

# A methodological approach towards high-resolution surface wave imaging of the San Jacinto Fault Zone using ambient-noise recordings at a spatially dense array

Philippe Roux,<sup>1</sup> Ludovic Moreau,<sup>1</sup> Albanne Lecointre,<sup>1</sup> Gregor Hillers,<sup>1</sup> Michel Campillo,<sup>1</sup> Yehuda Ben-Zion,<sup>2</sup> Dimitri Zigone<sup>2,\*</sup> and Frank Vernon<sup>3</sup>

<sup>1</sup>*Institut des Sciences de la Terre, Université Joseph Fourier, CNRS UMR 5275, Maison des Géosciences, F-38400 Saint Martin d'Hères, France.*

*E-mail: philippe.roux@ujf-grenoble.fr*

<sup>2</sup>*Department of Earth Sciences, University of Southern California, Los Angeles, CA 90089-0740, USA*

<sup>3</sup>*Scripps Institution of Oceanography, UC San Diego, La Jolla, CA 92093-0225, USA*

Accepted 2016 May 17. Received 2016 May 17; in original form 2015 December 22

## SUMMARY

We present a new technique for deriving detailed information on seismic velocities of the subsurface material from continuous ambient noise recorded by spatially dense seismic arrays. This method uses iterative double beamforming between various subarrays to extract surface wave contributions from the ambient-noise data in complex environments with unfavourable noise-source distributions. The iterative double beamforming extraction makes it possible to retrieve large amounts of Rayleigh wave traveltime information in a wide frequency band. The method is applied to data recorded by a highly dense Nodal array with 1108 vertical geophones, centred on the damage zone of the Clark branch of the San Jacinto Fault Zone south of Anza, California. The array covers a region of  $\sim 650 \times 700$  m<sup>2</sup>, with instrument spacing of 10–30 m, and continuous recording at 500 samples s<sup>-1</sup> over 30 d in 2014. Using this iterative double beamforming on subarrays of 25 sensors and cross-correlations between all of the station pairs, we separate surface waves from body waves that are abundant in the raw cross-correlation data. Focusing solely on surface waves, maps of traveltimes are obtained at different frequencies with unprecedented accuracy at each point of a 15-m-spacing grid. Group velocity inversions at 2–4 Hz reveal depth and lateral variations in the structural properties within and around the San Jacinto Fault Zone in the study area. This method can be used over wider frequency ranges and can be combined with other imaging techniques, such as eikonal tomography, to provide unprecedented detailed structural images of the subsurface material.

**Key words:** Spatial analysis; Tomography; Interferometry; Surface waves and free oscillations; Wave propagation; Fractures and faults.

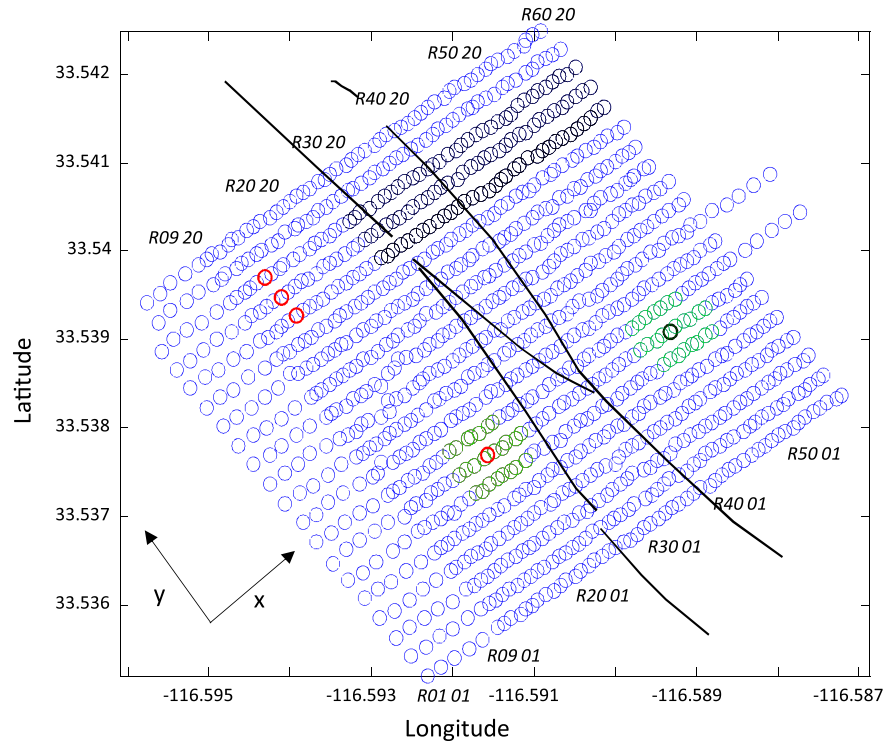
## 1 INTRODUCTION

In science-fiction films, the future is often described as a world of sensors. This is already true for the sciences in general, and for Earth sciences in particular. To analyse the vast amount of recorded data, large computers and fast algorithms are needed. More importantly, it is essential to develop new methodologies for extraction of the increasing range of signals that are recorded by dense arrays of sensors. The need for new methods is particularly true at geophysics scales, which can now benefit from continuous data acquisition using very dense arrays of seismometers. These arrays can sometimes include more than 10 000 sensors. Such sensor deployments were

nearly inconceivable only 10 yr ago, and even very recently they were limited to highly expensive geophysics experiments funded by the oil and gas industries (Slater & Hollis 2012). However, recent technological developments now allow academic institutions to obtain seismic data recorded by thousands of sensors (Lin *et al.* 2013; Ben-Zion *et al.* 2015; Hansen & Schmandt 2015).

This paper addresses the need for new methodologies in seismic imaging of complex environments based on data recorded by dense arrays of seismometers. We focus on methods that provide high spatial resolution of critical areas, such as fault zones, that are of particular importance for Earth sciences (Ben-Zion 2008). We modify and further develop imaging techniques imported from other fields than geophysics, such as medical ultrasonics and underwater acoustics that were designed for scales ranging from a few centimetres for medical applications to kilometres in ocean acoustics. The new array processing methods developed here do

\*Now at: Institut de Physique du Globe de Strasbourg, Université de Strasbourg, EOSt, CNRS, Strasbourg, France.



**Figure 1.** Map of the 1108 (10 Hz) geophones installed in a  $650 \times 700 \text{ m}^2$  configuration above the Clark branch (black lines) of the San Jacinto Fault (Southern California). Each row along the  $x$ -direction is composed of  $\sim 55$  sensors with an interdistance of 10 m, and the nominal separation between the rows in the  $y$ -direction is 30 m. Seismic ambient noise was recorded over more than one month, in May–June 2014. Some of the station names are given (*italics*). The red dot at the centre of a 25-element green subarray corresponds to the sensor positioned at row 23, column 7 (called R2307). The two green subarrays refer to the DBF processing described in Figs 8–11. The subarray centres at the North refer to a set of surrogate ‘sources’ (red) and a line of surrogate ‘receivers’ (black) for DBF results at different frequencies (see Fig. 12).

not rely on complicated models, but instead only on data quality. We aim to reach high resolution, although not at the price of robustness. If *a-priori* information is used at a given analysis stage, it should be obtained from the data and not from prior knowledge of the Earth structure, as this knowledge is limited and uncertain.

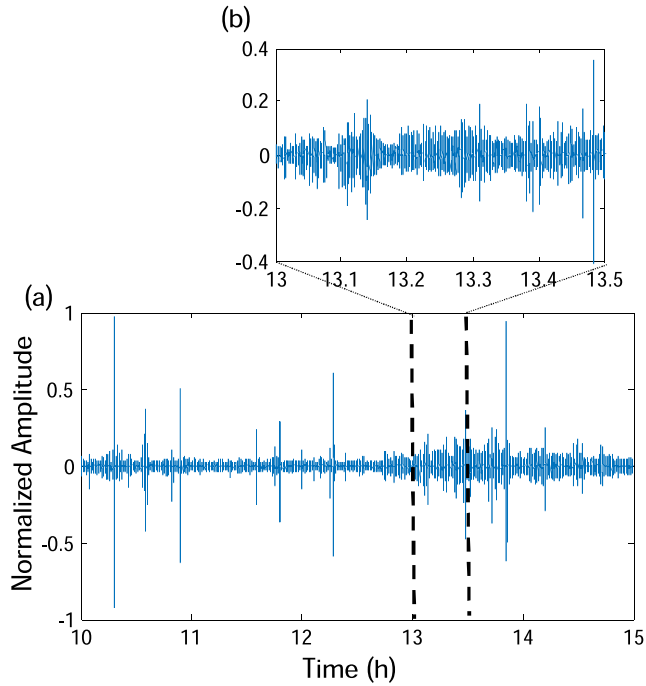
In parallel with the improvement of array-deployment technology and processing, the use of continuous ambient noise recorded on dense geophone arrays allows the creation of surrogate active sources from purely passive multi-element systems (Lin *et al.* 2013). Turning passive systems into active ones through basic signal processing has revolutionized the use of seismic arrays for subsurface imaging at local and regional scales (Sabra *et al.* 2005a,b; Shapiro *et al.* 2005; Roux *et al.* 2011). Taking advantage of the use of ambient seismic noise to produce coherent images, we develop here techniques to extract high-resolution subsurface structural models from spatially dense noise records.

Combining array processing and noise correlation, we analyse data recorded by a spatially dense array of seismometers positioned over the damage zone of the San Jacinto Fault Zone (SJFZ) in Southern California (Ben-Zion *et al.* 2015). The SJFZ is the most seismically active fault zone in Southern California (Hauksson *et al.* 2012), and it accounts for a large portion of the plate motion in the region (Johnson *et al.* 1994; Lindsey & Fialko 2013). Robust seismicity and a highly complex fault-zone structure with prominent lateral and vertical heterogeneities at various scales (Allam & Ben-Zion 2012; Zigone *et al.* 2015), along with fault-zone amplification effects (Kurzon *et al.* 2014), add significant challenges to the goals of this study.

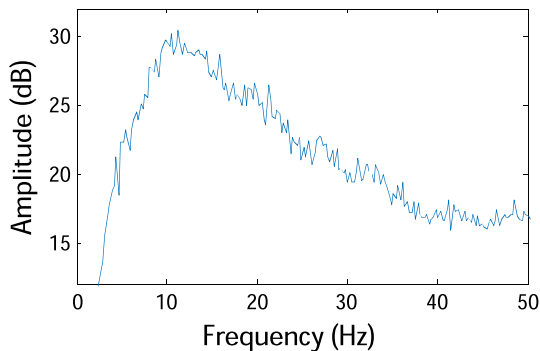
The paper is structured as follows. In Section 2, the properties of the ambient noise recorded by the dense array are analysed to optimize the outcome of the noise-correlation process. In particular, a spatial filtering algorithm is designed on the whole array to diminish the importance of body waves associated with local microseismicity just below the receivers. In Section 3, double beamforming (DBF) is applied to the noise-correlation signals between subarrays taken out of the dense array network. An iterative procedure allows the identification and extraction of surface waves from their slowness and group velocities. Finally, maps of the surface wave traveltimes are projected onto the seismic network, and tomography inversion is performed at discrete frequencies.

## 2 ARRAY GEOMETRY AND NOISE PROPERTIES

The dense seismic array that provides the data used in this study consists of 1108 (10-Hz) vertical geophones that were installed in a  $\sim 650 \times 700 \text{ m}^2$  configuration centred on the Clark branch of the SJFZ southeast of Anza, California. Each row along the  $x$ -direction has  $\sim 55$  sensors, with an inter-sensor distance of 10 m, and a nominal separation between the rows in the  $y$  direction of 30 m (Fig. 1). The array recorded the ambient seismic noise (and earthquakes) continuously over more than one month in May to June, 2014. Inspection of the recorded waveforms reveals numerous spikes and bursts that cover a large frequency band between 5 Hz and 40 Hz (Figs 2 and 3). In addition to distinguishable seismic events, the waveforms contain large sets of long bursts (Fig. 2b)



**Figure 2.** (a) Normalized ambient seismic data recorded over 5 hr on sensor R2307, and filtered below 50 Hz. The seismic signal shows both a continuous series of impulsive spikes that correspond to small local earthquakes and time-dispersed bursts that resemble creeping noise induced by the fault, as expanded in panel (b).



**Figure 3.** Averaged power-spectral density obtained for sensor R2307 for one day of recording (Julian Day 150).

that are typically grouped with the ambient seismic noise extending down to 1 Hz. Despite the incoherent noise generated by the ocean and by nearby sources (e.g. wind or human related), recent analyses have suggested that most of the detectable impulsive events are located below the surface (Ben-Zion *et al.* 2015).

As usual when dealing with ambient-noise correlations, the time and frequency pre-processing was extensively experimented with prior to the cross-correlation stage. The presence in the raw data of both impulsive spikes and time-dispersed bursts made the choice of the most appropriate pre-processing more complicated. We finally used one-bit time-domain normalization along with frequency whitening in limited frequency bands  $[F - \Delta F, F + \Delta F]$  with  $\Delta F < F/3$  to prevent the creation of signal harmonics caused by the one-bit clipping in the frequency band of interest. Given the numerous microseismic events in the noise recordings, the one-day

averaged cross-correlations show a strong contribution around time  $t = 0$  (Fig. 4a), which suggests that most of the coherent signals were incoming from below the array and with very high apparent velocities. Note that at frequencies between 3 and 5 Hz, the spatial coherency extends over the whole array (more than 600 m) when averaged over one day.

The main goal of this paper is to develop methodological tools for performing surface wave inversion to retrieve subsurface structures of fault zones despite the *a priori* unfavourable noise correlation pattern. Consequently, some post-processing must be performed to emphasize the surface wave contribution in the raw correlations. To this end, we use the spatial discrete Fourier transform (from the  $X$ – $Y$  position domain into the  $K_x$ – $K_y$  slowness domain) to mask the strong contributions of the phases with large apparent velocities (Fig. 4c).

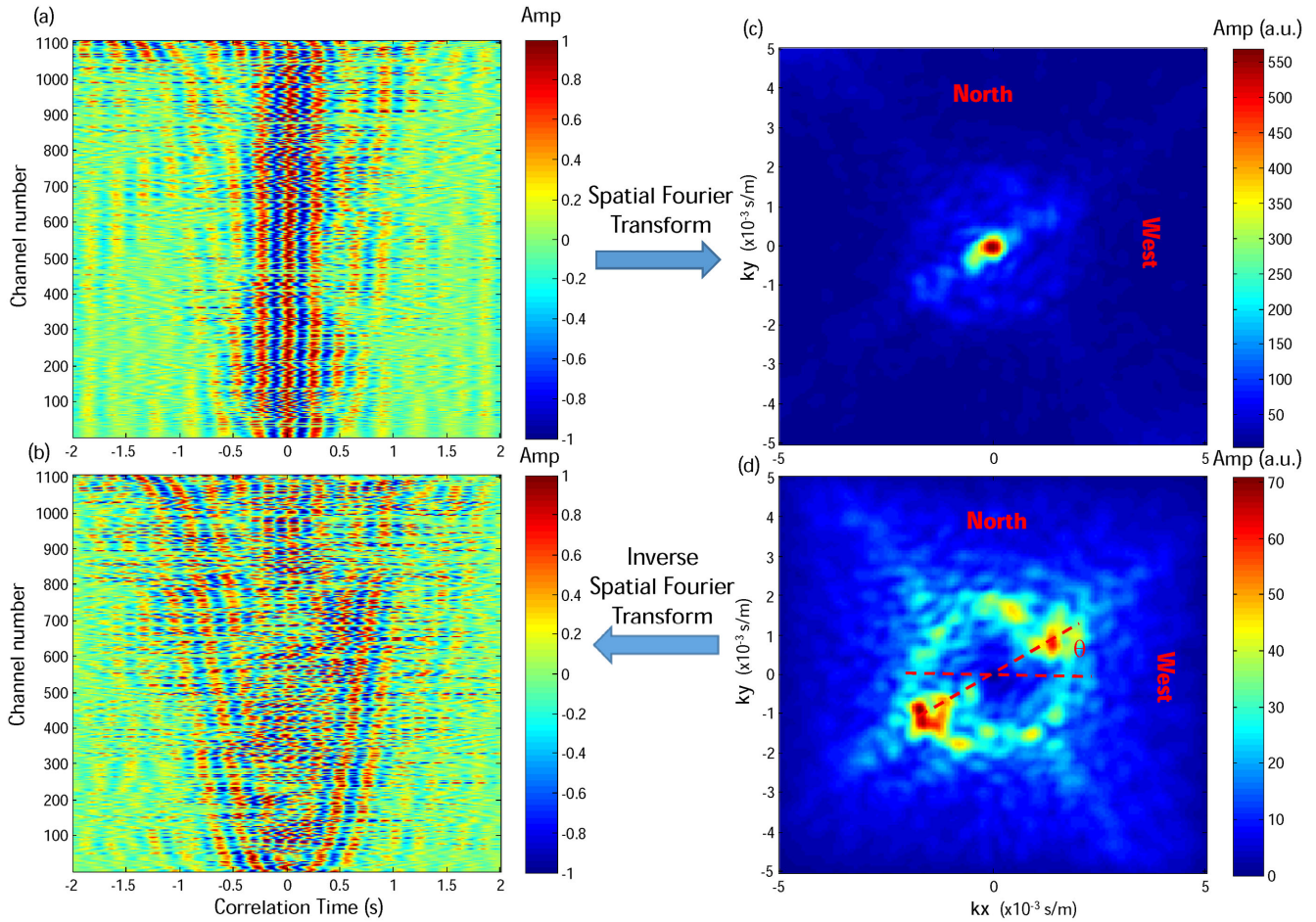
After cancelling the wavefield associated with phase velocities greater than  $1000 \text{ m s}^{-1}$  at each frequency, we see the emergence of a circle in the  $K_x$ – $K_y$  representation that is associated with surface waves incoming from all directions (Fig. 4d). However, this wavenumber filtering is imperfect due to strong local heterogeneities present at and around the fault. Nevertheless, this spatial filtering reduces (more than it cancels) the amplitude of high-apparent-velocity waves that dominate the surface wave amplitudes in the raw correlations (Fig. 4a). In practice, for example, the wavenumber filtering will not be efficient for incident low-velocity surface waves that are locally scattered into higher-velocity body waves due the medium heterogeneity. This bias is observed in the filtered  $K_x$ – $K_y$  representation (Fig. 4d) where the expected circle corresponding to the surface wave contribution is still polluted by a random speckle pattern that results from the surface wave interaction with local heterogeneities. Note also that some incident directions show higher intensities than others, as confirmed later by the beamforming results.

After the high-velocity mask is applied, we perform an inverse spatial Fourier transform to return into the  $X$ – $Y$  domain. This wavenumber filtering operation leads to surface waves that are now clearly visible in the cross-correlation pattern (Fig. 4b). To correct all of the cross-correlation pairs among the network, such spatial filtering must be applied to the 1108 cross-correlation patterns with every reference station. This involves a series of  $\sim 610\,000$  correlation pairs, from which the high apparent-velocity contributions have been removed. These spatially filtered correlations can be stacked in 4-m-distance bins to obtain a first-order spatial representation of the distance-versus-time correlated wavefield (Fig. 5). Such averaging into incremental bins independent of the position of the station pairs among the network amounts to considering a spatially homogeneous medium, which is clearly not consistent with existing knowledge of the fault-zone structure (Sharp 1967; Allam *et al.* 2014; Yang *et al.* 2014; Zigone *et al.* 2015). However, this approach allows us to confirm the effectiveness of the wavenumber filtering for the whole set of cross-correlation pairs, and to extract an averaged phase velocity of  $600 \text{ m s}^{-1}$  for Rayleigh waves at 4 Hz in the medium (Fig. 5b).

To investigate the additional general characteristics of the surface wavefield, beamforming restricted to velocities lower than  $1000 \text{ m s}^{-1}$  was performed on the ambient-noise data at different frequencies. To do so, a disk of 600 stations was defined at the centre of the array, and the beamforming pattern was averaged for 1-hr-long records over 24 hr (Fig. 6).

Fig. 6 shows a polar-plot representation of the plane-wave beamformer calculated for a surface wave model at the maximum velocity output. As expected, the noise incident directions match the





**Figure 4.** Representation of the correlation data in the position space and the wavenumber space before and after the spatial filter processing. (a) Normalized one-day cross-correlations of the 1108 stations, with reference station R2307 (see Fig. 1). The raw ambient-noise signals are frequency whitened in the [3–5 Hz] band before correlation. One-bit amplitude-related pre-processing was also applied to the raw data. The stations are sorted by increasing distance with respect to the reference station R2307. Strong spatial coherence spreads over the whole network with coherent arrivals corresponding to high effective velocities. (b) As for panel (a) after frequency–wavenumber filtering is applied to the raw correlations to cancel the low wavenumber (velocities  $> 1000 \text{ m s}^{-1}$ ) arrivals around  $t = 0$ . The surface wave contribution becomes visible, both on the positive and negative times of the correlations on the whole network in the [3–5 Hz] frequency interval. (c, d) Kx–Ky representations of the time-domain correlations shown in panels (a) and (b) after a 2-D spatial Fourier transform is applied. For the sake of the representation, the intensity patterns shown in panels (a) and (b) are averaged over nine successive frequencies between 3 and 5 Hz, with separation 0.25 Hz. To filter out the energy contribution that corresponds to low wavenumbers, a Gaussian-shaped mask is applied for velocities  $> 800 \text{ m s}^{-1}$  to the 2-D spatial Fourier transform at each frequency. As expected from panel (b), the frequency-averaged intensity pattern in panel (d) roughly reveals a circle, which corresponds to the surface waves, with some notable azimuthal dependence in intensity. Note that the low wavenumber filtering operation divides the maximum intensity in the Kx–Ky domain by a factor of eight. The angle  $\theta$  corresponds to the main directions obtained after beamforming of the surface waves (see Fig. 7).

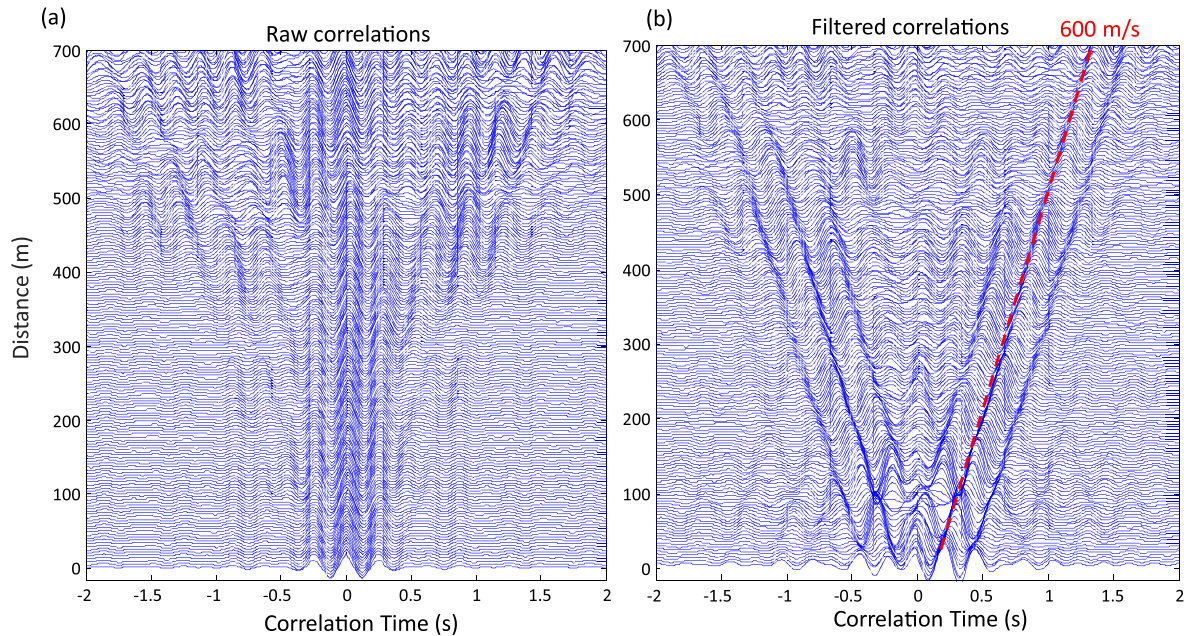
observations made in Fig. 4(d). The  $\sim 60^\circ$  azimuth direction with respect to North is typical of ambient noise in Southern California (Hillers *et al.* 2013), and consistent with previous studies on seismic noise at lower frequencies ( $< 1 \text{ Hz}$ ) that traditionally indicate the micro-seismic noise coming from the Pacific Ocean (Schulte-Pelkum *et al.* 2004; Sabra *et al.* 2005b; Gerstoft & Tanimoto 2007; Roux 2009; Hillers *et al.* 2013). On the other hand, we would expect much more angular dispersion in the back-scattered energy at these frequencies if the opposite direction emerged due to scattering or reflections inside the fault zone. Our interpretation is that, at these low frequencies, the backscattered signal is not due to the SJF itself but more likely to local topography (mostly parallel to the fault) or to other regional fault structures East of the SJFZ (as can be seen from Fig. 2 in Ben-Zion *et al.* 2015).

### 3 ITERATIVE WAVELET EXTRACTION USING DOUBLE BEAMFORMING

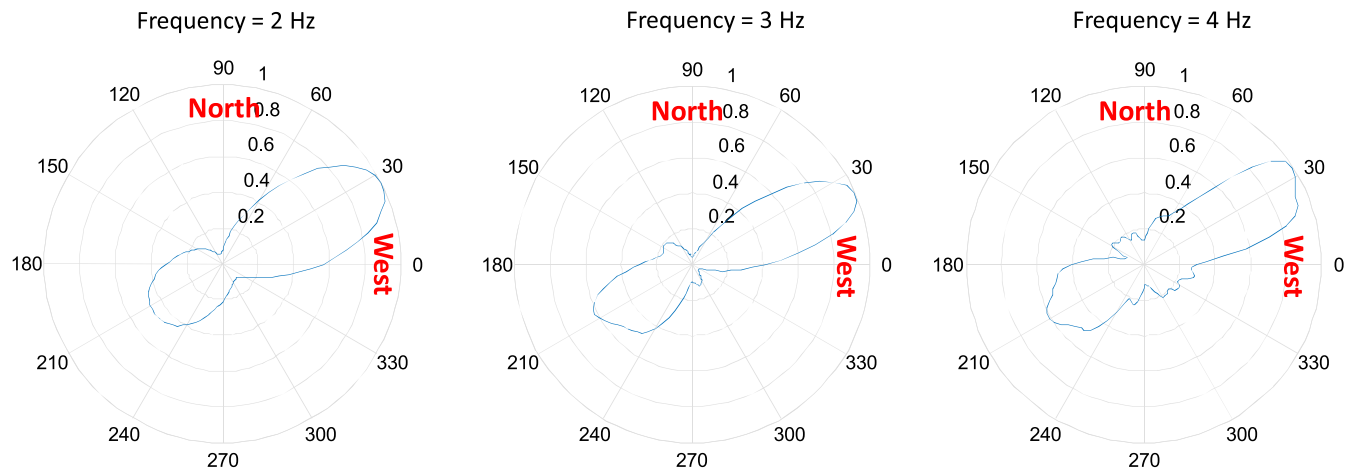
Given the beamforming results, it appears natural to focus on surface wave extraction within receiver pairs aligned with the noise incident directions. Benefitting from the dense array deployment, the strategy here is to combine two subarrays of receivers between which all of the cross-correlation pairs are computed to identify various wavelets of interest and to further constrain the wavelet extraction (Fig. 1, green circles).

The usefulness of DBF to identify and isolate different wave arrivals has been successfully demonstrated in seismology (Rost & Thomas 2002). When using source and receiver arrays, DBF results from underwater acoustics lead to high-resolution tomography from the extraction of numerous multiply reverberated eigenrays





**Figure 5.** Spatial-temporal representation of the averaged cross-correlations before and after spatial filtering. Fig. 4(a) shows the raw correlations with station R2307 taken as a reference. When the reference is sequentially taken as each one of the 1108 stations, the total number of cross-correlation pairs exceeds 610 000. Assuming in the first-order approximation spatially uniform elastic properties of the ground, these correlations can be averaged into 4-m incremental bins that range from 0 to 700 m. The bin-averaged raw correlations are shown in panel (a). As expected from Fig. 4(a), the dominant part is due to high-velocity waves around  $t = 0$ , which might be due to body waves coming from the Fault at a depth below the array. (b) Built following the same binning procedure as panel (a), but starting from the full set of low wavenumber filtered correlations as the single one generated in Fig. 4(b) using reference station R2307.



**Figure 6.** Polar-plot representation of the plane-wave beamforming output performed for Julian Day 150 and for maximum beamformer output at each frequency:  $800 \text{ m s}^{-1}$  at 2 Hz,  $650 \text{ m s}^{-1}$  at 3 Hz and  $590 \text{ m s}^{-1}$  at 4 Hz. The beamforming was computed from 600 neighbouring stations located at the centre of the network. The  $\sim 60^\circ$  azimuth with respect to North is typical of South California, and is consistent with previous studies on seismic ambient noise at lower frequencies ( $< 1 \text{ Hz}$ ) that traditionally indicate incident seismic noise coming from the Pacific Ocean. Note that the opposite direction also emerges from the ambient seismic noise, which might be due to local topography (mostly parallel to the fault) or to other regional fault structures East of the SJFZ.

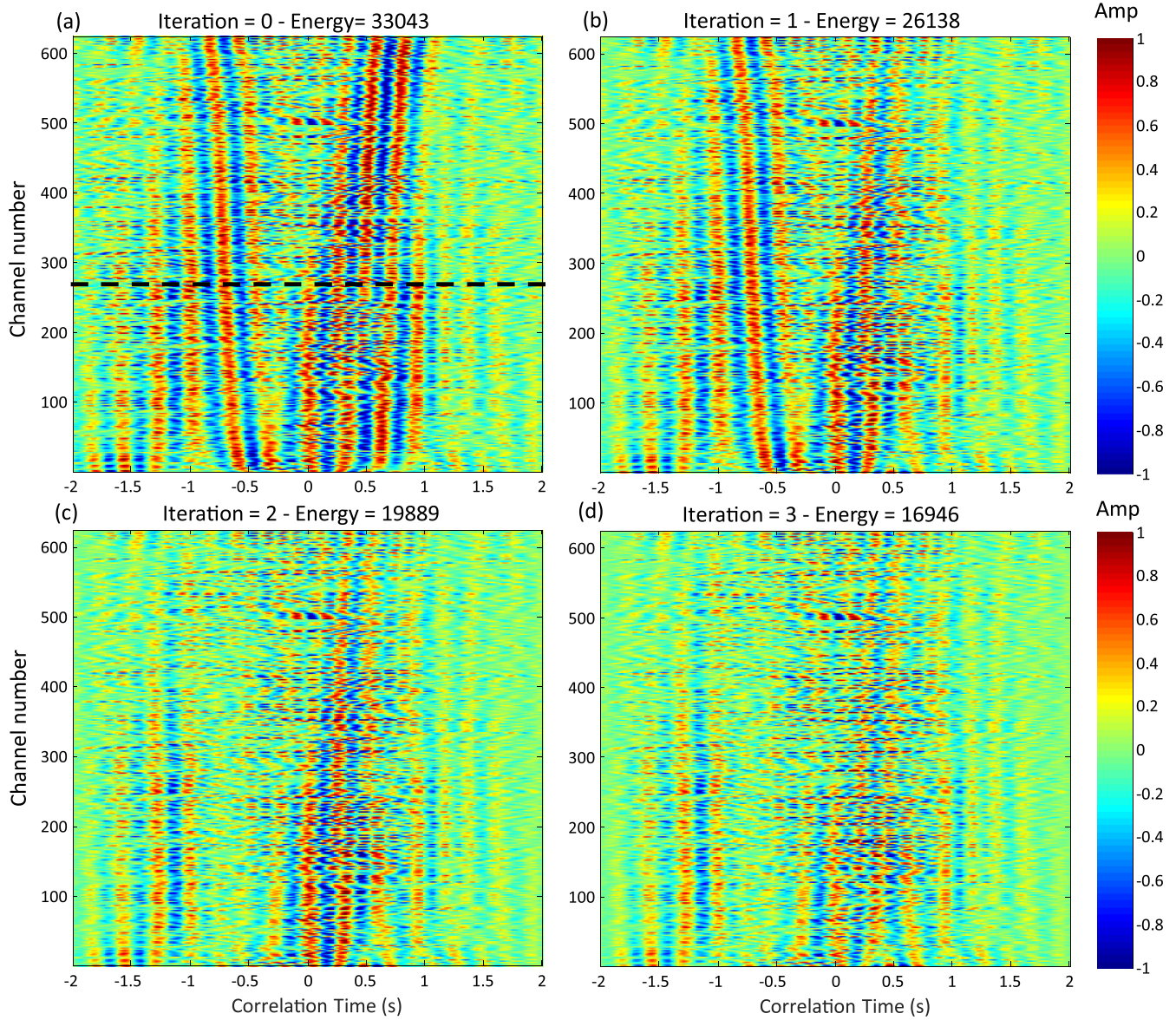
that propagate within a shallow-water ocean (Roux *et al.* 2008). This previous study was extended to geophysics applications using surface arrays at the laboratory scale (De Cacqueray *et al.* 2011). This was then generalized to ambient-noise surface wave imaging at the seismic scale using subarrays of seismometers among the USArray (Boué *et al.* 2014) and three arrays of geophones recently deployed on Piton de la Fournaise Volcano (Nakata *et al.* 2016).

Subarrays are classically chosen to avoid overlap between elements, which correspond to a  $\sim 100\text{-m}$  centre-to-centre distance for subarrays of 25 elements. The size and number of elements

in each subarray also results from a balance between several constraints:

(1) We first need slowness resolution in order to separate surface wave from body waves. By defining every subarray as the closest 25 sensors to every array element, we typically work with a subarray size on the order of 90 m (slightly more for subarrays centred on sensors located in the first 5 lines East of the Fault, see Fig. 1). This provides efficient wavelet separation for surface waves and other wave types as shown in the slowness-versus-time results (see Fig. 8).





**Figure 7.** (a) Iteration 0 corresponds to the initial set of 625 normalized cross-correlations between the two subarrays. In each panel, the cross-correlations are numbered from 1 to 625 (along the y-axis) and sorted by increasing distance. The dashed line corresponds to the correlation associated with the centre of the two green subarrays in Fig. 1 (centred in R2307 and R5207). Note that surface waves are clearly visible since the low wavenumber contribution is filtered out from the raw correlations in the previous step. After extraction of the dominant wave through DBF (see Fig. 8a), the DBF-extracted wavelet is removed from the 625 correlations to allow for the next DBF extraction. Here, iterations 0–3 [(a) to (d)] correspond to the DBF images in Figs 8(a)–(d). After each DBF extraction, the phase velocity, group velocity and amplitude of the extracted wavelet are recorded. For the sake of representation, panels (a) to (d) are normalized according to their maximum amplitudes, although the total energy decreases by a ratio of 2 between iteration 0 and iteration 3.

(2) Another constraint is the number of sensors in one subarray. With 25 elements, the DBF process is performed from  $25 \times 25 = 625$  cross-correlations which is sufficient to benefit from significant array gain on coherent signals and not too large to avoid having subarrays that spread on heterogeneous parts of the SJF.

(3) The last constraint involves the computation time of the whole DBF process that increases significantly with the number of elements taken for every subarray pair.

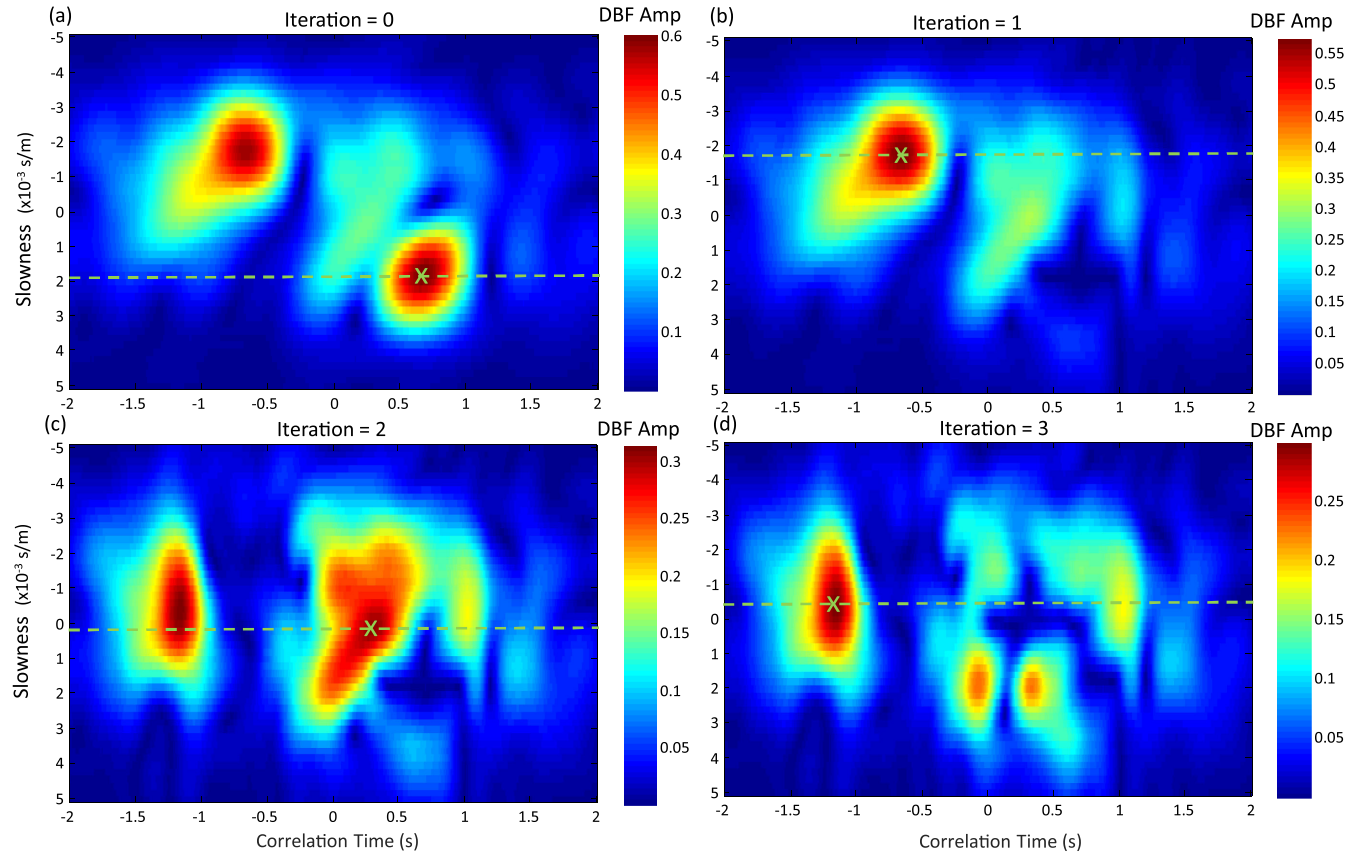
As for any beamforming process, DBF involves delaying and summing wavefields according to the appropriate phase velocities (or slowness):

$$C_{\text{DBF}}(t) = \sum_{i,j} C_{ij}(t - \delta t_{ij}), \quad (1)$$

where  $C_{\text{DBF}}$  is the DBF result obtained from the set of time-domain cross-correlations  $C_{ij}$  between every pair of sensors  $i, j$  among the two arrays, and  $\delta t_{ij}$  is the DBF time delay for each sensor pair. As two arrays are taken into account in the DBF process, local phase velocities and local azimuthal directions should be considered in the calculations of the time delays:

$$\delta t_{ij}(s_1, \theta_1; s_2, \theta_2) = s_1 \overrightarrow{A_i^1 A_c^1} \cdot \vec{u}_1 - s_2 \overrightarrow{A_j^2 A_c^2} \cdot \vec{u}_2, \quad (2)$$

where  $s_1$  and  $s_2$  are the local slowness at the two arrays, and  $\overrightarrow{A_i^1 A_c^1}$  is the position vector between the centre of array 1 ( $A_c^1$ ) and the  $i$ th element of the same array ( $A_i^1$ ). The same notations apply to array 2 for  $\overrightarrow{A_j^2 A_c^2}$ . The unit vectors  $\vec{u}_1 = (\cos \theta_1, \sin \theta_1, 1)$  and  $\vec{u}_2 = (\cos \theta_2, \sin \theta_2, 1)$  denote the azimuthal directions of the DBF



**Figure 8.** The DBF algorithm projects the set of  $25 \times 25 = 625$  correlations into the beam space. After DBF, each high-intensity spot in the slowness-versus-time space corresponds to one spatially coherent wavelet identified among the whole set of cross-correlation pairs. From the beam maximum at each iteration (green cross), one wavelet is computed in the time domain, as shown in Figs 9(b)–(e). At this stage, no correct identification of the extracted wavelet is performed. Note that the beam maximum intensity decreases as the iteration increases, as expected from the iterative subtraction process described in Fig. 7.

projection. The two first components of  $\vec{u}_1$  and  $\vec{u}_2$  refer to the horizontal plane, while the last dimension takes into account the difference in the altitudes between sensors. Assuming similar slowness ( $s_1 = s_2 = s$ ) and azimuthal directions ( $\theta_1 = \theta_2 = \theta$ ) at the two array locations, the DBF time delay simplifies to

$$\delta t_{ij}(s, \theta) = s \left( \vec{A}_i^1 \vec{A}_c^1 - \vec{A}_j^2 \vec{A}_c^2 \right) \cdot \vec{u} = s \left( \vec{A}_i^1 \vec{A}_j^1 - \vec{A}_c^1 \vec{A}_c^1 \right) \cdot \vec{u}, \quad (3)$$

where the position vectors in the right parentheses refer to the vector between two sensors  $i, j$  among each array, and the vector between the two array centres. Note that according to eq. (3), sensors located at the centre of the two arrays should not be time delayed.

Assuming a single slowness  $s$  and azimuthal direction  $\theta$  at the two subarrays (eq. 3) may lead to a bias in a very heterogeneous medium (as the SJFZ) as the effective slowness's may strongly vary depending on the subarray position. This bias was evaluated from a set of different subarrays among the network and was found to be small compared to the slowness beam size (see the size of the maximum intensity spot in Fig. 8a). We then validated this choice that significantly reduces the computational cost of the DBF process.

To further optimize the DBF processing, we match first the directions between the centre of the two subarrays 1 and 2 with the incident-noise direction (Fig. 1). Examining the set of  $25 \times 25 = 625$  cross-correlations between all of the pairs of 25 stations in each subarray, sorted with distance (Fig. 7a), it appears that a single apparent phase velocity can be used to compensate for travel-time

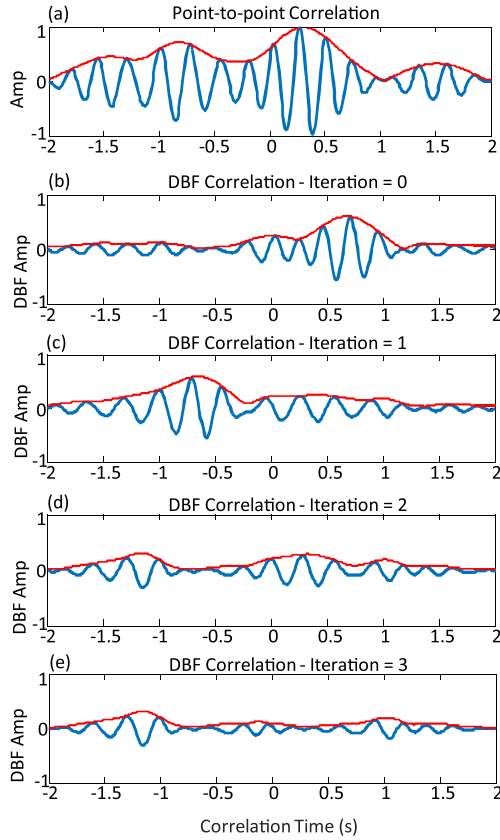
delays. We thus propose to use the following time delays in the DBF process:

$$\delta t_{ij}(s) = s \left( \left| \vec{A}_i^1 \vec{A}_j^1 \right| - \left| \vec{A}_c^1 \vec{A}_c^1 \right| \right). \quad (4)$$

Note that each of the 625 correlations is normalized prior to DBF processing to provide the same weight for every station pair in the beamforming process. This also means that a perfect slant-stack summation would result in a maximum DBF amplitude of 1.

As indicated in eqs (1) and (4), the DBF projects the cross-correlation pairs  $C_{ij}(t)$  from the distance-versus-time space (Fig. 7) to the slowness-versus-time domain (Fig. 8). At first glance, the DBF pattern  $C_{\text{DBF}}(t, s)$  reveals two intensity peaks at positive and negative slowness (Fig. 8a), which correspond to the two wave fronts observed in the cross-correlations for the positive and negative times (Fig. 7a). The DBF analysis sorts out wavefield contributions with different apparent phase velocities that might be blurred out in the individual point-to-point correlations (Fig. 7a). In particular, when comparing the point-to-point correlations between the centres of the two subarrays (Fig. 9a) to the DBF correlation (Fig. 8a, green cross, Fig. 9b), it is seen that the DBF processing removes some strong bias in the travel-time propagation of the surface waves. These ambiguities are due to weak arrivals in the raw correlations that are superimposed on the expected surface waves, as previously observed for a non-isotropic noise distribution in a complex and reverberating environment (Snieder & Fleury 2010; Colombi *et al.* 2014).





**Figure 9.** (a) Normalized point-to-point correlation between the centre of the two subarrays in Fig. 1, namely stations R2307 and R5207 (Fig. 7(a), dashed line). (b–e) DBF correlations obtained from iteration 0 to iteration 3. The traveltimes are measured from the maximum of the envelope (red). The group and phase velocities obtained for each DBF wavelet are as follows: iteration 0— $V_p = 574.6 \text{ m s}^{-1}$ ,  $V_g = 384.1 \text{ m s}^{-1}$ ; Iteration 1— $V_p = -576.5 \text{ m s}^{-1}$ ,  $V_g = -395.7 \text{ m s}^{-1}$ . The two first iterations correspond to Rayleigh waves from R2307 to R5207 (positive time) and from R5207 to R2307 (negative time), as expected from the correlation theorem in the case of omnidirectional noise. The two DBF wavelets extracted at iterations 2 and 3 are spurious, and might be due to pre-processing artefacts or the combination of the complex fault structure and the non-isotropic noise distribution. Note that the surface wave extraction would have been impossible from the point-to-point correlation in panel (a).

Since the DBF is a linear process, the wavelet extraction is not limited to the most energetic contribution, but can be iterated for several apparent phase velocities, as illustrated by the iteration process in Figs 7–10. The method works as follows. When one wavelet is identified through an intensity maximum in the slowness-versus-time DBF pattern (Fig. 8, green crosses), the DBF wavelet is computed using the set of time delays that are associated with the appropriate slowness (Figs 9b–e). This wavelet is then isolated in a time window and back-projected into the correlation-versus-time space for all of the correlation pairs. The duration of the time window is chosen according to the wave period and the frequency bandwidth of the noise-whitening process. This allows the reconstruction of the selected wavelet with constant amplitude for each station pair. However, as can be observed from the raw correlations between two subarrays (Fig. 7a), the amplitude pattern of each coherent wavelet is not uniform from one sensor pair to the next one in a  $25 \times 25$  configuration but shows a slow amplitude evolution with respect to the channel number axis. This is particularly true for the two wavelets selected at iterations 2 and 3 (Figs 7c and d), but it also

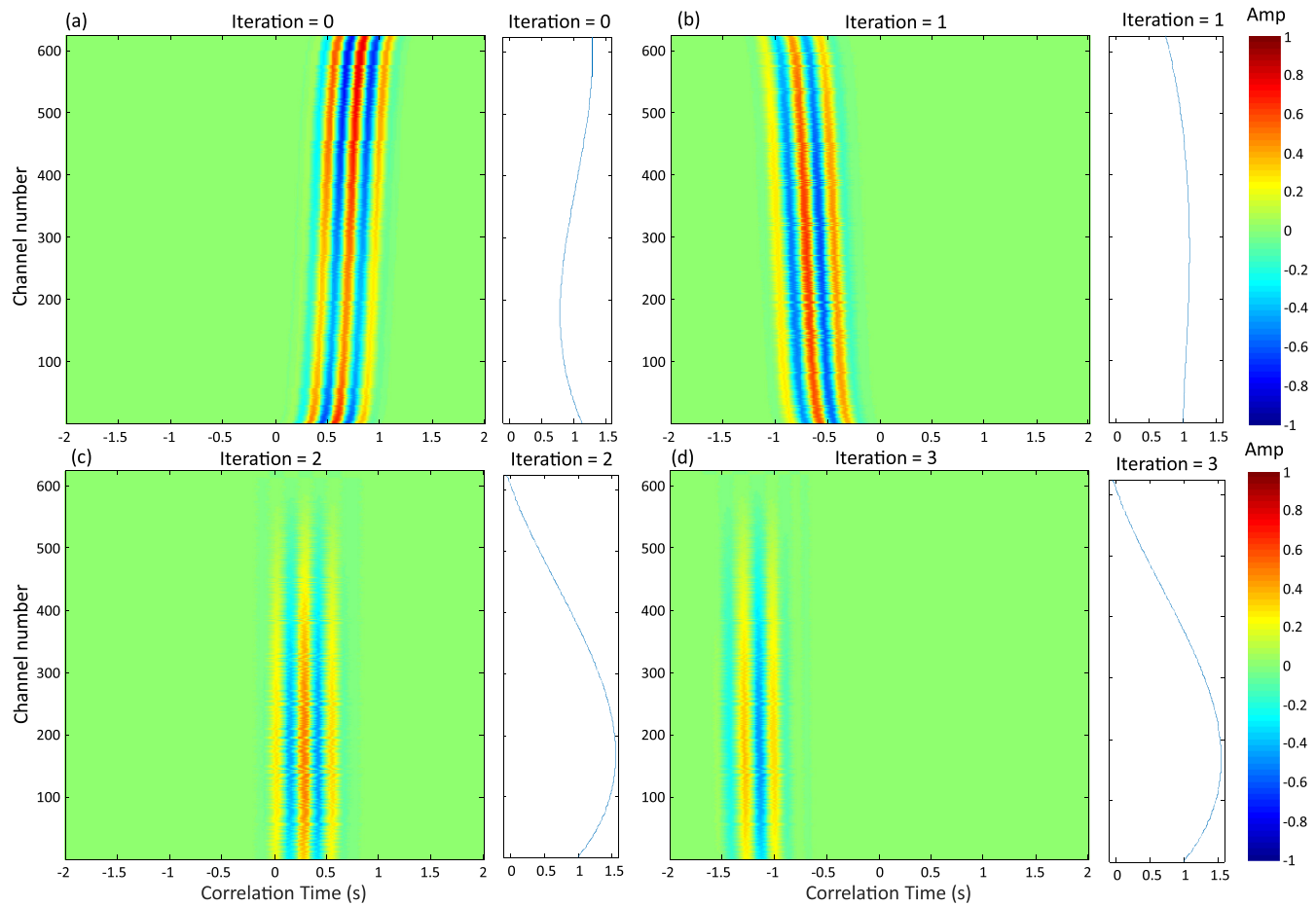
applies to the surface waves extracted at iterations 0 and 1. The reason for this amplitude evolution could be most likely due to the non-isotropic noise source distribution and the spatial heterogeneities in the medium.

In order to optimize the wavelet subtraction from the raw correlation data, this amplitude change from pair-to-pair has to be taken into account. One solution could have been to apply a mask on the slowness-versus-time DBF representation outside of the beam intensity associated with the selected wavelet (Fig. 8), and to reconstruct the wavelet pattern on each sensor pair with time and amplitude applying the inverse of the DBF transformation. This procedure (DBF transform + mask + inverse DBF transform) is very similar to wavenumber filter applied to the  $K_x$ – $K_y$  diagram in Fig. 4. However, the shape of the mask in the slowness-versus-time domain must adjust to the effective phase velocity of the selected wave. This wave-dependent mask makes the wavelet reconstruction and subtraction process unstable and inefficient. We therefore proceeded differently. We reconstructed the time-domain wavelet by (1) selecting the phase velocity maximum after DBF in the slowness-versus-time representation (see green cross at each iteration in Fig. 8) and then (2) adjusting the amplitude dependence of this wavelet along the channel number axis through a third-order polynomial fit (Fig. 10) that minimizes the residual energy after the wavelet subtraction process. The wavelet extraction and subtraction process becomes then automatic (not user-dependent), robust and stable, which is required to process the  $\sim 610\,000$  subarray pairs that can be taken out of this dense array recording.

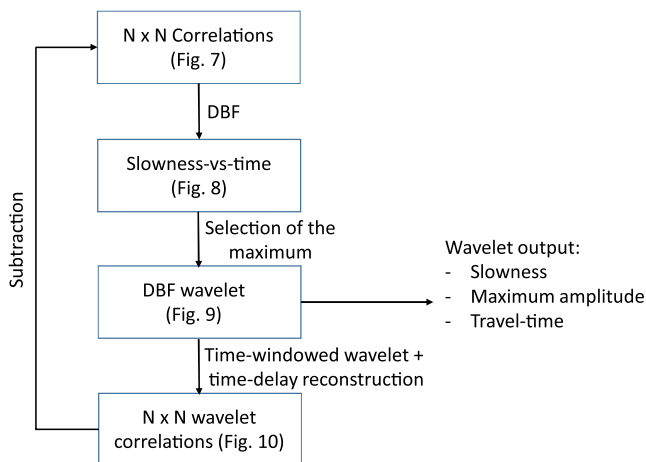
In practice, this means that the correlations at iteration 1 in Fig. 7(b) are the direct subtractions of the correlations at iteration 0 in Fig. 7(a) and the reconstructed wavelet at iteration 0 in Fig. 10(a). To summarize this sequence of operations at each iteration (Fig. 11): (1) DBF is performed on the correlations (Fig. 8); (2) one wavelet is extracted and isolated from the DBF maximum (Fig. 9); (3) the correlations-versus-time pattern is reconstructed for this wavelet and for all of the pairs with an optimized pair-to-pair amplitude (Fig. 10); and finally (4) these wavelet-driven correlations are subtracted from the original correlations (Fig. 7).

Going back to the correlation patterns at each iteration (Fig. 7), additional wavelets are clearly observed in the correlation data once the dominant surface waves have been subtracted, in both the positive and negative times. Some of these wavelets arrive at late times (i.e. with low group velocities) and with high apparent phase velocities (Fig. 10d). Some others arrive at early times and are superimposed with the direct surface wave arrivals (Figs 10a and c), resulting in a potential bias in the traveltimes extraction from a point-to-point analysis. As suggested by Figs 6–9 of Ben-Zion *et al.* (2015), these might be due to multiple reverberations from horizontal and vertical interfaces underneath and within the fault-zone structure. These might also result, at least partially, from pre-processing artefacts associated with the one-bit clipping applied to the ambient-noise signals before cross-correlation.

In this analysis, up to eight iterations are performed for each subarray pair, and the slowness, traveltimes (measured at the maximum of the DBF envelope) and amplitudes of the DBF maxima are systematically recorded at each iteration. To distinguish between physical and non-physical wavelet extractions, we focus on a set of subarrays that are centred on a line of receiver pairs that is parallel to the noise direction (Fig. 1). Each red and black dot represents the centre of a 25-geophone subarray, between which the DBF iterative procedure was performed. Group and phase velocity-versus-distance results are shown in Fig. 12 for frequencies  $F = 2$ ,



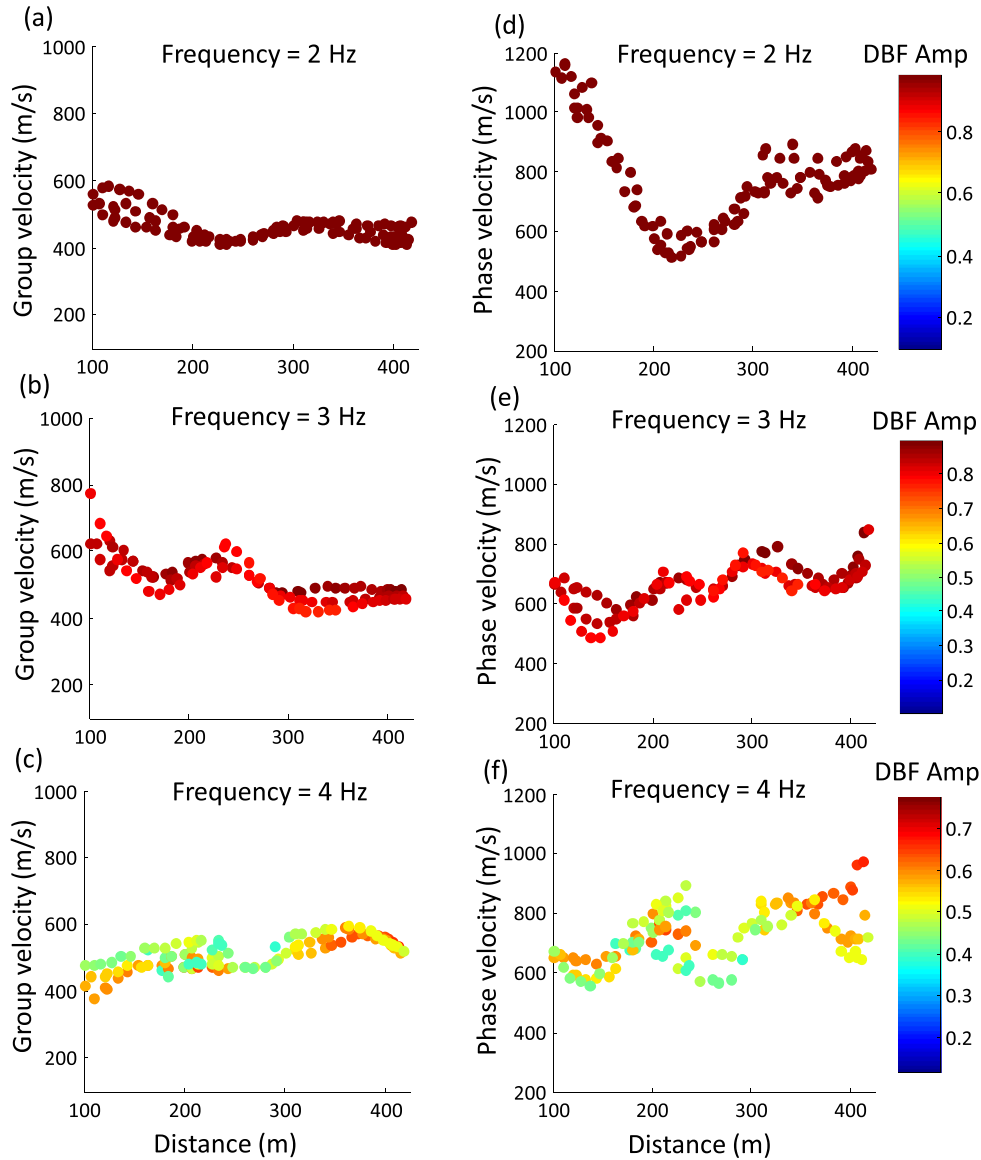
**Figure 10.** (a–d) Reconstructed wavelet for each of the 625 station pairs at each iteration (0 to 3), as obtained from the DBF extraction process. Each wavelet contribution should be compared to the corresponding data-based correlations at each iteration in Fig. 7. Note the amplitude dependence of the reconstructed wavelets along the 625 station pairs (especially for iterations 2 and 3) as the result of amplitude optimization based on a polynomial fit. The iterative subtraction process is such that the set of correlations at iteration 1 in Fig. 7(b) is obtained from the subtraction of the correlations at iteration 0 in Fig. 7(a) and the reconstructed wavelet at iteration 0 in Fig. 10(a).



**Figure 11.** Schematic representation using block diagrams of the iterative DBF processing applied to the correlation data. Namely: (1) DBF is applied on one subarray pair; (2) one wavelet is extracted and isolated for the slowness obtained at the DBF maximum; (3) the correlations are reconstructed for this wavelet and for all of the geophone pairs among the two subarrays; and finally, (4) these wavelet-driven correlations are subtracted from the original correlations before the next iteration is performed.

3 and 4 Hz. Group velocities are computed as the distance between the subarray centres divided by traveltimes obtained at the maximum of the DBF-extracted envelope for each wavelet (red curve in each panel in Fig. 9). The phase velocities are plotted as the inverse of the slowness, where the colour bar denotes the amplitude of each extracted wave. Note that both positive and negative traveltimes and slowness are superimposed on the same representation as the absolute values of the group and phase velocities.

In Fig. 12(a), only wavelets with strong amplitudes, most likely corresponding to Rayleigh waves, have been plotted with group velocities slowly evolving with distance in agreement with the location of the SJFZ. Phase velocities are generally higher, which is expected for dispersive waves in a medium with increasing velocity with depth. The DBF amplitudes of the extracted Rayleigh waves, which indicate the spatial coherence between the subarrays, decrease with increasing frequency. We also note a slight dispersion between the group and phase velocity measurements for Rayleigh waves at the same distance, which is due to weak fluctuations between the wavelet parameters in the positive and negative correlation times. These uncertainties reflect the potential error of the method, which appears to be limited to 10 per cent in this complex geophysical environment when the subarrays are aligned with the noise direction. Across the fault zone, variations in the group and



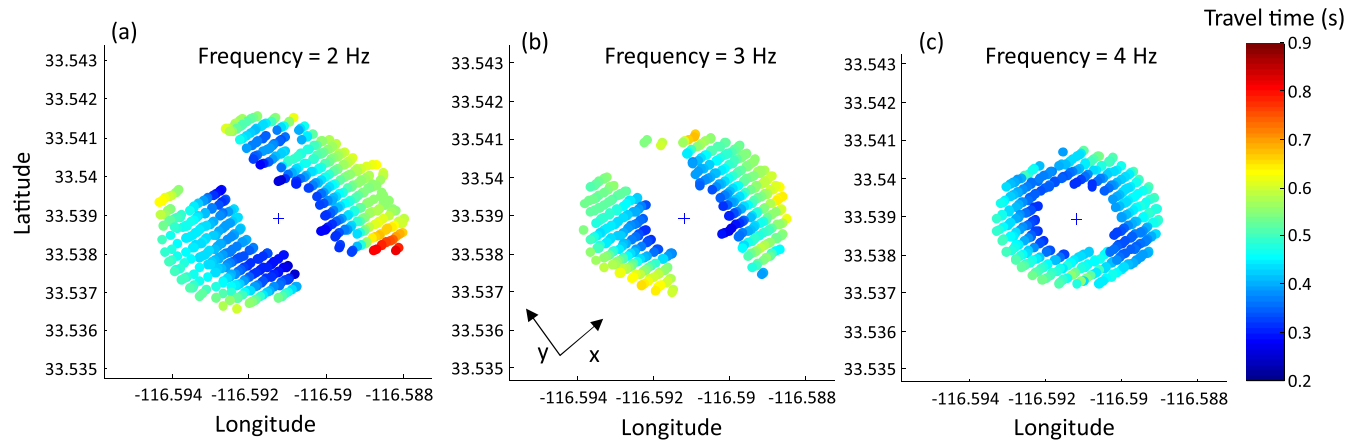
**Figure 12.** Representation of the DBF-extracted parameters plotted as velocity versus distance at different frequencies for the subarray centres (red, black) shown in Fig. 1. Group velocities are plotted in (a–c), and phase velocities are plotted in (d–f), for the DBF processing centred at 2, 3 and 4 Hz. At each frequency, the colour bar corresponds to the maximum amplitude of the DBF wavelet. As expected from Figs 9(a) and (b) and Figs 10(b) and (c), the high-amplitude DBF wavelets correspond to Rayleigh waves obtained from both the positive and negative times of the correlations. Note the smooth spatial and frequency dependence of the Rayleigh wave velocities that reflects the spatial heterogeneity of the elastic parameters perpendicular to the fault.

phase velocities with frequency and distance provide evidence for the strong subsurface heterogeneity of the SJFZ. This was seen in the early results of Ben-Zion *et al.* (2015) and should be further quantified by surface wave tomography.

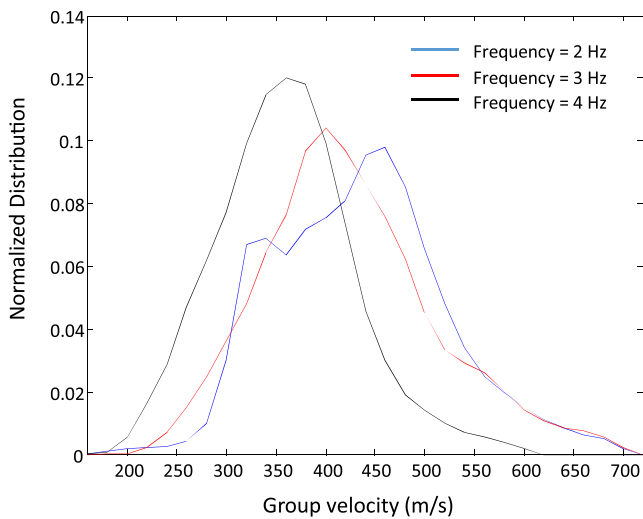
To illustrate the potential of the method, travel-time extraction is performed for Rayleigh waves between one reference geophone at the centre of the dense array and all of the possible geophones with an exclusion distance of 100 m. As for the DBF data shown in Fig. 12, each ‘geophone’ in Fig. 13 actually refers to the centre of a 25-element subarray. The exclusion distance was set to avoid superimposition between two subarrays at short distances. Only DBF results with amplitudes larger than a threshold value are kept in this analysis, thus focusing the wave selection on the dominant Rayleigh waves. This coherence threshold was set to 0.8 at 2 Hz, 0.7 at 3 Hz and 0.45 at 4 Hz. The smooth maps in Fig. 13 indicate that the travel-time extraction is consistent over the whole array, even

for pairs of subarrays that are not aligned with the ambient-noise direction. However, we note that subarray pairs that are orthogonal to the noise direction are missing at low frequencies (Figs 13a and b, blank areas), where the noise is very directional. These maps of the traveltimes at different frequencies (Figs 13a–c) can form the main ingredient for eikonal surface wave tomography (Roux *et al.* 2011; Lin *et al.* 2013; Mordret *et al.* 2013). If the medium is homogeneous, the travel-time maps would show concentric coloured circles centred at the reference station, which is not the case. The traveltimes elongations along the  $y$ -axis (parallel to the fault zone) confirm the structural heterogeneities perpendicular to the SJFZ. When applied successively to all of the reference geophones, a total of more than 400 000 Rayleigh wave traveltimes [484 000 at 4 Hz, 438 000 at 3 Hz and 427 000 at 2 Hz] are extracted among the 614 000 potential geophone pairs that can be taken out of the 1108 geophone array (Fig. 14).





**Figure 13.** Maps of the traveltimes for Rayleigh waves extracted from DBF processing between one subarray centred at the middle of the dense array and all possible subarrays selected among the dense array. The maps correspond to three different frequencies (a–c). Rayleigh waves are selected through their dominant amplitude in the DBF processing. Note that each subarray is represented by its central geophone as in Fig. 1 (red and black circles).



**Figure 14.** Normalized group velocity distribution at each frequency obtained after surface wave extraction from every subarray pair among the dense array network. The group velocities are calculated from the distance between the centre of each subarray pair and the extracted traveltimes (maxima of the wavelet envelopes at each iteration in Fig. 9). At each frequency, the distribution is calculated for 20 m s<sup>-1</sup> bin intervals and normalized by the total number of group velocities with 427 000 at 2 Hz, 438 000 at 3 Hz and 484 000 at 4 Hz.

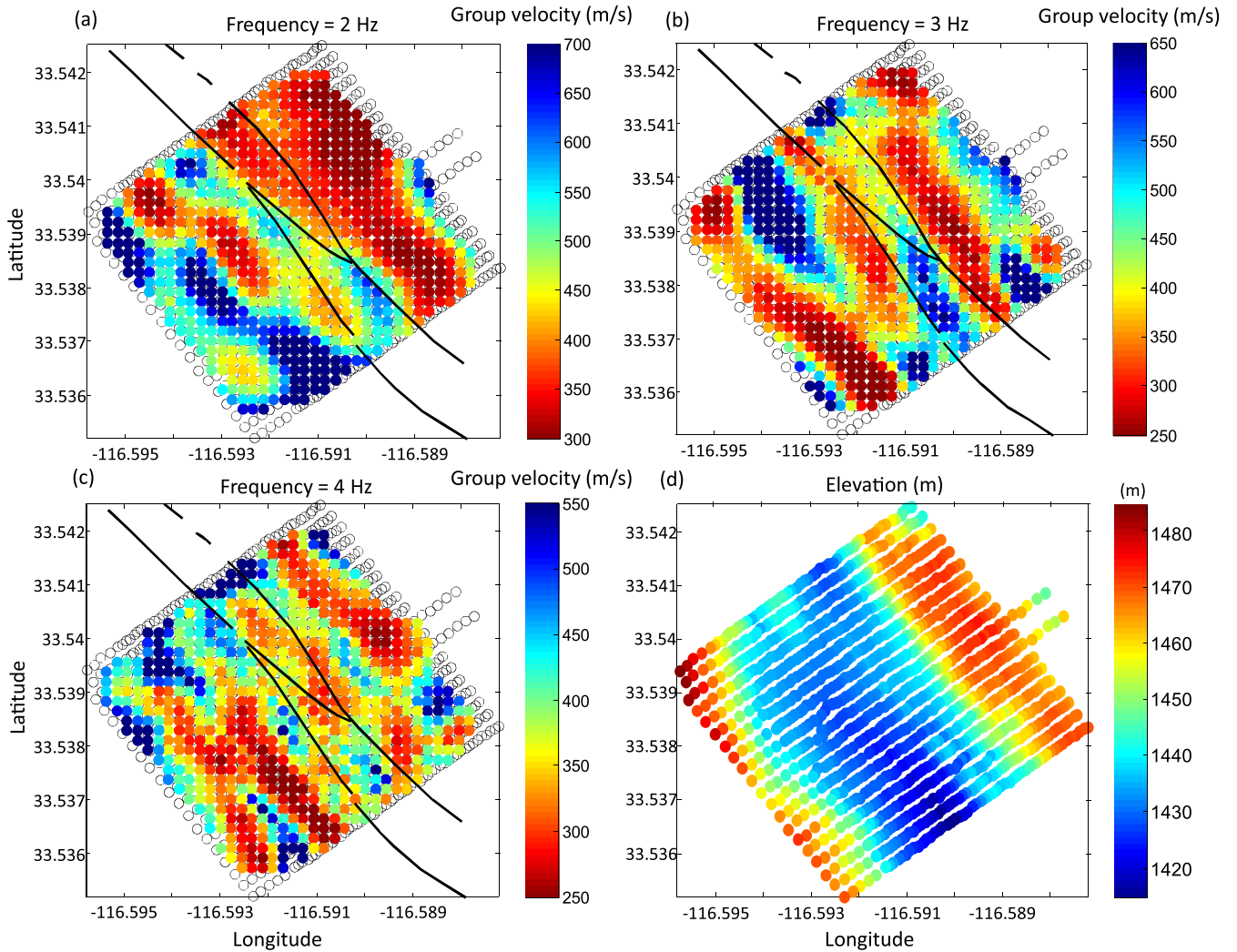
Fig. 15 shows the Rayleigh wave group velocity maps obtained at 2, 3 and 4 Hz as a result of the traveltime inversion on a regular horizontal grid with steps of 15 m. A simple linear inversion for the slowness was performed on a random selection of 20 000 arrival times (Barmin *et al.* 2001). This assumes straight rays as propagation paths, and an *a-priori* error covariance matrix that decreases exponentially with distances over 50 m. The weight of the spatial smoothing was classically chosen at the maximum curvature of the standard trade-off analysis (L-curve) based on the misfit value (Hansen & O’Leary 1993). The inversion started from a homogeneous initial model with a group velocity of 480 m s<sup>-1</sup> at 2 Hz, 425 m s<sup>-1</sup> at 3 Hz and 360 m s<sup>-1</sup> at 4 Hz, according to the distribution of group velocities plotted in Fig. 14. A total of 100 inversions were performed, and Fig. 15 shows the average group velocities for the positions where the standard deviation is less than 20 per cent of the average, which almost corresponds to the whole of the

surface array. At each frequency, each inversion produces a residual variance reduction of ~98 per cent relative to the arrival times for the homogeneous model. Note that similar surface wave inversion could be performed for phase velocities instead of group velocities from the collection of phase velocities extracted for each subarray pairs from the DBF process (Figs 12d–f). For sake of clarity and simplicity, we limited the representation to maps of traveltimes in Fig. 13 and to the corresponding map of group velocities after inversion in Fig. 15.

The ratio of the velocities divided by the frequency range used suggests that the observations characterize the top 75 m to 350 m of the crust. Topography was not taken into account in the inversion (as suggested in Pilz *et al.* 2013), but the correlation between the group velocity maps and the topography is not obvious (Fig. 15d). The results show strong lateral variations across the fault traces. The red low-velocity zone near the right fault line corresponds to the fault zone trapping structure identified in Ben-Zion *et al.* (2015) using active experiment data. The results for increasing frequencies have features that correspond progressively to shallower depth. The 2 Hz results (Fig. 15a) show a low velocity zone between the two fault traces in the northwest part of the array but not in the southeast. This discontinuous low velocity zone is less clear at the higher frequency images (Figs 15b and c), so it may have little evidence in the surface geology. The red zone to the right of the fault traces at 2 and 3 Hz is correlated with the small local topography of highly damaged (partially pulverized) rocks, and the red zone to the left of the fault traces (shown differently at the different frequencies) corresponds to a small sedimentary basin.

#### 4 DISCUSSION

This paper presents a new approach for wave separation that allows the derivation of detailed velocity images in complex environments from continuous dense recordings of ambient seismic noise. The core idea is associated with using an iterative technique that involves DBF between various subarrays to extract the different phases that are in the correlation results and to use the different phases for structural imaging. The initial results presented in Figs 12–15 illustrate both the structural complexity in the area and the richness of information that can be provided by the spatially dense data and the used imaging technique. The analysis in this study focuses on Rayleigh surface waves, but we envision that reflected and refracted body



**Figure 15.** (a–c) Group velocity maps obtained at 2, 3 and 4 Hz after surface wave inversion from more than 400 000 traveltimes on a 15-m-spacing regular grid. The circles in the background correspond to the sensors' location. The black lines indicate the surface traces of the faults and correspond to the fault directions associated with the SJFZ. Strong lateral velocity variations are observed across the fault structure that does not appear to correlate with the local topography (d).

waves can also be isolated between subarrays through DBF, and combined with surface waves for joint inversion of the subsurface material.

Additional structural information might be obtained by performing combined inversions between group and phase velocities using a wider range of discrete frequencies. The close spacing between the sensors allows the extraction of local velocity information on the shallow material at tens of Hz. Ben-Zion *et al.* (2015) provided an example of Rayleigh wave group velocities at 50 Hz obtained along a fault-normal line of the nodal array. This can be generalized with array processing of the type carried out here. The surface wave group velocities can be inverted to 3-D models of shear-wave velocities using Bayesian tomography inversion for all of the geophone pairs, similar to what was done in the context of Fig. 14, or using eikonal tomography performed from the travel-time maps obtained for each geophone (Lin *et al.* 2013; Mordret *et al.* 2013). A future study will focus on this objective, with an overall goal of imaging the 3-D structure of the SJFZ from the very shallow sub-surface to depths of  $\sim 500$  m.

From a first 3-D model of the SJFZ, we will be able to compute numerically higher-order surface wave arrivals and also locally scat-

tered body-waves. It will then be possible to revisit the weak arrivals extracted at iterations 2 and 3 (see Figs 8 and 9) and match them with the numerically predicted observations. We also note that a single day of ambient noise recordings was enough to retrieve high-amplitude correlations in the frequency band of interest. Equivalent results could be obtained with shorter ambient-noise intervals (down to 1 hr), as DBF provides significant array gain and should compensate for the loss of coherence due to correlations on limited time recordings. The production of repeated hourly based 3-D images of the SJFZ can be used to monitor the structural evolution of the shallow part of the fault. These topics will be the focus of future studies.

## ACKNOWLEDGEMENTS

ISTerre is part of Labex OSUG@2020. GH acknowledges support through a Heisenberg Fellowship from the German Research Foundation (HI-1714/1-2). Most of the computations presented here were performed using the CIMENT infrastructure (<https://ciment.ujf-grenoble.fr>), which is supported by the Rhône-Alpes

region (GRANT CPER07\_13 CIRA: <http://www.ci-ra.org>) and France-Grille (<http://www.france-grilles.fr>). We thank Dan Hollis and Mitchell Barklage from NodalSeismic. The paper benefitted from comments by Norimitsu Nakata and an anonymous referee.

## REFERENCES

- Allam, A.A. & Ben-Zion, Y., 2012. Seismic velocity structures in the Southern California plate-boundary environment from double-difference tomography, *Geophys. J. Int.*, **190**, 1181–1196.
- Allam, A.A., Ben-Zion, Y., Kurzon, I. & Vernon, F.L., 2014. Seismic velocity structure in the Hot Springs and Trifurcation Seismicity Cluster Areas of the San Jacinto Fault Zone from double-difference tomography, *Geophys. J. Int.*, **171**, 2993–3011.
- Barmin, M., Ritzwoller, M. & Levshin, A., 2001. A fast and reliable method for surface wave tomography, *Pure appl. Geophys.*, **158**(8), 1351–1375.
- Ben-Zion, Y., 2008. Collective behavior of earthquakes and faults: continuum-discrete transitions, evolutionary changes and corresponding dynamic regimes, *Rev. Geophys.*, **46**, RG4006, doi: 10.1029/2008RG000260.
- Ben-Zion, Y. *et al.*, 2015. Basic data features and results from a spatially-dense seismic array on the San Jacinto fault zone, *Geophys. J. Int.*, **202**, 370–380.
- Boué, P., Roux, P., Campillo, M. & Briand, X., 2014. Phase velocity tomography of surface waves using ambient noise cross-correlation and array processing, *J. geophys. Res.*, **119**(1), 519–529.
- Colombi, A., Boschi, L., Roux, P. & Campillo, M., 2014. Green's function retrieval through cross-correlations in a two-dimensional complex reverberating medium, *J. acoust. Soc. Am.*, **135**(3), 1034–1043.
- De Cacqueray, B., Roux, P., Campillo, M., Catheline, S. & Boué, P., 2011. Elastic-wave identification and extraction through array processing: an experimental investigation at the laboratory scale, *J. Appl. Geophys.*, **74**, 81–88.
- Gerstoft, P. & Tanimoto, T., 2007. A year of microseisms in southern California, *Geophys. Res. Lett.*, **34**, L20304, doi:10.1029/2007GL031091.
- Hansen, P. & O'Leary, D., 1993. The use of the L-Curve in the regularization of discrete ill-posed problems, *SIAM J. Sci. Comput.*, **14**, 1487–1503.
- Hansen, S.M. & Schmandt, B., 2015. Automated detection and location of microseismicity at Mount St. Helens with a large-N geophone array, *Geophys. Res. Lett.*, **42**, 7390–7397.
- Hauksson, E., Yang, W. & Shearer, P.M., 2012. Waveform relocated earthquake catalog for Southern California (1981 to June 2011), *Bull. seism. Soc. Am.*, **102**(5), 2239–2244.
- Hillers, G., Ben-Zion, Y., Landès, M. & Campillo, M., 2013. Interaction of microseisms with crustal heterogeneity: a case study from the San Jacinto Fault Zone area, *Geochem. Geophys. Geosyst.*, **14**(7), 2182–2197.
- Johnson, H.O., Agnew, D.C. & Wyatt, F.K., 1994. Present-day crustal deformation in Southern California, *J. geophys. Res.*, **99**(B12), 23 951–23 974.
- Kurzon, I., Vernon, F.L., Ben-Zion, Y. & Atkinson, G., 2014. Ground motion prediction equations in the San Jacinto Fault Zone: significant effects of rupture directivity and fault zone amplification, *Pure appl. Geophys.*, **171**, 3045–3081.
- Lin, F.C., Li, D., Clayton, R.W. & Hollis, D., 2013. High-resolution 3D shallow crustal structure in Long Beach, California: application of ambient-noise tomography on a dense seismic array, *Geophysics*, **78**(4), Q45–Q56.
- Lindsey, E.O. & Fialko, Y., 2013. Geodetic slip rates in the southern San Andreas Fault system: effects of elastic heterogeneity and fault geometry, *J. geophys. Res.*, **118**, 689–697.
- Mordret, A., Shapiro, N.M., Singh, S., Roux, P., Montagner, J.P. & Barkved, O.I., 2013. Azimuthal anisotropy at Valhall: the Helmholtz equation approach, *Geophys. Res. Lett.*, **40**, 2636–2641.
- Nakata, N., Boué, P., Brenguier, F., Roux, P., Ferrazzini, V. & Campillo, M., 2016. Body and surface wave reconstruction from seismic noise correlations between arrays at Piton de la Fournaise volcano, *Geophys. Res. Lett.*, **43**, 1047–1054.
- Pilz, M., Parolai, S. & Bindi, D., 2013. Three-dimensional passive imaging of complex seismic fault systems: evidence of surface traces of the Issyk-Ata fault (Kyrgyzstan), *Geophys. J. Int.*, **194**, 1955–1965.
- Rost, S. & Thomas, C., 2002. Array seismology: methods and applications, *Rev. Geophys.*, **40**(3), 1008, doi: 10.1029/2000RG000100.
- Roux, P., 2009. Passive seismic imaging with direct ambient noise: application to surface waves and the San Andreas Fault in Parkfield, CA, *Geophys. J. Int.*, **179**, 367–373.
- Roux, P., Cornuelle, B.D., Kuperman, W.A. & Hodgkiss, W.S., 2008. The structure of ray-like arrivals in a shallow-water waveguide, *J. acoust. Soc. Am.*, **124**(6), 3430–3439.
- Roux, P., Wathelet, M. & Roueff, A., 2011. The San Andreas Fault revisited through seismic-noise and surface-wave tomography, *Geophys. Res. Lett.*, **38**, L13319, doi: 10.1029/2011GL047811.
- Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A. & Fehler, M.C., 2005a. Extracting time-domain Green's function estimates from ambient seismic noise, *Geophys. Res. Lett.*, **32**, L03310, doi: 10.1029/2004GL021862.
- Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A. & Fehler, M.C., 2005b. Surface-wave tomography from microseisms in Southern California, *Geophys. Res. Lett.*, **32**, L14311, doi: 10.1029/2005GL023155.
- Schulte-Pelkum, V., Earle, P.S. & Vernon, F.L., 2004. Strong directivity of ocean-generated seismic noise, *Geochem. Geophys. Geosyst.*, **5**, Q03004, doi:10.1029/2003GC000520.
- Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **29**, 1615–1617.
- Sharp, R.V., 1967. San Jacinto Fault Zone in the Peninsular Ranges of Southern California, *Bull. geol. Soc. Am.*, **78**, 705–730.
- Slater, D. & Hollis, D., 2012. California's dense urban environment spawns friendlier 3D seismic survey design, *Oil Gas J.*, **110**(11), 54–57.
- Snieder, R. & Fleury, C., 2010. Cancellation of spurious arrivals in Green's function retrieval of multiple scattered waves, *J. acoust. Soc. Am.*, **128**, 1598–1605.
- Yang, H., Li, Z., Peng, Z., Ben-Zion, Y. & Vernon, F., 2014. Low-velocity zones along the San Jacinto Fault, Southern California, from body waves recorded in dense linear arrays, *J. geophys. Res.*, **119**, 8976–8990.
- Zigone, D., Ben-Zion, Y., Campillo, M. & Roux, P., 2015. Seismic tomography of the Southern California plate boundary region from noise-based Rayleigh and Love waves, *Pure appl. Geophys.*, **172**, 1007–1032.