The influence of an anticline structure on ambient noise spectral anomalies at an underground gas storage

Christine El Khoury, Alexandre Kazantsev, Damian Kula, Arthur Dartois, Hervé Chauris

Geophysical Journal International

Geophys. J. Int. (2024) **237,** 1061–1078
Advance Access publication 2024 March 6
GJI Seismology

https://doi.org/10.1093/gji/ggae089

The influence of an anticline structure on ambient noise spectral anomalies at an underground gas storage

Christine El Khoury [1], Alexandre Kazantsev,[2] Damian Kula,[3,*] Arthur Dartois[2] and Hervé Chauris[1]

[1] *Centre de Géosciences, MINES Paris – PSL Research University, Fontainebleau* 77300, *France. E-mail: christine_kh@hotmail.com*
[2] *Storengy SAS – EISE/DGSM, Engie Group, Bois-Colombes* 92270, *France*
[3] *Institut Terre et Environnement de Strasbourg, Université de Strasbourg/EOS/ENGEES, CNRS UMR* 7063, 5 *rue Descartes, Strasbourg F*-67084, *France*

Accepted 2024 March 5. Received 2024 February 21; in original form 2023 June 27

SUMMARY

The purpose of this study is to investigate the seismic ambient noise spectral anomalies that occur near gas reservoirs. These anomalies involve a significant spectral amplification of the vertical component for frequencies generally between 1.5 and 4 Hz and have been reported at various hydrocarbon sites worldwide. There are differing views on the mechanisms responsible for these anomalies. The guideline for this study is that many hydrocarbon reservoirs share a common geological feature: an anticline structure. It appears to cause site effects that influence the amplitude of the ambient noise wavefield. This research examines a dense real data set of ambient noise recorded at the Chémery underground gas storage site in France. The analysis identifies stable spectral anomalies between 1.2 and 2.4 Hz that are correlated to the position of the anticline structure, which also corresponds to the position of the gas bubble. We use a beamforming technique to study the composition and the origin of the ambient noise, and show that the variations of the spectral anomalies over time are correlated to changes in the source wavefield. Finally, we perform numerical simulations of Rayleigh wave propagation within a realistic 3-D velocity model of the Chémery site, while using source distributions directly extracted from real data analysis. The comparison of the simulated anomalies with real data yields a satisfactory qualitative fit. We conclude that the fundamental-mode Rayleigh wave site effect on the anticline is the main mechanism of the spectral anomaly.

Key words: Numerical modelling; Seismic noise; Site effects; Surface waves and free oscillations; Wave propagation.

1 INTRODUCTION

Seismic ambient noise refers to the low-amplitude continuous vibrations of the Earth that are not related to specific seismic events generated by individual sources. The sources of seismic ambient noise are categorized into two main groups: natural and human sources, primarily differing in their frequency content (Bonnefoy-Claudet *et al.* 2006). Natural noise at lower frequencies ($f < 1$ Hz) is mainly generated by ocean activities and large-scale meteorological conditions, while higher frequencies ($f > 1$ Hz) are likely dominated by human activities in dense enough populated areas. However, the accuracy of the 1 Hz limit varies, depending on the site and the noise sources dominating in the area (Guillier *et al.* 2007). The seismic ambient noise wavefield consists of two types of waves, body waves and surface (Rayleigh and Love) waves. Numerous studies have explored the composition of ambient noise wavefields at different locations worldwide. Nishida (2017) conducted a review focusing on the ambient seismic wavefield below 1 Hz. Bonnefoy-Claudet *et al.* (2006) compared the findings of multiple authors and concluded that the ambient wavefield usually comprises a mixture of *P* waves, fundamental-mode Rayleigh waves, and higher modes. While the fundamental mode typically dominates below 1 Hz, the wavefield at higher frequencies often becomes more complex. According to Bonnefoy-Claudet *et al.* (2006), the noise composition depends on various factors, including the distance and the nature of the source (natural or human), as well as site conditions and soil properties.

Dangel *et al.* (2003) reported ambient noise spectral anomalies characterized by spectral amplification peaks on the vertical component within a specific frequency band (1.5–4 Hz) above 15 hydrocarbon reservoirs at different locations in the world, mainly in the Middle East. The work of Dangel *et al.* (2003) has been a trigger for many researchers to reproduce the observations at new

*Independent researcher.

sites and to investigate on the possible mechanisms of the reported low-frequency microtremors. An industrial application has been attempted with the so-called 'low-frequency passive seismic' method which analyses the spectrum of the ambient seismic noise recorded with broad-band seismometers (see patents: Saenger 2008; Saenger 2009; Saenger et al. 2009; Podladchikov *et al.* 2010; Kelly *et al.* 2013). Useful information for locating potential fluid reservoirs was meant to be extracted from ambient noise spectral signatures of hydrocarbon microtremors, that were considered as 'Direct Hydrocarbon Indicators'. The method aimed at optimizing exploration well locations and possibly monitoring reservoirs.

The studies provide different and controversial explanations for the exact nature of the observed spectral anomalies. The focus of many authors was to identify and/or model an emission source or a filtering mechanism taking place inside the reservoir layer due to some fluid-specific effects (Steiner *et al.* 2008; Holzner *et al.* 2009; Frehner *et al.* 2009; Lambert *et al.* 2009; Saenger *et al.* 2009; Goertz *et al.* 2012). On the contrary, other authors tended to show that the observed spectral anomalies were not really significant or general, since they might be related to some local human noise sources or near-surface geology (Hanssen & Bussat 2008; Green & Greenhalgh 2010; Ali *et al.* 2010; De Vasconcelos Lopes & Nunes 2010; Martini *et al.* 2013).

From this perspective, our study focuses on a structural effect (anticline) which might be general enough to explain the observed spectral anomalies at several different reservoirs. In the following, we evaluate the hypothesis of an anticline-induced spectral anomaly using real data and numerical simulations. We stress that our purpose is to understand a phenomenon which is observed in a specific frequency band (1.5–4 Hz), rather than *selecting* an optimal frequency band for reservoir imaging/monitoring with ambient noise. For instance, the presence of surface wave effects in this frequency band implies that the observed spectral anomalies are more likely generated by shallower features in the local geology rather than a deep reservoir. On the contrary, mechanisms involving body waves would be needed if the observed spectral anomalies were related to a deep reservoir.

In this work, we choose to focus on the vertical component data only, even though 3-Component seismograms are available in the data set. We extract vertical component amplification maps at different frequencies using an analog of the standard spectral ratio (SSR) method (see Parolai 2012). Closely related in practice, the horizontal-to-vertical spectral ratio method (HVSR, see Bard *et al.* 2008; Lunedei & Malischewsky 2015) is a well-known ambient-noise application for shallow shear velocity profile characterization. Its most common application is to identify the first strong velocity contrast between the bedrock and the overburden. An important feature of the HVSR method is that the peak frequency of the horizontal component amplitude associated to the first strong velocity contrast is expected to be approximately the same for Rayleigh, Love, and body waves, which makes the method rather stable with respect to the ambient noise composition (see the discussion in Lambert 2010). A more extensive interpretation of the HVSR method is to invert the H/V spectral ratio curve over a range of frequencies, using a direct model based either on the surface wave theory (e.g. Rayleigh wave ellipticity, see application to inversion in Lin *et al.* 2012a) or on the diffuse wavefield assumption (see application to inversion in Spica *et al.* 2018). Interestingly, the H/V ratio has been adapted to the study of ambient noise amplification above hydrocarbon reservoirs, through the use of its inverse: the V/H ratio. The peaks of this new spectral ratio have often been interpreted as a sign of strong vertically polarized body waves emanating from a local source zone located directly beneath the receiver (Lambert *et al.* 2009; Saenger *et al.* 2009; Goertz *et al.* 2012). In a recent example, this interpretation has been applied to the study of mud volcanoes at much smaller scales (Antunes *et al.* 2022). In our case, as we are testing a surface-wave-based interpretation of the amplification phenomena, introducing the H/V or V/H spectral ratios would require to model Love waves in addition to Rayleigh waves, as well as to constrain the energy proportions between Rayleigh and Love modes from the real data. For this reason, we decided to discard the horizontal component vibrations from the scope of this study, in order to not overcomplicate both the modelling and the data processing parts. An extension of our study to H/V ratio modelling would be a valuable development in the future.

This work is the first real-data application of our previous modelling works on the interaction of Rayleigh waves with an anticline structure (El Khoury *et al.* 2023), where a combination of 1-D analytical and 3-D numerical simulations on simplified models was employed. The conclusion was that an anticline structure could indeed induce effects similar to the spectral anomalies reported in the literature. Here, instead of using simplified models, we perform 3-D numerical simulations in a realistic velocity model of the gas storage site.

Our work contains two parts: real data analysis and wave propagation modelling. We study the seismic ambient noise recorded at an underground gas storage site (UGS) located in Chémery, France (Sections 2 and 3). We are mainly interested in the vertical component recordings above 1 Hz. First, spectral anomalies are identified in the data and mapped, while the wavefield composition is studied over the same time windows. Second, numerical simulations are conducted within a realistic 3-D velocity model of the Chémery site in Section 4, and the obtained results are compared with real data. Finally, in the discussion part, the data from another UGS site (Saint-Illiers, France) are briefly presented to confirm the conclusions derived from the analysis performed on the Chémery site.

2 DATA AND METHODS

2.1 Data acquisition

The Chémery site is located inside the Paris sedimentary basin (Fig. 1). We refer to Peruzzetto *et al.* (2018) for a more detailed geological description. The reservoir is located at a depth of around 1100 m, while an anticline structure of relative elevation difference of approximately 100 m serves as a gas trap. This elevation difference between the top of the anticline and the surrounding area will be referred to as the 'anticline prominence'.

About 100 broad-band Trillium 40 s three-component seismometers (Nanometrics) recording with a sampling frequency of 100 Hz were deployed in 2010 April and 2010 November with various geometries. From April 20 to 24, and from November 4 to 11, most of the sensors were operational between 3 pm and 6 am local time (UTC+1), and were shifted by about 250 m every day to achieve the most dense coverage possible. There was an almost uniform spacing of 500 m between the sensors installed over a 3 km × 6 km area located above the reservoir. The recording configurations varied between rectangular arrays and two linear profiles (Fig. 2a). The stations are superposed on the elevation map of the reservoir top layer. The central zone with higher elevations constitutes an anticline shape below which the reservoir is located. A structural fault bounds the anticline structure in the North. The anticline deformation affects all the geological layers above the reservoir.

Figure 1. Location of the Chémery and Saint-Illiers UGS sites (source: maps.google.com).

2.2 Spectral anomalies

Traces recorded at different stations are sliced into 30-min chunks. The purpose is to study the signal and source characteristics at different periods of the day. Pre-processing steps are applied to the sliced traces such as instrument response removal, detrending, resampling, tapering, filtering and initial quality control. For the initial quality control, skewness and kurtosis are computed on each 30-min chunk of data samples in time domain. Chunks with skewness ouside $[-2, 2]$ or kurtosis ouside $[-100, 100]$ are rejected as a whole, which allows to discard some evident electronic parasites or periods with human operations in the immediate vicinity of the stations. To estimate the power spectral density (PSD) of the signals, we divide each 30-min chunk into 40 s subwindows $u_i(t)$ overlapping by 50 per cent for each component i, and apply a taper in the time domain. Then, we perform a fast Fourier transform to get $u_i(f)$ in frequency domain and we calculate the corresponding PSD as

$$\mathrm{PSD}_i(f) = \frac{2}{f_s N_f}[u_i(f)]^2, \tag{1}$$

with f_s the sampling frequency, N_f the number of samples in the frequency domain (equal to half the number of samples in the time domain, see Bormann & Wielandt 2013). We average the PSD of all the subwindows inside the 30-min chunk to get a PSD spectrum at each station every half-an-hour. For visual representation, we use 10 times the decimal logarithm of the PSD to transform it to a decibel scale with respect to 1 $\mathrm{m}^2\,\mathrm{s}^{-1}$.

Subsequently, a second stage of quality control is performed in order to reject stations with anomalous spectra within a *control* frequency band, for each 30-min chunk. In our analysis, we set this band to [0.4–1.5] Hz. This band is low-frequency enough to expect relatively stable spectra at the network scale, so that strong deviations would probably indicate a corrupted or non-representative signal. At each frequency f_i within the control frequency band, we calculate the mean $\mu(f_i)$ and the standard deviation $\sigma(f_i)$ of $10\log_{10}\mathrm{PSD}(f_i)$ over the stations. Stations with $10\log_{10}\mathrm{PSD}(\mathrm{f_i})$ values beyond $\mu(f_i) \pm 2.5\sigma(f_i)$ are flagged as outliers at the frequency f_i, and the procedure is iterated until there are no more stations to reject. After processing all the frequencies within the control frequency band, those stations which are flagged as outliers for more than 33 per cent of frequencies are completely rejected, while the others are kept for further analysis. This procedure usually results into rejecting 1–3 per cent of stations per data chunk of 30 min.

To better analyse the correlation of the spectral amplitudes with the anticline structure, we introduce the attribute p corresponding to the spectral anomaly. It is calculated by

$$p = 10\log_{10}\left(\frac{\mathrm{PSD}_{Z,s}}{\mathrm{PSD}_{Z,\mathrm{ref}}}\right), \tag{2}$$

with $\mathrm{PSD}_{Z,\,s}$ computed at a station s and $\mathrm{PSD}_{Z,\,\mathrm{ref}}$ computed at a reference station. In practice, since there were no fixed stations over the whole duration of the survey, $\mathrm{PSD}_{Z,\,\mathrm{ref}}$ is calculated as the mean of spectra of stations in a reference zone chosen in the north-west part of the survey area. The choice of the reference zone will be illustrated and discussed in Sections 3.1 and 5. Thanks to this normalization inspired by the SSR method (see Parolai 2012), spectral anomaly maps can be made independent from the absolute noise levels occurring at different dates.

2.3 Array analysis

To further analyse the ambient noise, we perform array analysis using a beamforming technique (see Rost & Thomas 2002, for an overview) on time windows where rectangular survey configurations were used. Beamforming analysis enables the retrieval of the incident plane wavefield power as an image in the slowness space (s_x, s_y) at each frequency, where body and surface waves modes can be clearly isolated based on their different velocities, provided enough radial resolution. The advantage of using a method from the beamforming family, compared to other higher resolution methods such as MUSIC (Multiple SIgnal Classification, Schmidt 1981, 1986) or HRFK (High Resolution Frequency Wavenumber, Capon 1969), is that it allows to retrieve an unbiased estimate of the wave power. Indeed, the beam maps obtained from the previous methods are not directly proportional to the physical wave power because they are using non-linear operators (Cornou 2002). Here, we also use a new deconvolution technique for array response function (ARF) removal. It is based on the Richardson–Lucy deconvolution method (RLC, Richardson 1972; Lucy 1974; Bertero & Boccacci 2005) previously proposed by Gal *et al.* (2016) for seismological applications. The key step in RLC is an iterative multiplicative update of the solution:

$$f^{(k+1)} = \frac{\mathbf{1}_{n,n}}{\alpha}\cdot\left(\frac{g}{f^{(k)} * A}\ast A^T\right).f^{(k)}, \tag{3}$$

where f is the iteratively updated solution image of size (n, n) (response-corrected beam power), g is the input image of same size (raw beam power), A is the ARF matrix of arbitrary large size and $\alpha = \mathbf{1}_{n,n} * A$, with $\mathbf{1}_{n,n}$ a square (n, n) matrix having all elements equal to 1. The convolution $X * Y$ stands for 2-D convolution of images X and Y, restricted to the size of X. Products and divisions are pixelwise in the image domain. We expect physically acceptable solutions for ambient noise to be arc portions of azimuthal extensions corresponding to the source zones, at radii corresponding to the local phase velocity of the waves. In order to prioritize physically acceptable solutions f, we smooth the multiplicative update term on the left of $f^{(k)}$ in eq. (3) with an ad-hoc family of kernels which enforces constant-by-arc updates, at each iteration. We start with updates which look like 'rings', and progressively reduce the angular extension of the smoothing kernels. Subsequently, to clarify notations, we call $B_0(f, s_x, s_y)$ the final beam power retrieved at the frequency f, and $P_0(f, s, \theta)$ its polar transform. This hierarchic

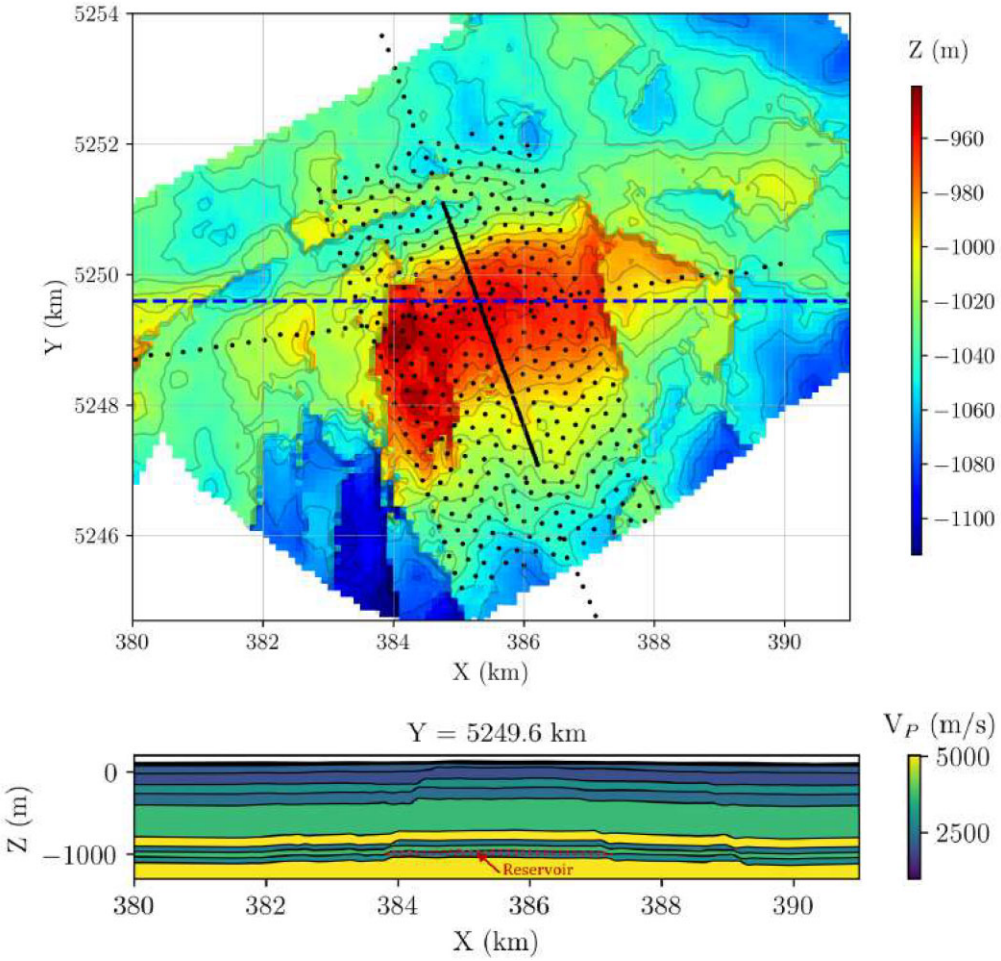

Figure 2. (a) Station network superposed on the elevation map of the reservoir top layer. Each black dot represents a temporary recording location. X and Y coordinates refer to the UTM metric system, zone 31T. (b) Vertical cross-section at $Y = 5249.6$ km of the V_P model of the Chémery site showing the reservoir inclusion with a velocity drop set to $\delta v_P/v_P = -18$ per cent with respect to the surrounding layer. The position of the section is indicated with the long dashed line on the top figure. The V_P model was build from 3-D reflection seismic data constrained by sonic logs, provided by Storengy.

approach allows to explain as much data as possible with physically acceptable solutions. Our modification enables a significant enlargement of the working frequency (or equivalently wavelength) interval of the array beyond the standard resolution and aliasing limits.

While a classical beamforming result is represented as a power map in the slowness space (s_x, s_y) for each frequency, an alternative representation is the slowness-frequency map. The latter gives access to the dispersion diagram of the signal. In order to achieve a clear visual representation of the dispersion diagram $D(f, s)$, we use a slightly customized normalization method at each frequency:

$$D(f,s) = s \cdot \frac{P_s(f,s)}{\max_s P_s(f,s)}, \tag{4}$$

where $P_s(f,s) = \int_0^{2\pi} \langle P_0(f,s,\theta)\rangle \mathrm{d}\theta$ is the beam power $\langle P_0\rangle$ (averaged over all time windows) in polar slowness coordinates integrated over backazimuths at frequency f. The normalization by $\max_s P_s(f,s)$ ensures consistency in the maximum value across all frequencies, regardless the absolute power spectrum. The multiplication by s comes from the polar integrand definition: the total beam power integrated over all the plane waves writes $P_{\mathrm{tot}}(f) = \int_{s_{\min}}^{s_{\max}} s \cdot P_s(f,s)\mathrm{d}s$. Thus, each column of $D(f, s)$ is proportional to the total beam power integrand over the slowness, at each frequency. The advantage of this representation is that it naturally preserves all the propagation modes, including those which are not dominant at a given frequency. Such secondary modes can be lost in representations based on peak slowness histograms, another common approach of constructing dispersion diagrams of ambient noise.

2.4 Clustering source distributions

Using the beamforming results, different types of incident waves can be identified, as well as their azimuthal origin. We use K-means clustering (MacQueen 1967; Hartigan & Wong 1979) to identify several source azimuthal distributions typical for the Chémery data set, and to associate each 30 min window of recording to one of these typical distributions. The clusters are constructed so that the source distributions within the same cluster are as similar as possible (minimizing intracluster variance), while the source distributions from different clusters are as dissimilar as possible (maximizing intercluster variance). We used the Elbow, the Silhouette and the gap statistic methods (Tibshirani *et al.* 2001; Kassambara 2017; Kaoungku *et al.* 2018) to get independent estimates of the optimal number of clusters. After constructing the clusters, we compute the average of both the spectral anomaly maps and the source distributions over the time windows belonging to the same cluster. In this

way, each average spectral anomaly pattern can be associated to an average source distribution, per cluster.

3 RESULTS

3.1 Spectral anomalies

The PSD_Z are computed for each station and for each 30-min chunk. In Fig. 3, the dots represent stations with the colour of the anticline elevation map at the corresponding location. Each PSD_Z curve is plotted with the corresponding colour, for three different days in April at 4 am, corresponding to the three main acquisition geometries. The amplification for frequencies above 1.5 Hz is clear for the stations in red, which are located above the anticline structure. The peak frequency of amplification ranges from 1.6 to 1.8 Hz for both the rectangular configuration and the east–west linear profile. Therefore, we conclude that there is some degree of correlation between the anticline top and the high spectral amplitudes.

Normalized spectral anomaly maps must be constructed in order to quantify this effect. For this purpose, eq. (1) requires the choice of a reference zone. This choice is illustrated in Fig. 4, where average PSD spectra between UTC 00:00 and 04:00 within the reference zone are superimposed onto the full 'cloud' of spectra recorded by the network during the four days corresponding to the rectangular survey configurations in 2010 April. The reference zone contains between 6 and 8 stations each day. This zone is not affected by the anticline deformation, which makes it an appropriate reference without any expected anticline-induced spectral anomalies. Stations within the reference zone present generally smooth and monotonic spectra between 1.2 and 2.4 Hz, that is, the lower end of the *a priori* frequency band of interest (1.5–4 Hz). The reference spectra are located on the lower edge of the amplitude distribution within the highlighted frequency band (1.2–2.4 Hz). Consistently with this observation, the average PSD spectral anomalies in this frequency band, represented in Fig. 4, are mostly positive. As already observed from Fig. 3, they tend to outline the anticline structure in the centre of the network. Beyond 2.4 Hz, it seems that the spectra become less stable in space and time.

Once the reference zone has been fixed, normalized spectral anomaly maps are constructed at different frequencies for the rectangular survey configurations, and then interpolated on a finer grid in order to analyse the spectral anomaly patterns in space. Based on the previous paragraph and Fig. 4, our main frequency band of interest is 1.2–2.4 Hz, though we will extend our analysis slightly beyond. The positive spectral anomalies for a 30 min time window on April 21 at 4 am (Fig. 5) are concentrated in the central zone of the map. They reach a maximum amplitude of 6 dB between 1.7 and 1.9 Hz. As a next step, we calculate the average PSD_Z for all the time windows in April for the days where a rectangular survey configuration was used (from April 20 to 23). We then compute the corresponding spectral anomaly maps (Fig. 5b). While the positive spectral anomaly remains concentrated above the anticline, we also see that time averaging allows to achieve smoother maps. The maximum spectral anomaly is about 4 dB after averaging.

As previously mentioned, two surveys were performed for the Chémery site, one in April (minimum gas volume in the UGS) and the other in 2010 November (maximum gas volume in the UGS). The type of wave present in the ambient noise is also likely to differ between these two periods because of seasonal variations. Therefore, it is interesting to compare the differences in noise spectra between these surveys to evaluate the extent to which gas content and/or seasonal noise variations influence the spectral anomalies. Similarly to Fig. 5(b), we compute the averaged spectral anomalies over all the time windows in November (from November 4 to 10, Fig. 6). The spectral anomaly patterns exhibit similarities between April and November, and the correlation with the geological structure is visible for the two periods. However, the spectral anomaly amplitudes are higher in April than in November: a maximum anomaly of 4 dB is reached in April compared to 1.3 dB in November (Fig. 6). It must be stressed that the spectral anomalies also exhibit daily variations which have been smoothed out through averaging in Fig. 6 (3 d in April and 6 d in November). This aspect will be discussed in the next section, where we will assess the correlation between variations in spectral anomaly strength and daily or seasonal variations of the ambient wavefield composition in terms of Rayleigh and body waves.

3.2 Array analysis

Using the beamforming results computed for each 30-min chunk, we present the slowness-frequency maps averaged over the 3 d of the April survey, and the 6 d of the November survey (Fig. 7). Beamforming analysis is limited by the upper frequency bound of 3 Hz due to aliasing and lack of coherence at higher frequencies. Though some artefacts appear due to aliasing effects, the fundamental-mode Rayleigh wave (called R0 in the rest of the paper) is clearly identified above 1 Hz for both April and November surveys. The associated dispersion curves also extend to lower frequencies but with lower amplitudes. Higher modes are dominant between 0.5 and 1 Hz. In November, low slowness waves (probably *P* waves) are more pronounced above 1 Hz, and can have a power comparable to that of R0. In April, in contrast, R0 dominates above 1 Hz. These observations are consistent with those of Peruzzetto *et al.* (2018) who applied different array analysis techniques to the same data set. This analysis shows that Rayleigh waves are consistently present (both in April and November) within the previously identified spectral anomaly band, between 1.2 and 2.4 Hz, while body waves are intermittent (strong in November and nearly absent in April). On the other hand, the spectral anomalies are significantly stronger in April (see Fig. 6). This is a first indication of the fact that Rayleigh waves might play a key role in the spectral anomaly mechanism.

The link between the wave composition of the ambient noise and the spectral anomalies can be further studied thanks to detailed analysis of continuous time-series of both amplitude and wavefield composition estimated over short time windows. In Fig. 8, we represent simultaneously the beam power and the corresponding spectral anomaly map for each 30 min time window during 24 hr in April. The time windows between 7:30 and 14:30 are omitted because stations were being moved in the field from one position to another at this time of the day. Each beam power is normalized by the maximum, but one should keep in mind that the beam power tends to be higher during daytime hours due to human activity. The previously mentioned fast *P* waves appear as a dot or a very small circle in the middle of the beam power map (low slowness values), while Rayleigh wave modes appear as larger circles (large slowness values). Here, we must stress that body waves, despite their generally low amplitude in comparison to Rayleigh waves at 1.5 Hz in April (see Fig. 7), are still visible in specific time windows (Fig. 8). A first observation is that the time windows with strong Rayleigh waves (e.g. 15:00) tend to exhibit a spectral anomaly concentrated above the anticline, while time windows with weaker Rayleigh waves (e.g. 01:30) tend to exhibit a weaker and more diffuse spectral anomaly.

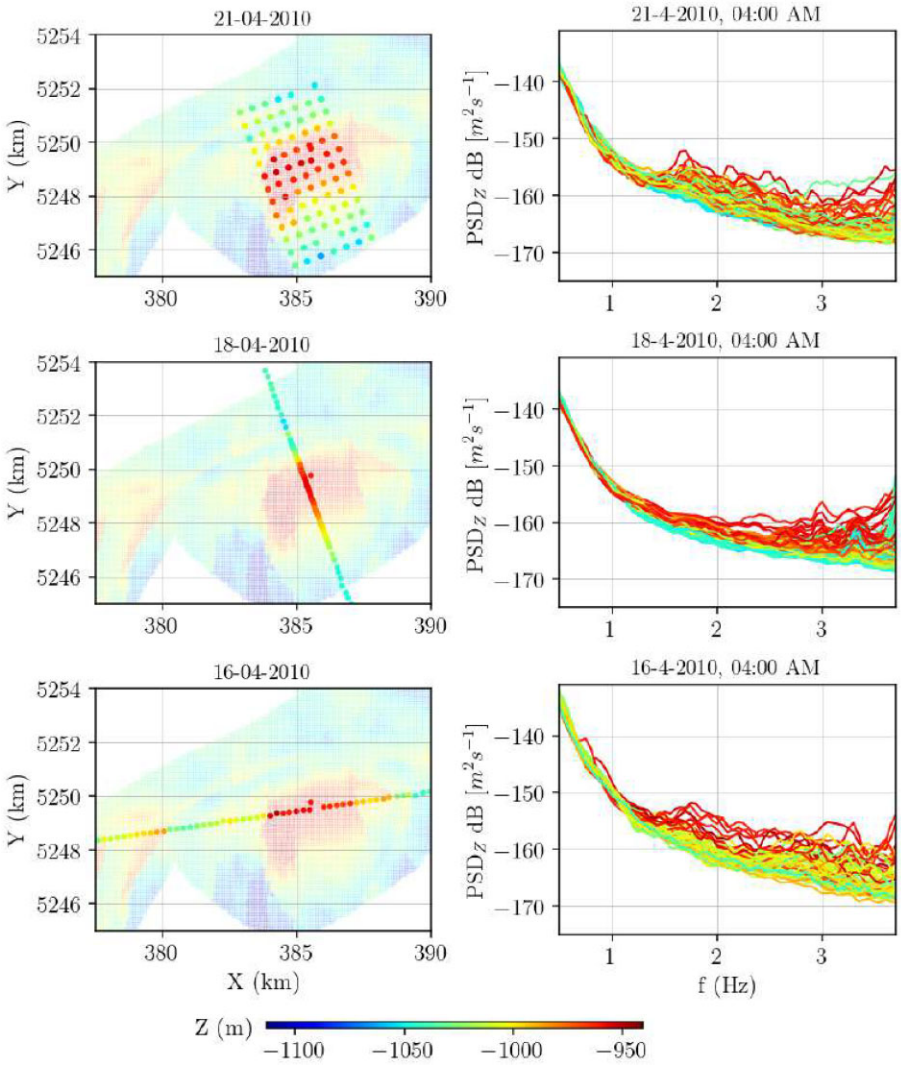

Figure 3. Left-hand column: three station configurations for three different days in April, superposed on the structural map of the reservoir top layer. Right-hand column: PSD$_Z$ calculated for the different stations of each configuration. The stations' colour in the left-hand column and the associated spectrum colour in the right-hand column correspond to the reservoir top elevation at each location.

This observation, made at the daily scale, is similar to what was observed above at the seasonal scale. Indeed, the spectral anomaly was stronger in April, when the Rayleigh waves were dominant, compared to November, when the body and Rayleigh waves have similar energies above 1 Hz. This tends to confirm that dominant Rayleigh waves are necessary to generate a strong spectral anomaly.

If we focus on the Rayleigh wave sources, we note two main backazimuths for R0: south and north-west (Fig. 8). We also observe that for some time windows, the positive spectral anomaly (red colour) spreads out to the south. These time windows correspond to dominant sources located mainly in the south (e.g. at 2:30, 3:00, 4:00, 19:00, 21:30, 00:30). For other time windows, the positive spectral anomaly is concentrated in the central zone with dominant sources located at two backazimuths: south and north-west (e.g. from 15:00 to 18:00). However, deriving reliable conclusions from visual analysis alone can be challenging, which underscores the need for a clustering approach outlined in the next section.

3.3 Clustering source distributions

In this section, we focus on the azimuthal source distribution of R0 and investigate how it is correlated with the spectral anomaly patterns. 97 time windows of 30 min are available in April for the 3 d where rectangular survey configurations were used. From the beam power maps, we extract the R0 distribution over the backazimuths for each time window at a specified frequency. We choose to perform this analysis at $f = 1.5$ Hz because this frequency is minimally affected by aliasing, as well as it presents intermittently dominant body and surface waves. Tests performed at different frequencies showed similar results, though more contaminated by beamforming artefacts. The slowness of the R0 mode ranges between 1 and 1.3 s km^{-1}. Thus, the azimuthal distribution of R0 at this frequency is obtained by integrating the beam power over the slowness range [1, 1.3] s km^{-1}. We fit each distribution with a function $A(\phi)$ expressed as a sum of cosine functions with increasing

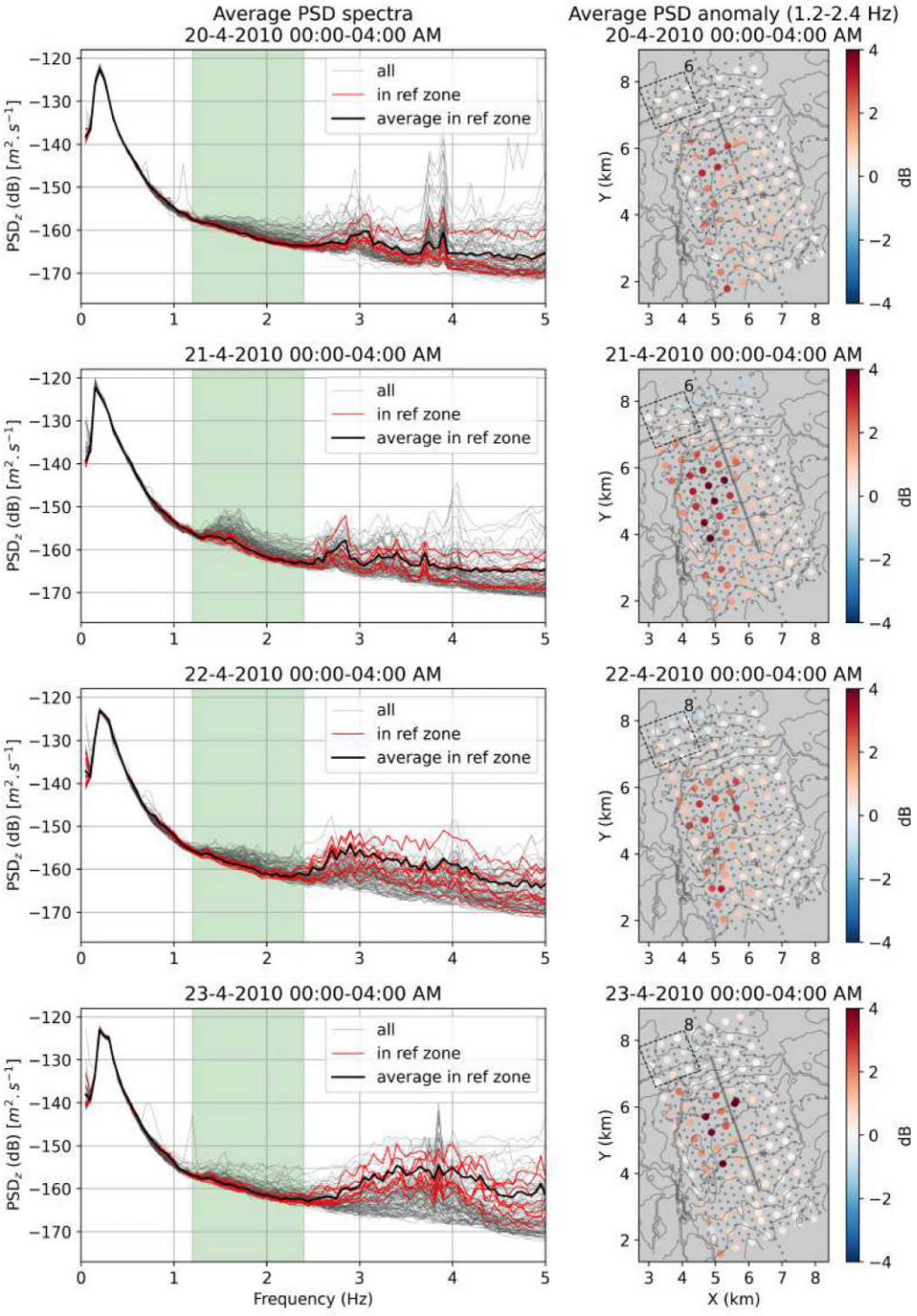

Figure 4. Left-hand column: PSD spectra averaged between UTC 00:00 and 04:00 for the 4 d corresponding to rectangular network configurations. The frequency band of interest for the spectral anomaly analysis is highlighted in green. Right-hand column: average normalized spectral anomalies over the frequency band of interest. The reference zone is shown with the dashed rectangle with the number of stations inside. The grey dots correspond to all the recording locations on different days. Here and in the following plots, X and Y are relative to a chosen UTM reference coordinate (379.9 km, 5243.6 km), while the background corresponds to contour lines of the elevation map of the reservoir top layer.

angular frequencies (Weaver *et al.* 2009):

$$A(\phi) = a_0 + a_1 \cos\left(\frac{\phi}{2}\right) + a_2 \cos\left(2\frac{\phi}{2}\right) + a_3 \cos\left(3\frac{\phi}{2}\right) + ..., \tag{5}$$

with ϕ the trigonometric angle. Note that in all the next figures, we represent the trigonometric angle ϕ and not the backazimuth defined as $\pi/2 - \phi$. In Fig. 9, we show an example of the original distribution as a function of ϕ for one time window, as well as the corresponding fitted $A(\phi)$ function. We found that a reasonable fit

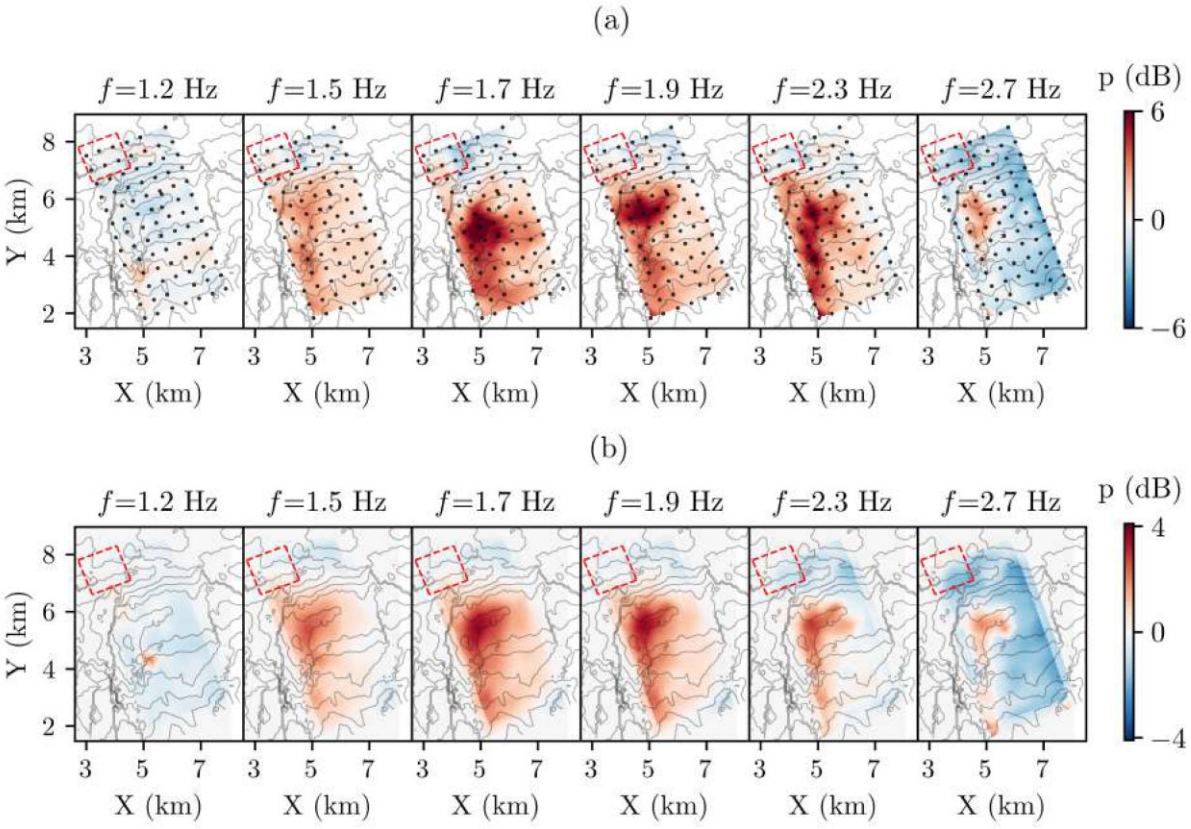

Figure 5. (a) Spectral anomaly maps at different frequencies calculated for a 30 mi time window in April 21 at 4 am. The black dots correspond to the station network and the red dashed square is the reference zone for the spectral anomaly calculation. (b) Same as for (a), but for frequencies averaged over 3 d in April.

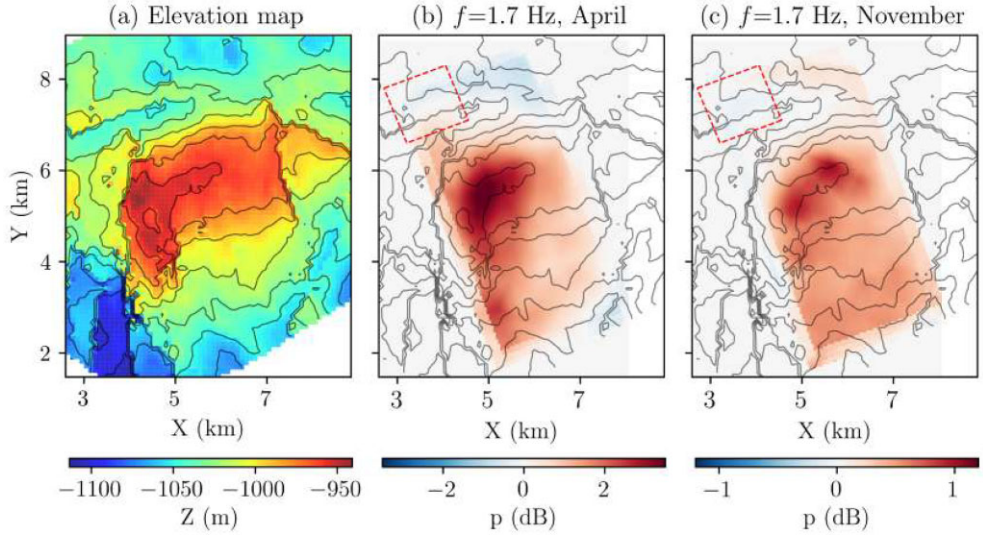

Figure 6. (a) Elevation map of the reservoir top layer. (b) and (c) Spectral anomaly maps averaged over 3 d in April and 6 d in November for $f = 1.7$ Hz. Beware of the colour scale difference between April and November.

(RMS values smaller than 0.13) for all time windows can be obtained with seven coefficients a_i in the function $A(\phi)$. The use of the $A(\phi)$ fitting function allows to reduce the dimensionality of the clustering problem. The clustering is applied on the seven coefficients a_0 to a_6 of the function $A(\phi)$, instead of the full azimuthal grid. For each coefficient a_i, 73 observations are available, corresponding to the total number of time windows. The optimal number of clusters chosen via the statistical methods mentioned earlier is eight.

As a final result, we obtain eight groups of R0 distributions (Fig. 10). The black curves represent the average over all distributions in each cluster. The colour of each individual distribution curve corresponds to an hour of the day (following the colour bar). An interesting observation is that some clusters are correlated to the

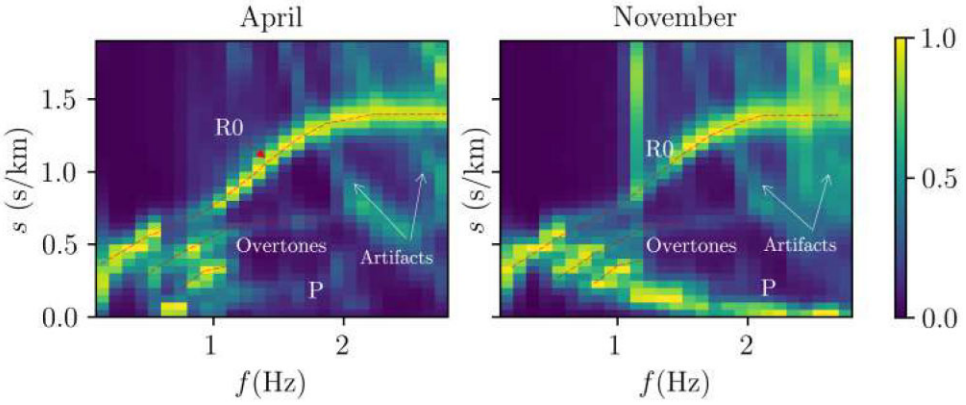

Figure 7. Slowness-frequency maps averaged over the April survey (left) and the November survey (right).

hour of the day. This is mostly visible for clusters C1, C2, C5 and C8. Most of the distributions in C1 correspond to mid-night hours, with sources having one dominant backazimuth (South). The same applies for cluster C2 which corresponds to night hours. C5 mostly corresponds to the late afternoon, with a secondary source zone starting to emerge in the north-western direction. C8 concentrates between noon and 3 pm, with north-western sources dominating over the southern sources.

The final step is to average the spectral anomaly maps corresponding to time windows within the same cluster (Fig. 10). The maps are different from one cluster to another. For clusters C1 and C2, where the sources have a dominant backazimuth in the south, the positive spectral anomaly spreads out to the south. In other clusters where the north-western sources are well visible, the spread out of the spectral anomaly to the south is weaker. Finally for cluster C8 where the north-west sources become dominant over the southern sources, the spectral anomaly is concentrated in the centre of the map.

The clustering analysis allowed us to obtain a reference spectral anomaly map per typical R0 source distribution, assuming that R0 sources are the main contributors to the spectral anomaly generation. The time windows in November were not included in the study due to their lower spectral amplitude anomalies and the influence of body waves, which could bias the results. The reference spectral anomaly maps will be compared in Section 4 with the spectral anomaly maps obtained from numerical simulations using a realistic 3-D velocity model of the Chémery site, as well as reference source distributions extracted from the clustering analysis.

4 NUMERICAL SIMULATIONS

While this paper focuses on a real-data example, Rayleigh wave spectral distortions induced by an anticline structure were studied on synthetic data sets in a previous paper (El Khoury *et al.* 2023) using both 1-D analytical and 3-D numerical modelling for the case of R0 sources placed in simplified models. Before we move to simulations in a realistic 3-D velocity model of the Chémery site, it might be useful to briefly recall the main results obtained for simplified anticline structures. First, we found that for uniformly distributed sources, the signature of the anticline is a positive spectral anomaly in the anticline zone for a large frequency range, reaching an amplitude comparable with real data observations (several dB). Second, a parametric study using 1-D analytical equations in a two-layer model has shown that a change in the layers elastic properties or in the anticline geometry causes a change in the spectral anomaly maximum amplitude and frequency, but the general shape of the spectral anomaly is preserved. Third and last, we also found that for sources located in a specific narrow backazimuth range, the wavefield interaction with the anticline generates (de)focusing effects. This superimposes extra positive and negative spectral anomaly patterns onto the pure 1-D spectral anomaly described above. More details can be found in El Khoury *et al.* (2023).

As in that previous work, we are using the spectral element method (SPECFEM3D software, Komatitsch & Vilotte 1998; Komatitsch *et al.* 1999) for 3-D wave propagation simulations presented in this paper. The novelty here is that we use a complex realistic model, to possibly explain the real data.

4.1 Simulations in the Chémery 3-D velocity model

To study the effect of the structure, we first neglect the presence of the gas inside the reservoir, which normally causes a limited change in the medium properties at the location of the reservoir (Fig. 2b). The only heterogeneity in the model that could cause spectral anomalies will thus be the anticline structure.

We perform simulations in a 3-D model derived from active seismic data and constrained by well logs (Fig. 2). We use the coupling technique described in El Khoury *et al.* (2023) to inject a 1-D modal solution of Rayleigh wave propagation at the boundaries of the 3D model used by SPECFEM3D. This approach allows to model pure Rayleigh waves eliminating any incident body wave energy. Moreover, it enables to study each Rayleigh wave mode seperatly as well as to significantly reduce the 3-D model size. We use point sources (with Ricker source wavelet of central frequency $f_0 = 1.5$ Hz) located outside the 3-D model. The sources are distributed along a circle of constant radius measured from the centre of the 3-D model, with a step of 10°. Thus, 36 simulations are performed in total (one per source). With a mesh size of 100 m and a model of dimensions 12 km × 12 km × 1.45 km, one simulation took around 45 min when parallelized on 100 CPU. For each simulation corresponding to one source, the output signals are recorded on a regular array of receivers at the surface separated by 100 m along the *X*- and *Y*-directions.

The PSD_Z on the vertical component are computed at each station (eq. 1) and the spectral anomaly p (eq. 2) is determined with respect to a reference zone set up in the real data analysis part (north-west of the survey area). However, in a 3-D simulation of finite duration and using a finite number of sources, p would not be zero even

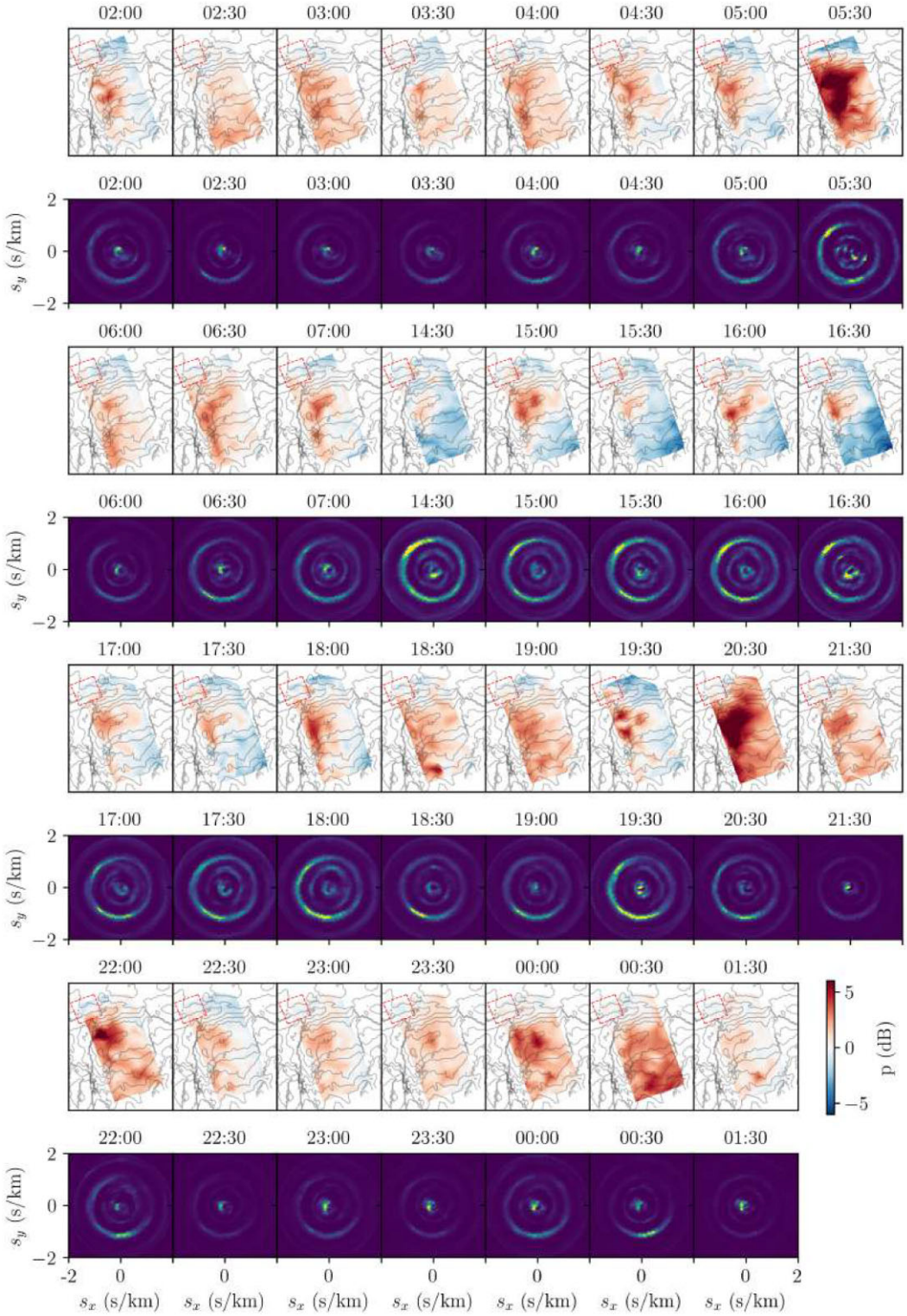

Figure 8. Spectral anomaly map variations and beam power variations between 30 min time windows along 24 hr in April at $f = 1.5$ Hz. The time windows between 7:30 and 14:30 are omitted due to parasitic noise caused by moving the stations while switching from one survey configuration to another.

in a laterally homogeneous model. This is due to the geometrical spreading across the simulation zone, and also to the small parasitic reflections on the model boundaries. In order to isolate the pure effect of the anticline structure, we perform one extra simulation for each source using a 1-D (laterally homogeneous) model without the anticline deformation. Then, we compute a corrected spectral anomaly p^* (eq. 6) by removing the parasitic anomalies existing in the 1-D model

$$p^*(x, y) = p_{3D}(x, y) - p_{1D}(x, y). \quad (6)$$

Before comparing the numerical spectral anomalies, under realistic source distributions, with the real data spectral anomalies, we

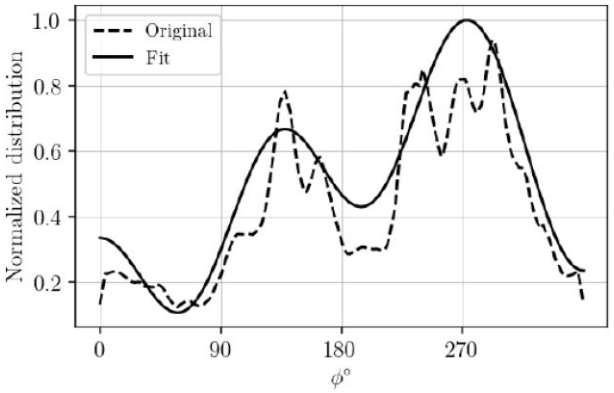

Figure 9. Original distribution of R0 sources along the trigonometric angle ϕ (dashed line) and the fitted function $A(\phi)$ (solid line) corresponding to one time window.

first consider the case of R0 sources uniformly distributed around the model. We average the PSD_Z of the 36 simulations at each station, and we calculate the corresponding spectral anomaly maps at different frequencies. Note that performing a simulation per source backazimuth is different from performing a simulation including all the sources at different backazimuths at once, as we are averaging the PSD and not the signal over different sources. Averaging PSD implicitely assumes uncorrelated noise sources, and thus the absence of interference between contributions of different backazimuths. Such interference would be inevitably present if we were to estimate the PSD from a single simulation containing all the sources. An alternative would be to perform several simulations containing a large number of randomly triggered sources in space and time, but in practice this would require a large number of simulations before the stabilization of the PSD. Moreover, this would not allow to easily test different scenarios for the source azimuthal distributions.

Results are shown in Fig. 11. The reference zone is shown with the red dashed square. The positive spectral anomaly in the central zone starts at $f = 1.5$ Hz and continues at higher frequencies, with a maximum spectral anomaly values of 5 dB around a frequency of 2 Hz. The correlation of the spectral anomaly maps with the anticline location (Fig. 6a) is very clear; the zone with highest spectral anomaly corresponds to the zone of the highest elevation of the anticline structure. The values reached by the positive spectral anomaly, as well as the frequency range and the frequency of maximum spectral anomaly, are very close to the ones extracted from the real data analysis (Section 3.1). Now that the key expected effect has been confirmed with idealized sources, we can move on to compare numerical and real data spectral anomaly maps with source distributions extracted from beamforming analysis on the Chémery site.

4.2 Comparison with real data spectral anomaly maps

To make a comparison between numerical and real data spectral anomalies, we use the R0 source distributions derived from beamforming analysis on real data. In Section 3.3, a clustering technique with the *K*-means algorithm was used to derive groups of source distributions presenting similarities. A reference spectral anomaly map was associated to each reference distribution (average of the distributions within each cluster). We choose three particular distributions among the eight clusters: C1, C5 and C8. C1 distribution is a typical case where the sources are mainly located in the South. For C5, the main backazimuth is the south but also other sources contribute from the north-west with lower amplitudes. Finally for C8, both south and north-west sources are present with a dominant north-west backazimuth (Figs 12a–c).

We use our database of 36 SPECFEM3D simulations performed using one source at each particular backazimuth. For each cluster distribution, we combine the simulations from the database by summing the PSD spectra at each station from all the simulations. The PSD spectra of each simulation are assigned a weight determined by the amplitude of the source distribution for the corresponding backazimuth (this approach again assumes uncorrelated sources from each backazimuth). Spectral anomaly maps derived from numerical simulations are plotted for the three cases (Figs 12g–i). They can be directly compared to real data averaged spectral anomalies (Figs 12d–f). This comparison is performed at a frequency of 1.5 Hz. It is the frequency for which the cluster analysis is performed and for which the resolution of the real data beamforming (used for source distribution extraction) is the most satisfying.

Real data and numerical spectral anomaly maps present important similarities. First, in both cases the amplitude of the spectral anomalies reaches the same maximum value of around 3.5 dB at $f = 1.5$ Hz. For C1, the positive spectral anomaly spreads out to the South with higher spectral anomalies in the anticline zone; this is also observed in the numerical spectral anomaly maps. For C5 and C8, the spectral anomaly in concentrated in the anticline zone with the highest contrast between positive and negative spectral anomalies for C8. To some extent, this is observed in numerical simulations. However, obvious discrepancies are observed between the real data and the simulations results. The real amplification maps are considerably smoother compared to the simulated ones. This might be because of (i) an insufficient number of simulations in our database (36 backazimuths only), as well as (ii) the absence of scattering in the wave propagation simulations. The first aspect might explain why we fail to completely cancel out (de)focusing patterns specific to each individual source backazimuth. The second aspect might explain why the (de)focusing patterns are overestimated in each individual simulation, compared to what a single-azimuth source would produce in the real data.

Note that for the comparison, we considered the Chémery real data spectral anomalies obtained for April data only, and we did not consider November. The reason is that the November data contains a high fraction of body wave energy at our frequencies of interest (>1 Hz). On the other hand, we recall that in our numerical simulations, we only model pure Rayleigh waves with a boundary injection technique. Thus, our simulations might have not been representative of the November data. The choice of modelling pure Rayleigh waves is due the fact we have demonstrated in Section 3 that R0 sources are the main contributors to the spectral anomalies. Modelling a mixture of body and Rayleigh waves can be done in the future, but is not within the scope of the current work.

4.3 Effect of adding a reservoir

The initial objective of analysing spectral anomalies above gas reservoirs was to understand their monitoring potential for the underground gas amount/location. Kazantsev (2018) studied the effect of the reservoir through 2-D numerical modelling. He modelled the reservoir as an elastic inclusion with a compression wave velocity drop $\delta v_P/v_P = -18$ per cent and a density drop $\delta\rho/\rho = -10$ per cent.

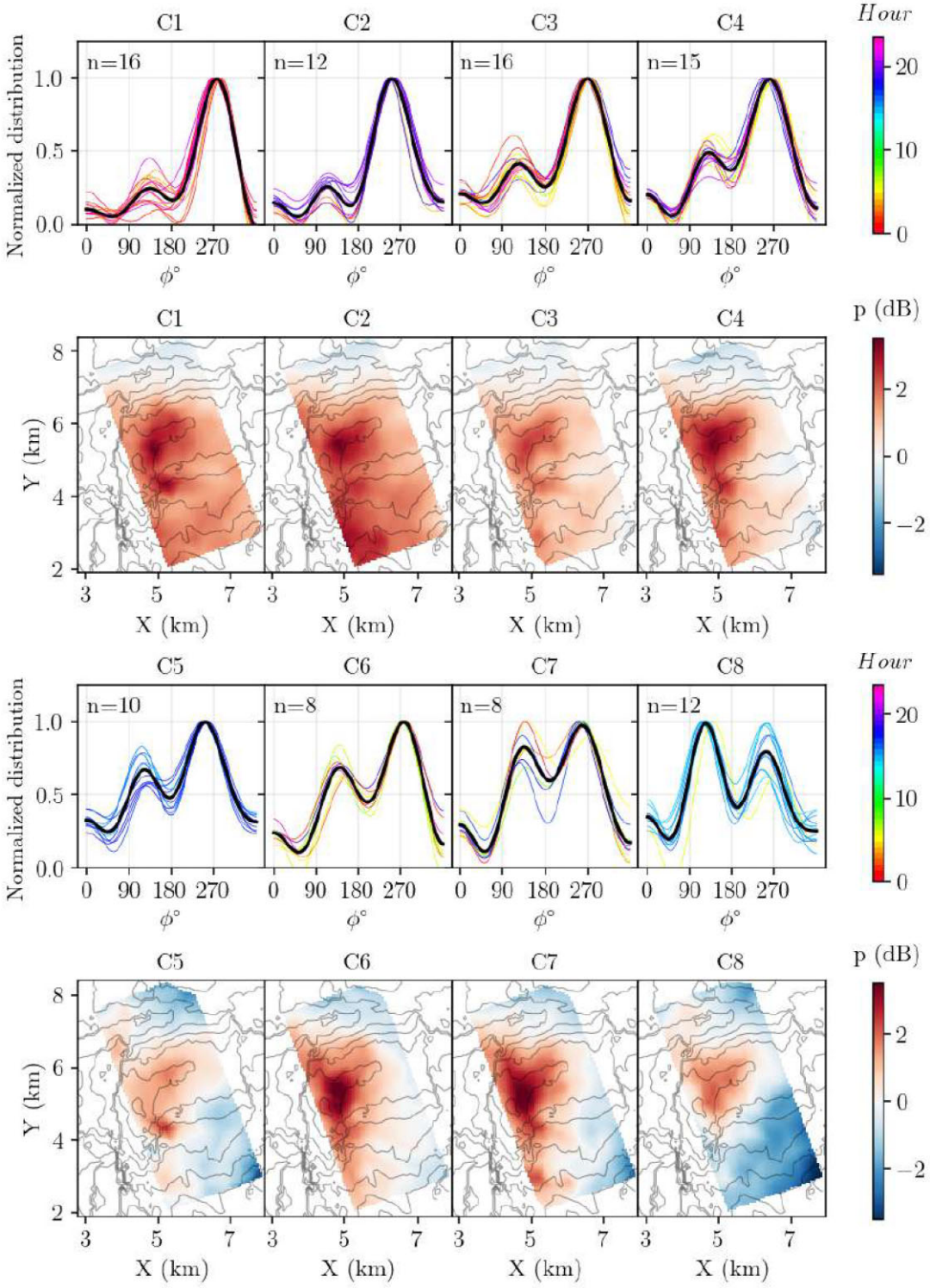

Figure 10. Results of R0 source distribution clustering (eight clusters), and the corresponding spectral anomaly maps. The curves are coloured according to the hour of the day and the thick black curve is the average of the distributions within the same cluster. The number *n* indicated in the top left of each subplot is the number of observations constituting each cluster. For each cluster, the corresponding spectral anomaly map is the average of the different time windows within the cluster.

These values are representative of a full substitution of water by gas in a sandstone reservoir similar to Chémery (Vidal *et al.* 2002). They give the maximum expected changes within the reservoir, since the full gas stock of the reservoir is never withdrawn/injected on yearly basis. We perform the same analysis as Kazantsev (2018), but using our 3-D modelling approach explained above. The reservoir maximum thickness is 60 m, while it is located at a depth of 1000 m with a lateral extent of up to 2000 m (Fig. 2b).

In Fig. 13, we show the anomaly map differences between simulations including the reservoir or not, using uniformly distributed

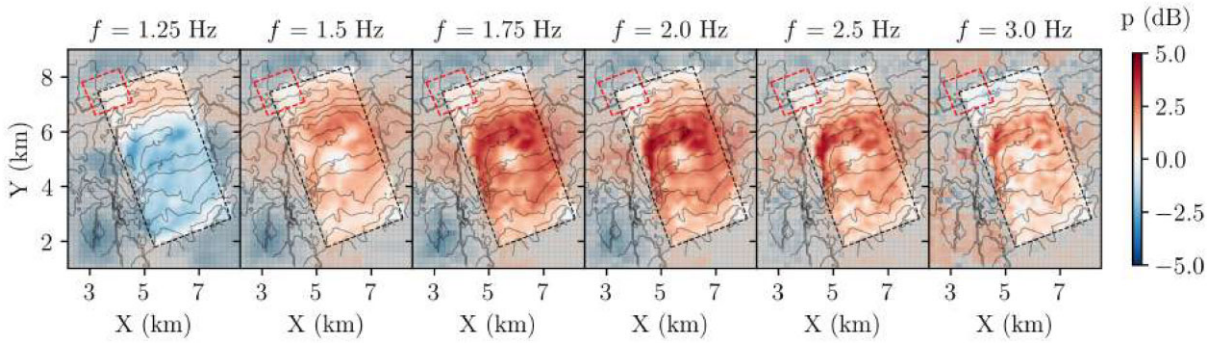

Figure 11. Spectral anomaly maps p^* at different frequencies obtained from numerical simulations for the Chémery site using R0 sources uniformly distributed along the backazimuths.

sources. The amplitude of the residual spectral anomaly generated by the reservoir inclusion is very low compared to the real data spectral anomalies and the simulated anticline-induced spectral anomalies. Also, the spectral anomaly spatial pattern is not correlated with the reservoir location, except for $f = 0.75$ Hz, where we have a red colour (positive anomaly) in the central zone; however it is not the frequency at which the real data spectral anomalies occur. Thus, our study tends to show that the spectral anomaly is mainly due to a static anticline-induced site effect, and the variation between a period of time and another is mainly related to the variation of the sources in the ambient noise wavefield.

5 DISCUSSION

In this paper, we have analysed seismic ambient noise data recorded at the Chémery UGS site and detected the presence of spectral anomalies on the vertical component (amplification) above the reservoir zone. Similar phenomena have been described at different hydrocarbon reservoirs (both oil and gas) around the world (see Dangel *et al.* 2003, and other works mentioned in the Introduction). Since the spectral anomalies on all the studied sites show remarkable similarities in terms of spectral characteristics and frequency range, the mechanisms behind these spectral anomalies could be common to all those sites. A generic mechanism could be either related to the fluid reservoir itself or a structural effect due to some typical geological feature of hydrocarbon reservoirs. Our work brings extra light to the debate between those who interpreted the spectral anomalies as direct hydrocarbon presence indicators and those who were tending to reject the use of ambient noise spectral amplitude as a potential exploration tool (see Introduction for an overview). The use of a dense array of seismometers allowed us to establish a correlation between the elevation of the anticline structure hosting the reservoir and the spectral anomalies. On the other hand, a cross-analysis of PSD and beamforming time-series showed that the spectral anomalies are the most pronounced for wavefields dominated by R0. Numerical simulations of Rayleigh waves interaction with the anticline structure confirmed that this mechanism was likely to explain the spectral anomalies observed in Chémery.

An important aspect to discuss is the choice of the reference station/zone and of the frequency band of interest, both of which are at the basis of the whole analysis procedure. In practice, this choice results from a preliminary set of trial and error tests. The main requirement is to have a reference spectrum which remains stable relative to the distribution of the spectra recorded by the full network, whatever the absolute noise level, within the frequency band of interest, as shown in Fig. 4. As we measure spectral anomalies on a logarithmic scale, the absolute location of the reference spectrum within the full spectra distribution does not have much importance, as it would only shift the colour scale of the spectral anomaly maps. However, it seems logical to choose a reference station/zone far enough from the imaging target, so that the target effects can be quantified with respect to this reference. It would have been preferable to install one or several reference stations at fix locations, which has not been done in Chémery, forcing us to introduce a reference zone. Instead, we could have used the mean or the median spectrum of the full network as a reference, since the network covers the same area and has a uniform station density at each of the rectangular recording stages. However, for field experiments not fulfilling these two conditions, using a mean/median spectrum would induce bias. Therefore, we believe that using a fixed reference station/zone is a more universal approach.

The example of anticline-induced spectral anomalies observed in Chémery does not pretend to give a general interpretation of all the spectral amplification phenomenology observed at different hydrocarbon sites. First, in order to get an anticline-induced spectral anomaly at frequencies of several Hz in a Rayleigh-wave-dominated wavefield, the anticline must affect relatively shallow layers (i.e. at a depth of several hundred metres). This situation is far from being common to all the sites described in the literature. For example, at the Voitsdorf field described in Lambert *et al.* (2009), oil and gas (at a depth between 2 and 2.5 km) are rather trapped on a fault, while the shallow layers do not seem to be affected by the same fault system (see their fig. 1b). This is obviously a very different geological setting, and we cannot transpose our conclusions to this case. Yet spectral amplification anomalies were reported on this site (Lambert *et al.* 2009). Second, our example shows a Rayleigh-wave-driven spectral anomaly, while several previous works considered that it was rather body-wave driven (Lambert *et al.* 2009; Saenger *et al.* 2009; Goertz *et al.* 2012). In our opinion, the wavefield composition should be systematically analysed with beamforming techniques in order to have a clear understanding of the temporal correlation between the body/surface wave balance and the spectral anomalies, such as we propose in Fig. 8 and Section 3.3. This imposes some constraints on the survey geometry: the ARF should be suitable for working in the frequency band of the observed spectral anomalies. If the latter occur at relatively high frequencies (>3 Hz e.g.), the deployment of small ad-hoc arrays might be helpful.

It has sometimes been debated whether the spectral anomalies could be due to local human sources at the industrial facilities associated to oil/gas production at the site, as suggested, among others, by Hanssen & Bussat (2008). The frequency band above above 1 Hz is indeed the one where human activities are visible. Moreover, our

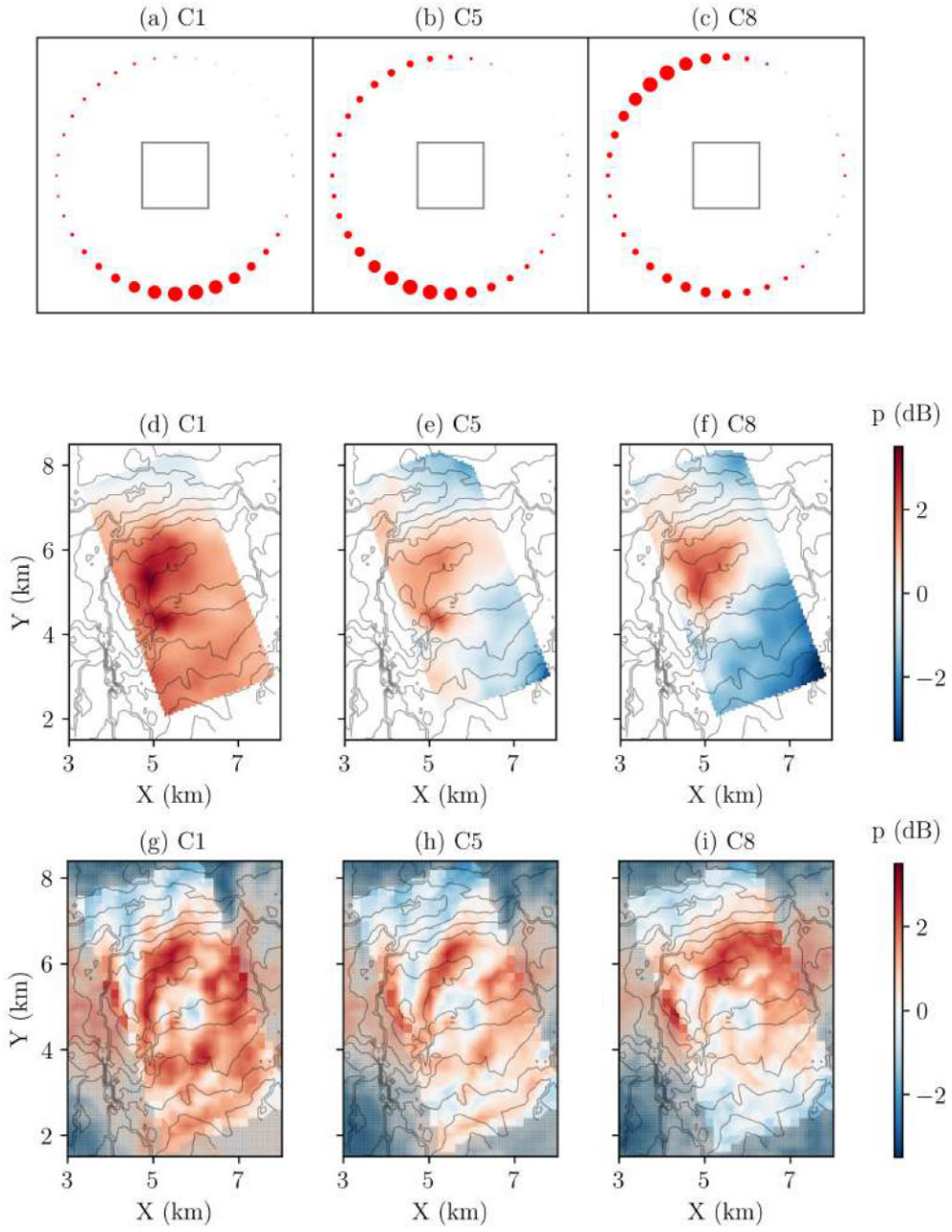

Figure 12. (a)–(c) R0 source distributions for clusters C1, C5 and C8 (see clustering analysis performed on real data in Section 3.3). (d)–(f) Real data averaged spectral anomaly maps over time windows belonging to the same clusters. (g)–(i) Spectral anomaly maps p^* derived from numerical simulations on the Chémery site using source distributions of clusters C1, C5 and C8.

analysis shows that the hours of high human activity correlate with high levels of spectral anomaly. However, our conclusion is that there is nearly no link between the industrial site and the observed amplification in Chémery. This conclusion comes from the clustering analysis (Fig. 10). For example, the cluster C8 corresponds to time windows between noon and 3 pm, and yields a moderate amplification amplitude, while the cluster C1 corresponds to night hours, with a higher amplification amplitude. As the human activity stops during night hours at the site, such a situation would not be possible if the observed amplification was due to the site self-generated noise.

We remind that the objective of the study was *not* to find the frequency band where the gas reservoir is most likely to influence surface wave propagation. The objective was rather to find a mechanism which would be able to explain an amplification phenomenon observed at frequencies >1 Hz. We have demonstrated that the amplification was stronger when fundamental Rayleigh waves were dominating the wavefield. This implicitly rules out a reservoir effect in this frequency band. For completeness, numerical simulations presented in Section 4.3 were performed to explicitly study the reservoir effect. They confirmed that a 1 km deep reservoir placed into a surface-wave- dominated wavefield affects lower frequencies (around 0.75 Hz, see Fig. 13)

Our work shows that the spectral anomalies can probably not be used as an UGS monitoring tool. This is a direct consequence of interpreting them as a static structural effect. Even if this

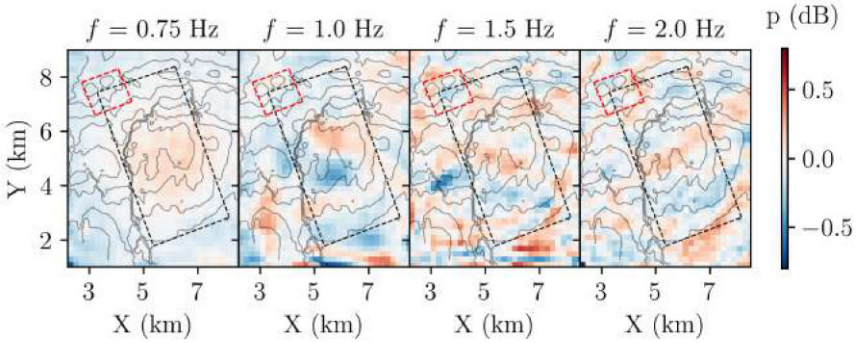

Figure 13. Reservoir induced spectral anomaly maps (differences between spectral anomaly maps with/without the presence of the reservoir) at different frequencies obtained from numerical simulations on the Chémery site using R0 sources uniformly distributed along the backazimuths.

interpretation were to be erroneous for some reason, our analysis shows that the spectral anomalies are stronger in April, at minimum gas storage level, as compared to November, at maximum storage level. From a purely phenomenological point of view, this points out how sensitive the spectral amplification is with respect to the seasonal variability of the ambient wavefield. On the other hand, our analysis shows that ambient noise amplification maps *do* carry physical information, in our case they underline a structural anticline deformation. We believe that this is a quite valuable result in the exploration context. One advantage, in comparison with standard ambient noise imaging techniques such as surface wave tomography (see Mordret *et al.* 2013, for one of the first application examples), is that the 'raw' PSD computed at each station is a very low-level result which is close to the data and carries almost no uncertainties related to data processing and quality control thresholds. If spectral amplification maps are correlated with tomography results, it brings extra confidence on the significance of the retrieved velocity models. If one is able to extract Rayleigh-wave-dominated windows from the data, the most efficient use of the spectral anomalies might be a joint traveltime/amplitude tomography, such as described by Lin *et al.* (2012b) with using teleseismic arrivals at a continental scale.

Going back to more technical aspects of our interpretation (anticline-induced spectral anomaly), it is important to stress that the numerical simulations using R0 sources predict a negative spectral anomaly at the anticline below 1 Hz, as it can be seen in Fig. 11, and as it has been discussed by El Khoury *et al.* (2023). This attenuation pattern is not observed in the data. Our understanding of this apparent contradiction is that the overtones, rather than the fundamental mode, dominate the wavefield between 0.5 and 1 Hz (Fig. 7). Another point is that the few days of recording in 2010 April and November might be statistically insufficient to draw conclusions about the relation between the spectral anomaly variation and the source variation over time, which is the key point of our analysis.

To compensate for this lack of recording time, a similar ambient noise analysis was performed at another UGS site located in Saint-Illiers, France. The Saint-Illiers survey was performed over a long continuous period of time (2.5 yr), allowing us to study the long-term variation of the spectral anomalies and of the wavefield composition. However, in contrast to the Chémery survey, only a limited number of seismometers (Trillium 20 s broad bands) were used (maximum 12, and usually 8–9 working correctly at the same time, Fig. 14a). While we only present the key results of this data set analysis here, as summarized in Fig. 14, more details can be found in the PhD work by El Khoury (2022). As in Chémery, we observe a clear amplification in the PSD_Z for stations SI37 and SI09 (not available in 2019 November) located in the vicinity of the reservoir zone, compared to stations outside the reservoir zone, at frequencies ranging from 1.3 to 4.3 Hz (see spectra in Figs 14c and d). Long-time continuous evolution of the amplification (integrated between 1.3 and 4.3 Hz) at the receiver SI37 relative to the receiver SI03 is shown in Fig. 14(e). The figure separates 'weekdays-day' and 'weekends-night' hours, used as proxies for respectively the maximum and the minimum human activity. The two red circles indicate the two dates for which the spectra are shown in Figs 14(c) and (d), and for which the beamforming results are shown in Figs 14(f) and (g). While the amplification occurs for nearly all the time windows, we see that the spectral anomalies exhibit clear seasonal, weekly and daily variations. The spectral anomalies are the lowest during the winter, weekend and night hours, while they are the highest during the summer, weekdays and day hours (Figs 14c–e). On the other hand, our analysis of the dispersion curves obtained through beamforming revealed that the summer, weekdays and day hours are those when the wavefield above 1 Hz is dominated by R0 sources (Figs 14f and g). The study of the source backazimuths described in El Khoury (2022, not shown here) showed that R0 sources are mainly due to human sources located in the Paris region, about 60 km away from the site, while *P* waves are mainly due to natural sources excited in the North Atlantic ocean. These observations support the hypothesis that the observed spectral anomalies are likely to be associated with R0 waves excited by human activities, and interacting with an anticline structure, confirming the Chémery analysis conclusions. Given that natural noise sources in the Atlantic are typically stronger in winter (Stutzmann *et al.* 2012), the lower values of the spectral anomalies in the late Autumn, Winter and early Spring are likely due to body waves from distant natural sources masking a part of the R0 energy, and thus masking the anticline-induced amplification. It is important to insist that this seasonality might appear correlated to the gas level in the reservoir, which is usually maximum around November and minimum around April. Our analysis suggests, however, that the spectral amplification time variations are due to the source wavefield and not the gas level in the reservoir.

Finally, going back to the absence of the attenuation anomaly below 1 Hz, predicted by the numerical simulations, we see in Figs 14(f) and (g) that the wavefield in Saint-Illiers is also dominated by the overtones between 0.5 and 1 Hz (despite a lower resolution as compared to Fig. 7, due to a less dense array of seismometers). This might be a feature of the ambient noise propagation through the Paris Basin where both Chémery and Saint-Illiers sites are located. In the future, it might be interesting to study the case of some other anticline structure placed into a wavefield fully dominated by R0

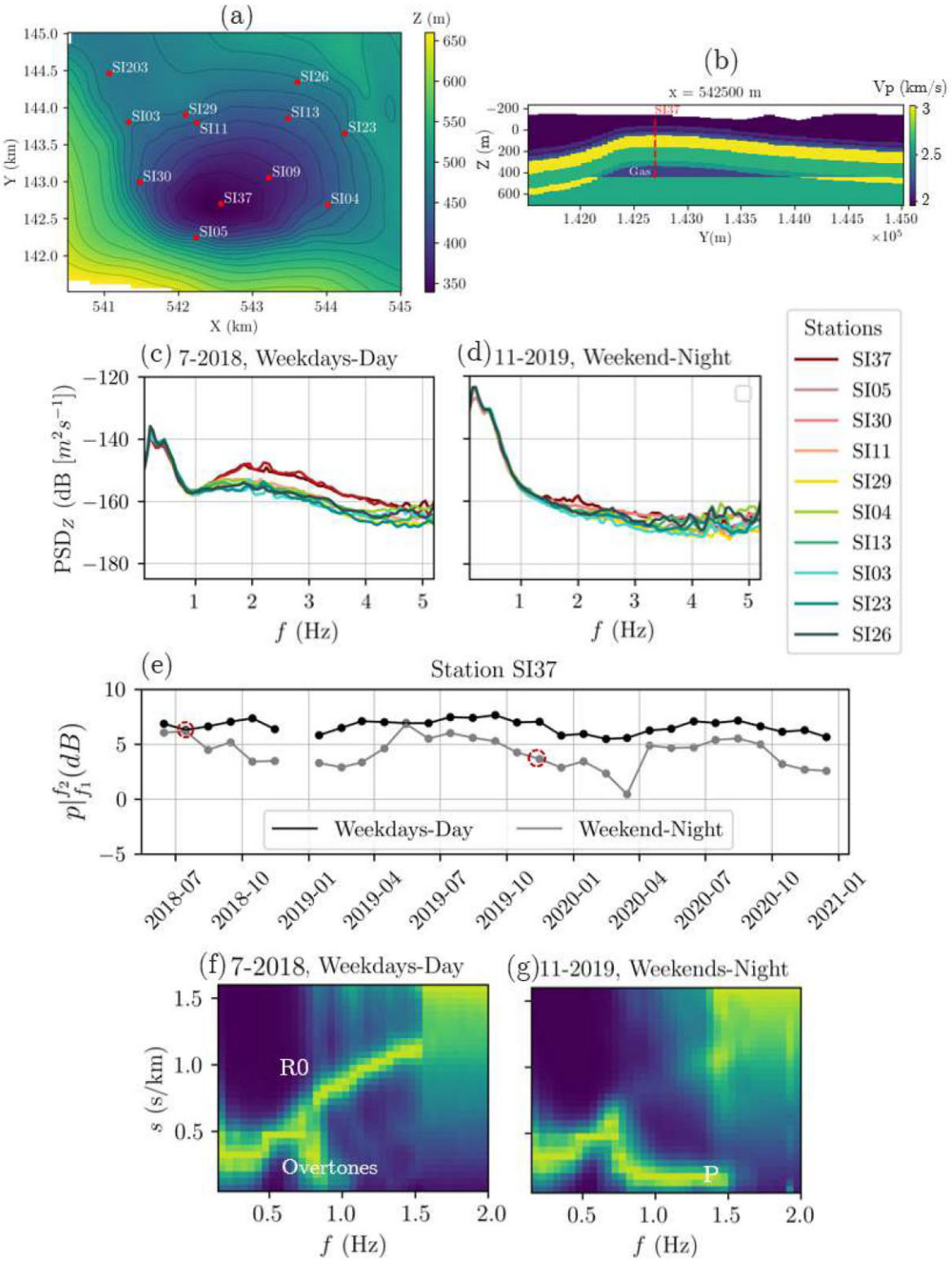

Figure 14. (a) Reservoir top (depth below sea level) and station locations for the Saint-Illiers site. (b) N-S cross-section of the V_P velocity passing by the station SI37. (c) and (d) PSD$_Z$ for different stations averaged over two different time periods (weekdays-daytime in 2018 July and weekends-nighttime in 2019 November). (e) Relative spectral anomaly (average between 1.3 and 4.3 Hz) variation over 2.5 yr at the station SI37 with respect to the station SI03. Each dot represents the average over one month. The two curves correspond to 'weekdays-daytime' and 'weekends-nighttime' data selections. (f) and (g) Dispersion diagrams averaged over the two time periods shown respectively in (c) and (d), and spotted with red circles in (e).

from the microseism frequency (around 0.1 Hz) up to several Hz, and to check if the attenuation anomaly prediction is verified.

6 CONCLUSION

We have analysed the ambient noise data recorded for several days in 2010 April and November by a dense surface array of seismometers at the Chémery UGS site in the Paris Basin, France. We observed spectral anomalies (amplification) up to 4 dB peaking between 1.6 and 1.8 Hz above the reservoir zone, also associated with an anticline trap affecting multiple geological layers. We extracted both spectral amplification and beamforming results for time windows of 30 min. The analysis of these time-series lead us to the conclusion that variations of the spectral anomaly over time are correlated with the variations of the ambient noise wavefield composition, the strongest amplification corresponding to the time windows dominated by R0. 3-D numerical simulations in a realistic 3-D velocity model with source distributions extracted from the data suggest that the observed spectral anomalies are consistent with the R0 interaction with the anticline structure in Chémery. The analysis of a second data set recorded at the Saint-Illiers UGS site over 2.5 yr confirms the phenomena observed in Chémery, and supports the idea that the systematic spectral anomaly decrease in winter is associated with stronger incident body waves of natural origin masking the more local human-generated fundamental-mode Rayleigh waves. Without claiming any general explanation to the world-wide observations of ambient noise spectral amplification above hydrocarbon reservoirs, our study suggests a possible interpretation for fluid reservoirs associated with a prominent anticline trap also affecting shallow layers. Furthermore, it provides guidelines for data analysis of local ambient noise amplification, which must rely on a thorough study of the ambient wavefield composition and evolution over time. While we discard the use of spectral amplification for reservoir monitoring, we believe that it can be a powerful supplement to surface wave imaging techniques used for passive exploration, either as a qualitative validation of the location of structural heterogeneities, or as an extra observable for a joint inverse problem.

ACKNOWLEDGMENTS

The authors thank Storengy for funding the project and granting access to the data and their presentation.

DATA AVAILABILITY

Due to confidentiality agreements, supporting data cannot be made openly available.

REFERENCES

Ali, M., Berteussen, K., Small, J. & Barkat, B., 2010. Low-frequency passive seismic experiments in Abu Dhabi, United Arab Emirates: implications for hydrocarbon detection, *Geophys. Prospect.,* **58,** 875–899.

Antunes, V. *et al.*, 2022. Insights into the dynamics of the Nirano Mud Volcano through seismic characterization of drumbeat signals and V/H analysis, *J. Volc. Geotherm. Res.,* **431,** 107619. doi: 10.1016/j.jvolgeores.2022.107619.

Bard, P.-Y., Anastasiadis, A., Atakan, K. & Azzara, R. M., 2008. SESAME-HV-User-Guidelines, *Bull. Earthq. Eng.,* **63,** 1–62.

Bertero, M. & Boccacci, P., 2005. A simple method for the reduction of boundary effects in the richardson-lucy approach to image deconvolution, *Astron. Astrophys.,* **437,** 369–374.

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**(3–4), 205–227.

Bormann, P. & Wielandt, E., 2013. *Seismic Signals and Noise,* pp. 1–62, Deutsches GeoForschungsZentrum, GFZ.

Capon, J., 1969. High-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE,* **57**(8), 1408–1418.

Cornou, C., 2002. *Traitement d'antenne et imagerie sismique dans l'agglomération grenobloise (Alpes françaises) : implications pour les effets de site,* PhD thesis, Laboratoire de Géophysique Interne et Tectonophysique.

Dangel, S., Schaepman, M.E., Stoll, E.P., Carniel, R., Barzandji, O., Rode, E.D. & Singer, J.M., 2003. Phenomenology of tremor-like signals observed over hydrocarbon reservoirs, *J. Volc. Geotherm. Res.,* **128**(1), 135–158.

De Vasconcelos Lopes, A.E. & Nunes, L.C., 2010. Pitfalls of tremor-like signals for hydrocarbon exploration in producing oil fields in Potiguar Basin, northeast Brazil, *Leading Edge,* **29**(7), 826–830.

El Khoury, C., 2022. *Anomalies d'amplitude du bruit sismique ambiant induites par une structure anticlinale : application au contexte de réservoirs de gaz,* PhD thesis, Mines Paris - Université Paris sciences et lettres.

El Khoury, C., Chauris, H., Kazantsev, A. & Monteiller, V., 2023. Rayleigh wave spectral distortions induced by an anticline structure, *Geophys. J. Int.,* **233**(3), 2067–2083.

Frehner, M., Schmalholz, S.M. & Podladchikov, Y., 2009. Spectral modification of seismic waves propagating through solids exhibiting a resonance frequency: a 1-D coupled wave propagation-oscillation model, *Geophys. J. Int.,* **176**(2), 589–600.

Gal, M., Reading, A., Ellingsen, S., Koper, K., Burlacu, R. & Gibbons, S., 2016. Deconvolution enhanced direction of arrival estimation using 1- and 3-component seismic arrays applied to ocean induced microseisms, *Geophys. J. Int.,* **206**(1), 345–359.

Goertz, A., Schechinger, B., Witten, B., Koerbe, M. & Krajewski, P., 2012. Extracting subsurface information from ambient seismic nose - A case study from Germany, *Geophysics,* **77,** 13. doi: 10.1190/geo2011-0306.1.

Green, A.G. & Greenhalgh, S., 2010. Comment on 'Low-frequency microtremor anomalies at an oil and gas field in Voitsdorf, Austria' by Marc-André Lambert, Stefan Schmalholz, Erik H. Saenger and Brian Steiner, *Geophys. Prospect.,* **57,** 393–411 *[Geophys. Prospect.*, **58**(2), 335–339].

Guillier, B., Chatelain, J.-L., Bonnefoy-Claudet, S. & Haghshenas, E., 2007. Use of ambient noise: from spectral amplitude variability to H/V stability, *J. Earthq. Eng.,* **11,** 925–942.

Hanssen, P. & Bussat, S., 2008. Pitfalls in the analysis of low frequency passive seismic data, *First Break,* **26**(6). doi: 10.3997/1365-2397.26.1288.28412.

Hartigan, J.A. & Wong, M.A., 1979. Algorithm AS 136: a K-means clustering algorithm, *J. R. Stat. Soc. Ser. C (Appl. Stat.),* **28**(1), 100–108.

Holzner, R., Eschle, P., Dangel, S., Frehner, M., Narayanan, C. & Lakehal, D., 2009. Hydrocarbon microtremors interpreted as nonlinear oscillations driven by oceanic background waves, *Commun. Nonlinear Sci. Numer. Simul.,* **14,** 160–173.

Kaoungku, N., Suksut, K., Chanklan, R., Kerdprasop, K. & Kerdprasop, N., 2018. The silhouette width criterion for clustering and association mining to select image features, *Int. J. Mach. Learn. Comput.,* **8**(1), 69–73.

Kassambara, A., 2017. *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning,* CreateSpace Independent Publishing Platform, S.l.

Kazantsev, A., 2018. *Ambient noise spectral amplitude distortions above heterogeneities: feasability study for multi-fluid reservoir exploration and monitoring,* PhD thesis, MINES ParisTech - Université de recherche Paris Sciences et Lettres PSL Research University, Paris, France.

Kelly, M.C., Riahi, N., Ruiz, M. & Weiwei, 2013. *Bayesian DHI for seismic data*. https://patents.google.com/patent/US8358561B2/en.

Komatitsch, D. & Vilotte, J.-P., 1998. The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures, *Bull. seism. Soc. Am.,* **88**(2), 368–392.

Komatitsch, D., Vilotte, J.-P., Vai, R., Castillo-Covarrubias, J.M. & Sánchez-Sesma, F.J., 1999. The spectral element method for elastic wave equations—application to 2-D and 3-D seismic problems, *Int. J. Numer. Methods Eng.,* **45**(9), 1139–1164.

Lambert, M.-A., 2010. *Characterization of ambient ground-motion using spectral analysis techniques,* PhD thesis, Dissertation, Eidgenössische Technische Hochschule ETH Zürich . Nr. 19047, 2010.

Lambert, M.-A., Schmalholz, S.M., Saenger, E.H. & Steiner, B., 2009. Low-frequency microtremor anomalies at an oil and gas field in Voitsdorf, Austria, *Geophys. Prospect.,* **57**(3), 393–411.

Lin, F.-C., Schmandt, B. & Tsai, V.C., 2012a. Joint inversion of Rayleigh wave phase velocity and ellipticity using USArray: Constraining velocity and density structure in the upper crust: Rayleigh wave ellipticity across usarray, *Geophys. Res. Lett.,* **39**(12), doi: 10.1029/2012GL052196.

Lin, F.-C., Tsai, V.C. & Ritzwoller, M.H., 2012b. The local amplification of surface waves: a new observable to constrain elastic velocities, density, and anelastic attenuation, *J. Geophys. Res. Solid Earth.,* **117**(B6), doi: 10.1029/2012JB009208.

Lucy, L.B., 1974. An iterative technique for the rectification of observed distributions, *Astron. J.,* **79,** 745. doi: 10.1086/111605.

Lunedei, E. & Malischewsky, P., 2015. A review and some new issues on the theory of the H/V technique for ambient vibrations, in Ansal, A., ed., *Perspectives on European Earthquake Engineering and Seismology. Geotechnical, Geological and Earthquake Engineering, Vol. 39, ch. 15 of Geotechnical, Geological and Earthquake Engineering,* Springer International Publishing, Cham. 371–394.

MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations, in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1: Statistics,* **5.1,** pp. 281–298. University of California Press, Berkeley.

Martini, F. *et al.*, 2013. A passive low-frequency seismic experiment in the albertine graben, Uganda, *Geophys. Prospect.,* **61,** 39–61.

Mordret, A., Landes, M., Shapiro, N.M., Singh, S.C., Roux, P. & Barkved, O.I., 2013. Near-surface study at the Valhall oil field from ambient noise surface wave tomography, *Geophys. J. Int.,* **193**(3), 1627–1643.

Nishida, K., 2017. Ambient seismic wave field, *Proc. Japan Acad. Ser. B,* **93**(7), 423–448.

Parolai, S., 2012. Investigation of site response in urban areas by using earthquake data and seismic noise, in *New Manual of Seismological Observatory Practice 2 (NMSOP2),* Potsdam : Deutsches GeoForschungsZentrum GFZ, pp. 1–38. https://doi.org/10.2312/GFZ.NMSOP-2_ch14.

Peruzzetto, M., Kazantsev, A., Luu, K., Métaxian, J.-P., Huguet, F. & Chauris, H., 2018. Broad-band ambient noise characterization by joint use of cross-correlation and MUSIC algorithm, *Geophys. J. Int.,* **215**(2), 760–779.

Podladchikov, Y., Lambert, M.-A., Dewarrat, R. & Schmalholz, S., 2010. *VH Reservoir Mapping*. https://patents.google.com/patent/US7676326/en.

Richardson, W.H., 1972. Bayesian-Based iterative method of image restoration, *J. Opt. Soc. Am. (1917-1983),* **62,** 55–59.

Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.,* **40**(3). doi: 10.1029/2000RG000100.

Saenger, E. *et al.*, 2009. A passive seismic survey over a gas field: analysis of low-frequency anomalies, *Geophysics,* **74**(2), O29–O40.

Saenger, E.H., 2008. *Seismic attributes for reservoir localization*. https://patents.google.com/patent/US8219320/en.

Saenger, E.H., 2009. *VH signal integration measure for seismic data*. https://patents.google.com/patent/US7725265B2/en.

Schmidt, R., 1986. Multiple emitter location and signal parameter estimation, *IEEE Trans. Antennas Propag.,* **34**(3), 276–280.

Schmidt, R.O., 1981. *A signal subspace approach to multiple emitter location and spectral estimation,* PhD thesis, Stanford University.

Spica, Z., Perton, M., Nakata, N., Liu, X. & Beroza, G.C., 2018. Shallow VS imaging of the Groningen Area from joint inversion of multimode surface waves and H/V spectral ratios, *Seismol. Res. Lett.,* **89**(5), 1720–1729.

Steiner, B., Saenger, E. & Schmalholz, S., 2008. Time reverse modeling of low-frequency microtremors: application to hydrocarbon reservoir localization, *Geophys. Res. Lett.,* **35**(3), doi: 10.1029/2007GL032097.

Stutzmann, E., Ardhuin, F., Schimmel, M., Mangeney, A. & Patau, G., 2012. Modelling long-term seismic noise in various environments, *Geophys. J. Int.,* **191**(2), 707–722.

Tibshirani, R., Walther, G. & Hastie, T., 2001. Estimating the number of clusters in a data set via the gap statistic, *J. R. Stat. Soc. Ser. B,* **63,** 411–423.

Vidal, S., Longuemare, P., Huguet, F. & Mechler, P., 2002. Reservoir parameters quantification from seismic monitoring integrating geomechanics, *Oil Gas Sci. Technol.,* **57**(5), 555–568.

Weaver, R., Froment, B. & Campillo, M., 2009. On the correlation of non-isotropically distributed ballistic scalar diffuse waves, *J. acoust. Soc. Am.,* **126,** 1817–26.