior information to set an initial number of clusters. Moreover, DBSCAN is able to classify data depending on their density, which suits perfectly our need since FAS relative density is changing when isolating transient signals. The only two parameters for this clustering procedure are chosen empirically after manual inspection for some examples. One parameter is the maximum Euclidian distance between the sample point and the neighbourhood boundary. We set this parameter to 150. The other is the density threshold, which corresponds to the minimal number of sample points in the neighbourhood to form a cluster. We set this parameter equal to the total number of spectra divided by 60.