



# Pre-processing ambient noise cross-correlations with equalizing the covariance matrix eigenspectrum

Léonard Seydoux, Julien de Rosny, Nikolai M. Shapiro

## ► To cite this version:

Léonard Seydoux, Julien de Rosny, Nikolai M. Shapiro. Pre-processing ambient noise cross-correlations with equalizing the covariance matrix eigenspectrum. *Geophysical Journal International*, 2017, 210, pp.1432-1449. 10.1093/gji/ggx250 . insu-03748842

**HAL Id: insu-03748842**

**<https://insu.hal.science/insu-03748842>**

Submitted on 10 Aug 2022

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

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

# Pre-processing ambient noise cross-correlations with equalizing the covariance matrix eigenspectrum

Léonard Seydoux,<sup>1,2</sup> Julien de Rosny<sup>2</sup> and Nikolai M. Shapiro<sup>2,3</sup>

<sup>1</sup>*Institut de Physique du Globe de Paris, UMR CNRS 7154, F-75005 Paris, France. E-mail: leonard.seydoux@gmail.com*

<sup>2</sup>*ESPCI Paris, CNRS, Institut Langevin, PSL Research University, F-75005 Paris, France*

<sup>3</sup>*Institute of Volcanology and Seismology FEB RAS, 9 Piip Boulevard, 683006 Petropavlovsk-Kamchatsky, Russia*

Accepted 2017 June 6. Received 2017 May 22; in original form 2017 February 15

## SUMMARY

Passive imaging techniques from ambient seismic noise requires a nearly isotropic distribution of the noise sources in order to ensure reliable traveltimes measurements between seismic stations. However, real ambient seismic noise often partially fulfils this condition. It is generated in preferential areas (in deep ocean or near continental shores), and some highly coherent pulse-like signals may be present in the data such as those generated by earthquakes. Several pre-processing techniques have been developed in order to attenuate the directional and deterministic behaviour of this real ambient noise. Most of them are applied to individual seismograms before cross-correlation computation. The most widely used techniques are the spectral whitening and temporal smoothing of the individual seismic traces. We here propose an additional pre-processing to be used together with the classical ones, which is based on the spatial analysis of the seismic wavefield. We compute the cross-spectra between all available stations pairs in spectral domain, leading to the data covariance matrix. We apply a one-bit normalization to the covariance matrix eigenspectrum before extracting the cross-correlations in the time domain. The efficiency of the method is shown with several numerical tests. We apply the method to the data collected by the USArray, when the M8.8 Maule earthquake occurred on 2010 February 27. The method shows a clear improvement compared with the classical equalization to attenuate the highly energetic and coherent waves incoming from the earthquake, and allows to perform reliable traveltimes measurement even in the presence of the earthquake.

**Key words:** Seismic interferometry; Seismic tomography; Statistical seismology.

## 1 INTRODUCTION

Fifteen years ago, Weaver & Lobkis (2001) demonstrated that the Green's function between two receivers probing a diffuse acoustic wavefield could be retrieved from the cross-correlation of the receivers' records. First applications of this theory in seismology (Campillo 2003; Shapiro & Campillo 2004; Sabra *et al.* 2005a; Shapiro *et al.* 2005) and oceanography (Roux *et al.* 2004; Sabra *et al.* 2005b) have led to a class of routinely used passive imaging methods in both domains. Ambient-noise based surface wave tomography studies have been done at global scale using hum (Nishida *et al.* 2009; Haned *et al.* 2015), at regional scale (Shapiro *et al.* 2005; Sabra *et al.* 2005a; Behm *et al.* 2016), at local scales (e.g. for the study of faults and volcanoes Brenguier *et al.* 2007; Roux 2009; Jaxybulatov *et al.* 2014; Hillers *et al.* 2016; Lanza *et al.* 2016) and in industrial contexts (Mordret *et al.* 2013, 2014; de Ridder *et al.* 2014). Body-waves can also be extracted from the cross-correlation of ambient noise at global scale (Boué *et al.* 2013; Lin *et al.* 2013)

and at local scale (Nakata *et al.* 2015). Also, several applications have been done in passive monitoring (known as 4-D tomography) of active zones such as faults (Wegler & Sens-Schönfelder 2007; Brenguier *et al.* 2008a; Wu *et al.* 2016) or volcanoes (Sens-Schönfelder & Wegler 2006; Duputel *et al.* 2009; Brenguier *et al.* 2008b, 2014, 2015).

A full recovery of the Green's functions from the noise cross-correlations is based on the assumption that the wavefield is equipartitioned, that is, all the modes are uncorrelated and equally excited. To reach this regime, the seismic noise should be generated by uncorrelated sources evenly distributed in the medium (Derode *et al.* 2003; Wapenaar 2004). Nevertheless, this condition is often not verified: seismic events such as earthquakes (that are localized in time and space) often produce strong-amplitude transient signals which are highly coherent. Also, the ambient microseismic noise (Longuet-Higgins 1950) is generated at specific regions of the Earth's surface, and shows a seasonal dependence due to its generating mechanisms (e.g. Stehly *et al.* 2006). While the primary

microseisms are likely due to the direct interaction between oceanic swells and the seafloor in shallow water (e.g. Traer *et al.* 2012; Ermer *et al.* 2015), the secondary microseisms can be both generated in deep ocean or coastal areas by the non-linear interaction of oceanic gravity waves propagating in opposite directions, from which are generated pressure oscillations on the seafloor responsible for seismic waves generation (e.g. Longuet-Higgins 1950; Landès *et al.* 2010; Bromirski *et al.* 2013). Also, localized and persistent sources may exist and can contaminate the cross-correlations. This is the case of the 26-s microseismic source observed in Holcomb (1980), Shapiro *et al.* (2006), Xia *et al.* (2013) and Seydoux *et al.* (2016b). The effect of non-isotropically distributed noise sources have been discussed in Tsai (2009), Weaver *et al.* (2009), Froment *et al.* (2010), Kimman & Trampert (2010), Tian & Ritzwoller (2015) and Fichtner *et al.* (2016). They show that anisotropic seismic source distributions contribute to produce spurious arrivals or asymmetry of the cross-correlations functions, resulting in a bias of the travel-time measurements.

Pre-processing techniques therefore are usually applied to the seismic data in order to reduce the effect of noise anisotropy and strong events influence. First, the data time-segments that contain strongly coherent signals are clipped above a given threshold or down-weighted (e.g. Sabra *et al.* 2005a; Hillers *et al.* 2014; Wu *et al.* 2016). These data time-segments can even be rejected (e.g. Rhie & Romanowicz 2004; Gouédard *et al.* 2008; Seats *et al.* 2011) which results in a decrease of exploitable data. Also, the individual traces are often equalized in both spectral and temporal domains as described in Bensen *et al.* (2007). This attenuates the strong directional sources. Nevertheless, even if these pre-processing techniques strongly improve the seismic wavefield isotropy, some anisotropic features of the seismic wavefield may remain in the cross-correlations, as it has been recently discussed in Seydoux *et al.* (2016a,b). In particular, earthquakes and strong microseismic events most of the time resist to this amplitude equalization process which is performed on individual traces.

More recently, array-based filtering techniques have been developed in order to enhance the recovery of Green's functions from the cross-correlations (Gallot *et al.* 2012; Leroy *et al.* 2012; Menon *et al.* 2012; Carrière *et al.* 2013). A passive inverse filter was proposed in Gallot *et al.* (2012) to enhance the isotropy of surface waves in the ambient seismic noise, and the results are compared with an active seismic experiment. In ocean acoustics, Leroy *et al.* (2012) proposed to improve the emergence rate of Green's function with a spatio-temporal filter built up from the first singular vectors of a reference covariance matrix. Finally, Carrière *et al.* (2013) proposed a spatial-filter approach based on an angular filtering of volume noise using least-square truncated pseudo-inversion of a steering matrix. This last study allowed to filter unwanted directive sources from the cross-correlation, such as shipping-related signals in underwater acoustics, but cannot be used automatically because the filter has to be designed with respect to the source back-azimuth.

Our approach is inspired by Gallot *et al.* (2012) and Carrière *et al.* (2014), and is directly based on pre-processing the covariance matrix computed from array seismic data. This is an advantage compared with Gallot *et al.* (2012), because it can be applied to a pre-existing cross-correlation database. We propose to extend the widely used spectral equalization process presented in Bensen *et al.* (2007), which is applied to the frequency spectrum of individual seismograms, to the equalization of the covariance matrix spectrum (i.e. the ordered eigenvalues of the covariance matrix). Because the entries of the covariance matrix have been measured at different spatial locations, we interpret our approach as a *spatial* equal-

ization process. This method allows to correct for the wavefield anisotropy in two ways: (1) the influence of strong directive sources is substantially attenuated, and (2) the weakly excited modes are reinforced, allowing to partially recover the conditions that ensure a good-quality Green's function retrieval. It presents the advantage to be automatic, and does not require the knowledge of the source back-azimuth, contrary to Carrière *et al.* (2014).

In the real data, uncorrelated noise may be present in the covariance matrix. It has been shown in many studies (Bienvenu & Kopp 1980; Wax & Kailath 1985; Gerstoft *et al.* 2012; Seydoux *et al.* 2016a) that the highest eigenvalues of the covariance matrix are related to coherent information, whereas the weakest are related to the uncorrelated noise. Similarly to Gallot *et al.* (2012) and Carrière *et al.* (2014), our technique requires to be performed on a selected subspace of the covariance matrix, in order not to take into account uncorrelated noise (such as self-sensor electronic noise) in the equalization process, which could result in strong artefacts in the cross-correlation functions. In the present paper, we show how the appropriate truncation can be derived from the number of degrees of freedom of the seismic wavefield from Walsh & Vekua (1969).

Finally, we present an eigenvector-based spatial filter useful to distinguish between surface and body waves. This filter is to be used together with the equalization of the eigenvalue spectrum. We apply it in the case of the equalization of earthquake-related signals when high-amplitude body-waves are present, and show that we can remove them from the cross-correlation functions.

The paper is organized as follows: Section 2 presents the theoretical statements used to properly explain our approach. The equalization of the covariance matrix spectrum is defined therein. Section 2 also presents the theoretical derivation of the eigenvalue cut-off used in the equalization process, and with a more detailed derivation presented in Appendix A. Section 3 introduces the seismic array used in both numerical and real experiments. The method is then illustrated with a synthetic example in Section 4 where a source embedded in idealistic isotropic noise is equalized. In Section 4, we also investigate the effectiveness of our method with using 2-D numerical experiments with heterogeneous velocity model. We show that our technique allows to substantially reduce the impact of a strong seismic source embedded in isotropic seismic noise in the cross-correlations. In particular, we investigate the effectiveness of our approach. To that end, we quantify the traveltime measurement error remaining after having applied the method. Finally, we apply the technique to the data recorded USArray during 2010 in Section 5. We focused our analysis on 30 d of continuous data centred around the 2010 M8.8 Maule earthquake (e.g. Delouis *et al.* 2010), and show that the equalization of the covariance matrix eigenspectrum strongly attenuates the effect of the earthquake in the cross-correlations.

## 2 METHOD

### 2.1 Green's function retrieval from ambient noise cross-correlation and array covariance matrix

It is a well-established theory (Weaver & Lobkis 2001; Derode *et al.* 2003; Wapenaar 2004; Gouédard *et al.* 2008) that the correlation  $\mathcal{R}_{ij}(\tau)$  of the ambient noise recorded at two sensors located at positions  $\mathbf{r}_i$  and  $\mathbf{r}_j$  leads to

$$\frac{d\mathcal{R}_{ij}(\tau)}{d\tau} = \frac{\sigma^2}{4a} (G_{ij}(-\tau) - G_{ij}(\tau)), \quad (1)$$

where  $\tau$  is the correlation time,  $\sigma$  is the noise variance, and where  $a$  is the attenuation of the medium. Here,  $G_{ij}(\tau)$  is the time-domain Green's function between the two sensors, that is, the signal that would be recorded at sensor  $j$  if an impulsive source were located at sensor  $i$ .

The theorem presented in eq. (1) holds when the noise is generated by uncorrelated sources evenly distributed in the medium (e.g. Wapenaar 2004). The statistical definition of the cross-correlation between signals  $u_i(t)$  and  $u_j(t)$  recorded by two sensors  $i$  and  $j$  is given by

$$\mathcal{R}_{ij}(\tau) \equiv \mathbb{E}[u_i(t)u_j(t - \tau)], \quad (2)$$

where  $\mathbb{E}$  denotes expected value. Often, eq. (2) is computed in the frequency domain,  $R_{ij}(\tau) = \mathcal{F}^{-1}[\mathcal{C}_{ij}(f)]$ , from the array covariance matrix  $\mathcal{C}_{ij}(f)$  defined as

$$\mathcal{C}_{ij}(f) \equiv \mathbb{E}[u_i(f)u_j^*(f)], \quad (3)$$

where  $u_i(f) = \mathcal{F}[u_i(t)]$  is the Fourier transform of the seismogram recorded in station  $i$  and  $*$  represents complex conjugation. The covariance matrix is a square  $N \times N$  matrix, where  $N$  is the total number of seismic sensors of a given seismic array.

## 2.2 Estimation of the array covariance matrix

The array covariance matrix is estimated from the array data vector  $\mathbf{u}(f, t)$  such as

$$\mathbf{u}(f, t) = [u_1(f, t), u_2(f, t), \dots, u_N(f, t)]^T, \quad (4)$$

where  $u_i(f, t)$  is the Fourier transform of the data time-segment recorded by the station  $i$  of duration  $\delta t$  and starting at time  $t$ . We then estimate the array covariance matrix from a finite number  $M$  of time segments, that is, over a total duration  $\delta t(M - 1)/s$ , where  $s$  is the overlapping ratio of each time-segments

$$\mathbf{C}(f, t) = \frac{1}{M} \sum_{m=0}^{M-1} \mathbf{u}(f, t + m\delta t) \mathbf{u}^\dagger(f, t + m\delta t). \quad (5)$$

The estimated array covariance matrix  $\mathbf{C}$  differs from the theoretical array covariance matrix  $\mathcal{C}$  because it estimates the statistical relationship between the entries of the array data vectors from a finite number of windows. It converges towards  $\mathcal{C}$  as the number of subwindows  $M \rightarrow \infty$ . We can therefore define a convergence criterion to ensure the estimated covariance matrix to be close enough to the theoretical covariance matrix. As a rule of thumb, many studies claim that having  $M$  number of estimating time-segments larger than three times the number of array elements  $N$  might be sufficient to properly estimate the array covariance matrix from real data (e.g. Menon *et al.* 2012). In the present study, we consider the case when the covariance matrix is estimated with  $M \gg N$  (i.e.  $M \approx 10N$ ), in order to avoid any estimation problems.

## 2.3 Equalization of the covariance matrix eigenspectrum

The covariance matrix is a positive semi-definite Hermitian matrix by definition. It can therefore always be diagonalized, such as

$$\mathbf{C} = \mathbf{\Psi} \mathbf{\Lambda} \mathbf{\Psi}^\dagger, \quad (6)$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$  is the diagonal matrix formed by the non-negative eigenvalues  $\lambda_i$  and where  $\mathbf{\Psi}$  is the matrix formed by the corresponding eigenvectors  $\psi_i$ . We propose to construct

the covariance matrix  $\hat{\mathbf{C}} = \mathbf{\Psi} \hat{\mathbf{\Lambda}} \mathbf{\Psi}^\dagger$  from its eigenvectors with the equalized eigenvalues  $\hat{\mathbf{\Lambda}}$  defined with

$$\hat{\Lambda}_{ii} = \begin{cases} \lambda_i = 1 & \text{if } i \leq L \\ \lambda_i = 0 & \text{otherwise} \end{cases}, \quad (7)$$

where  $L$  is the eigenvalue cut-off above which we consider the eigenvectors to be noise-related only, similar to Gallot *et al.* (2012). We derive theoretical expression for  $L$  in the next section.

## 2.4 Theoretical derivation of the eigenvalue cut-off

Obviously, the eigenvalue cut-off  $L$  of  $\hat{\mathbf{\Lambda}}$  has to be chosen carefully. Indeed, if it is chosen too small, the wavefield information is missed and a too large cut-off will amplify the influence of electronic noise. The optimal value depends on both the geometry of the seismic wavefield that is considered (surface or volume) and on the physical number of degrees of freedom of this wavefield, which is the number of coefficients that are useful to fully describe the seismic wavefield in an arbitrary basis of functions. This number of coefficients is related to the slowness of the medium  $\gamma$ , to the analysing frequency  $f$  and to the typical radius  $\bar{r}$  of the seismic array, which we define as

$$\bar{r} = \frac{2}{N(N-1)} \sum_{i=1}^N \sum_{j>i}^N r_{ij}, \quad (8)$$

where  $r_{ij}$  is the great-circle distance between sensors  $i$  and  $j$ .

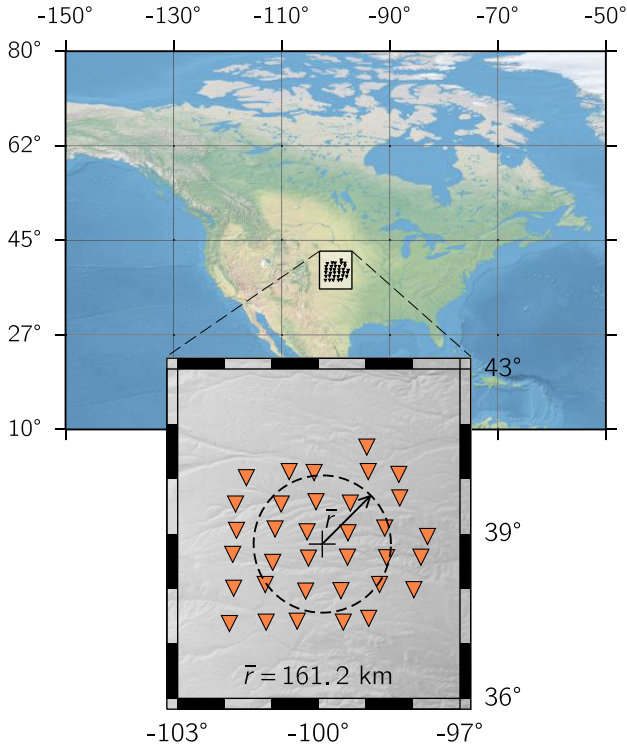
A fully detailed derivation of the eigenvalue cut-off is presented in Appendix A, where we use the cylindrical harmonic decomposition for the 2-D case and the spherical harmonic decomposition in the 3-D case. We then use the theoretical result presented in Moiola *et al.* (2011) in order to count the number of functions that contribute to a full description of the seismic wavefield. This first approach only holds when the wavefield is oversampled by the seismic array (that is when the interstation distance is smaller than half the wavelength). When the wavefield becomes undersampled, we infer in Appendix A that the eigenvalue cut-off should be saturated to  $N/2$  in order to maximize the quality of reconstruction of the covariance matrix. This leads to the following eigenvalue cut-off for the 2-D and 3-D cases:

$$\begin{cases} L^{2D}(f) = \min \{2\lceil 2\pi f \gamma \bar{r} \rceil + 1, N/2\} \\ L^{3D}(f) = \min \{(\lceil 2\pi f \gamma \bar{r} \rceil + 1)^2, N/2\} \end{cases} \quad (9)$$

As different type of waves may exist in an elastic medium, involving different slownesses, we should treat each of these waves separately, that is, we should focus on surface waves or body waves only. Also, for different types of body waves, we should consider only  $P$  or  $S$  waves. We finally observe that the eigenvalue cut-off depends on the slowness, which is supposed to be an unknown quantity. A rough estimation of the average slowness  $\gamma_0$  is nevertheless sufficient to extend our approach to the heterogeneous case.

## 3 US TRANSPORTABLE ARRAY GEOMETRY

The present study investigates the data continuously collected by the vertical channel of the USArray seismic stations during 2010. In order to observe the isotropy of the seismic wavefield, we use the conventional beamforming technique (e.g. Veen & Buckley 1988), that estimates the coherence of the seismic wavefield with respect



**Figure 1.** Seismic stations from the US Transportable Array used in this study. The 34 seismic stations are shown with triangles on the global and on the regional maps. The seismic array aperture is  $\sim 350$  km, and the typical spacing between neighbouring seismic stations is about 55 km. The average interstation spacing  $\bar{r}$  is represented with a vector, and the circle of radius  $\bar{r}$  is shown with black dashed line.

to the slowness and back-azimuth. The beamforming  $\mathbf{B}(f, \boldsymbol{\gamma})$  can be expressed as the projection of the covariance matrix  $\mathbf{C}(f)$  onto plane-wave beamformers  $\mathbf{b}(f, \boldsymbol{\gamma}) = e^{-2i\pi f \boldsymbol{\gamma} \cdot \mathbf{r}}$  as following

$$\mathbf{B}(f, \boldsymbol{\gamma}) = \mathbf{b}^\dagger(f, \boldsymbol{\gamma}) \mathbf{C}(f) \mathbf{b}(f, \boldsymbol{\gamma}), \quad (10)$$

where  $f$  is the frequency,  $\boldsymbol{\gamma} = \gamma_x \mathbf{e}_x + \gamma_y \mathbf{e}_y$  is the 2-D apparent slowness vector in the seismic array plane, and where  $\mathbf{r} = x \mathbf{e}_x + y \mathbf{e}_y$  is the seismic stations coordinate vector. The maxima of the beam power amplitude  $|\mathbf{B}(f, \boldsymbol{\gamma})|^2$  indicates the energy of the incoming waves as a function of the slowness and the back-azimuth encoded in  $\boldsymbol{\gamma}$ .

The beamforming resolution (size of the main-lobe) is inversely proportional to the array aperture (e.g. Veen & Buckley 1988; Rost 2002). If the seismic array shape shows a larger extent in a specific direction, then the slowness resolution is anisotropic. This is particularly inconvenient when our goal is to observe the isotropy of the seismic wavefield, which is to observe the energy of the seismic waves in the slowness space with respect to the azimuth.

We therefore select a nearly square-shaped seismic array formed by 34 seismic stations of the USArray, shown with inverted triangles in Fig. 1. The typical interstation spacing is  $\sim 50$  km and the spatial extent of the seismic array is about 500 km in both east–west and north–south directions. The typical radius  $\bar{r}$  that corresponds to the average interstation spacing of the selected seismic array is 161.2 km, and the corresponding circle is represented with dashed line on Fig. 1.

## 4 NUMERICAL EXPERIMENTS

### 4.1 The homogeneous case

This preliminary test aims at illustrating the method in the simple case where the medium is homogeneous with space-independent slowness  $\gamma_0 = 0.25 \text{ s km}^{-1}$ , and where we set the frequency  $f_0$  to 0.02 Hz in order not to induce aliasing artefacts in the beamforming output. The selected slowness is the typical Rayleigh wave slowness within the Earth around this frequency.

First, we calculate the idealistic covariance matrix  $\mathbf{C}_{\text{ideal}}$  directly from the theoretical expression of the 2-D isotropic seismic noise (e.g. Cox 1974)

$$C_{ij, \text{ideal}} = J_0(2\pi f_0 \gamma_0 |\mathbf{r}_i - \mathbf{r}_j|), \quad (11)$$

where  $\mathbf{r}_i$  and  $\mathbf{r}_j$  are the position vectors for stations  $i$  and  $j$  respectively, so that  $|\mathbf{r}_i - \mathbf{r}_j|$  represents the interstation distance. The eigenvalue spectrum of  $\mathbf{C}_{\text{ideal}}$  is presented in Fig. 2(a). We observe that the eigenvalues are decaying steadily, and vanish with high eigenvalue indexes. This indicates that the number of degrees of freedom of the seismic wavefield  $L^{2D}$  is smaller than the number of sensors  $N$ . Indeed, in this particular case, one gets  $L^{2D} = 2\lceil 2\pi f_0 \gamma_0 \bar{r} \rceil + 1 = 13$  (as illustrated in Fig. A1a), which is in accordance with the obtained eigenvalue spectrum.

The corresponding beamforming obtained from  $\mathbf{C}_{\text{ideal}}$  is presented in Fig. 2(d). We see that the energy is located near the slowness circle of radius  $\gamma_0 = 0.25 \text{ s km}^{-1}$ . This quasi-circular shape reflects that the energy incomes from all azimuths with the same energy. The slight differences in amplitude around azimuths  $45^\circ$  and  $225^\circ$  are due to the non-perfectly uniform shape of the selected USArray stations.

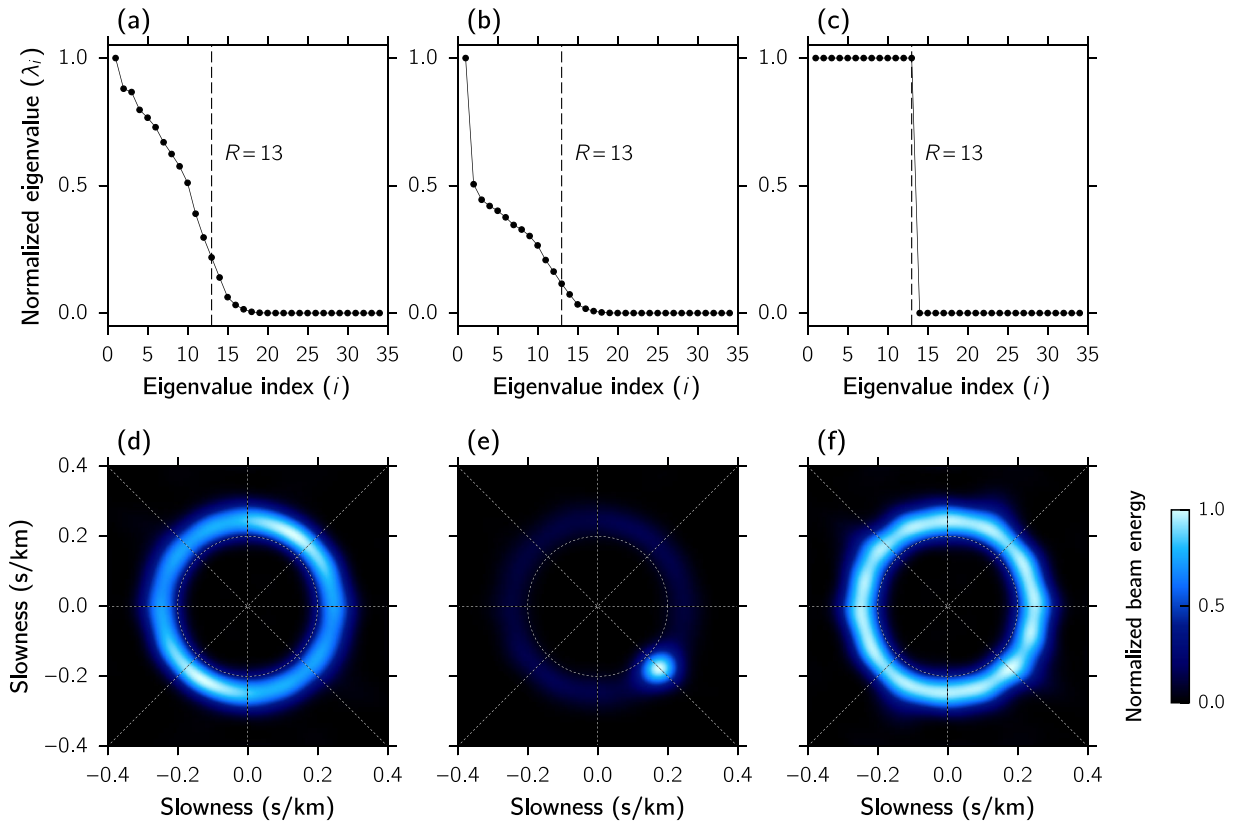
We then calculate a second covariance matrix  $\mathbf{C}_{\text{source}}$ , where we embed a strong source in the aforementioned idealistic 2-D isotropic noise. This second covariance matrix aims at illustrating the case where the seismic data contain an earthquake-related signal, embedded in background seismic noise. We set the amplitude of the source  $a_s = 10$  times higher than the one of the seismic noise. Assuming that both the source and the seismic noise are statistically independent, this second covariance matrix is obtained with

$$\mathbf{C}_{\text{source}} = \mathbf{C}_{\text{ideal}} + a_s^2 \mathbf{s} \mathbf{s}^\dagger, \quad (12)$$

where  $\mathbf{s}$  is the source-related array data vector such as  $s_i = e^{2i\pi f_0 \gamma_0 (\cos \theta x_i + \sin \theta y_i)}$ , and  $\theta$  is the source back-azimuth. We use  $\theta = 135^\circ$  in this first test. The eigenvalue spectrum of  $\mathbf{C}_{\text{source}}$  is presented in Fig. 2(b), where we clearly observe the strong dominance of the first eigenvalue, which is mainly related to the high-amplitude source. The corresponding beam power in Fig. 2(e) exhibits a high-amplitude spot located at azimuth  $135^\circ$  related to the source, and a background circular-shaped pattern related to the idealistic noise.

We finally obtain the equalized covariance matrix  $\mathbf{C}_{\text{equalized}}$  from the truncated summation of the  $L^{2D} = 13$  firsts eigenvectors, with the equalized eigenvalues as visible in Fig. 2(c).

We see in Fig. 2(f) that the amplitude of the source and the one of the background noise have been equalized, and the circular shape recovered. This simple test illustrates our equalization technique, with considering expected covariance matrices of idealistic noise and signal related to a strong far-field single source. One notices that the resulting beamforming energy has been nearly uniformly equalized close to the slowness circle corresponding to  $0.25 \text{ s km}^{-1}$  in Fig. 2(f), and that we do not observe anymore the slight asymmetry due to the arrays shape in the idealistic case (Fig. 2d). This indicates that small errors are induced with the equalization of



**Figure 2.** Spatial equalization of a synthetic source embedded in an idealistic 2-D isotropic seismic noise. The top panels show the covariance matrix eigenvalues, sorted in decreasing order (i.e. the covariance matrix spectra). The corresponding beamforming is presented in each case at the bottom. They are obtained from the covariance matrix in each case with eq. (10). We fixed the slowness  $s_0 = 0.25 \text{ s km}^{-1}$ , and the period  $T = 50 \text{ s}$ . (a,d) Idealistic 2-D seismic noise obtained from eq. (11). (b,e) Idealistic 2-D noise plus a 10 times higher seismic source located at azimuth  $135^\circ$ . (c,f) Case (b/e) where the covariance matrix is reconstructed from its first 11 eigenvectors with the corresponding eigenvalues set to 1.

the covariance matrix spectrum. These errors are investigated in terms of traveltimes measurements from the corresponding cross-correlation in the following section that considers the heterogeneous case.

## 4.2 The heterogeneous case

### 4.2.1 Numerical setup

We perform a 2-D acoustical simulation which models Rayleigh waves propagation only. We used 2-D finite-differences for solving the 2-D wave equation.

The numerical setup of the simulation is presented in Fig. 3(a). In order to use the geometry of the selected USArray seismic stations presented in Fig. 1, we performed the simulation on a  $2500 \times 2500 \text{ km}^2$  grid. We chose a spatial grid step  $d_x = 1.6 \text{ km}$ , and a time step  $d_t = 0.065 \text{ s}$  in order to decrease numerical dispersion. The reflexions of the pressure waves on the borders of the numerical domain are cancelled out using perfectly matched layers derived from Liu (1997) and shown with darker zones in the borders of the spatial domain in Fig. 3(a).

The velocity model is changing linearly in the  $x$ -direction from 2 to 6  $\text{km s}^{-1}$ . It is presented in colour scale in Fig. 3(a) relatively to the average velocity  $v_0 = 4 \text{ km s}^{-1}$ . The positions of the  $N = 34$  seismic stations are shown with blue triangles. In order to simulate noise sources, we used a number of  $S = 200$  seismic sources

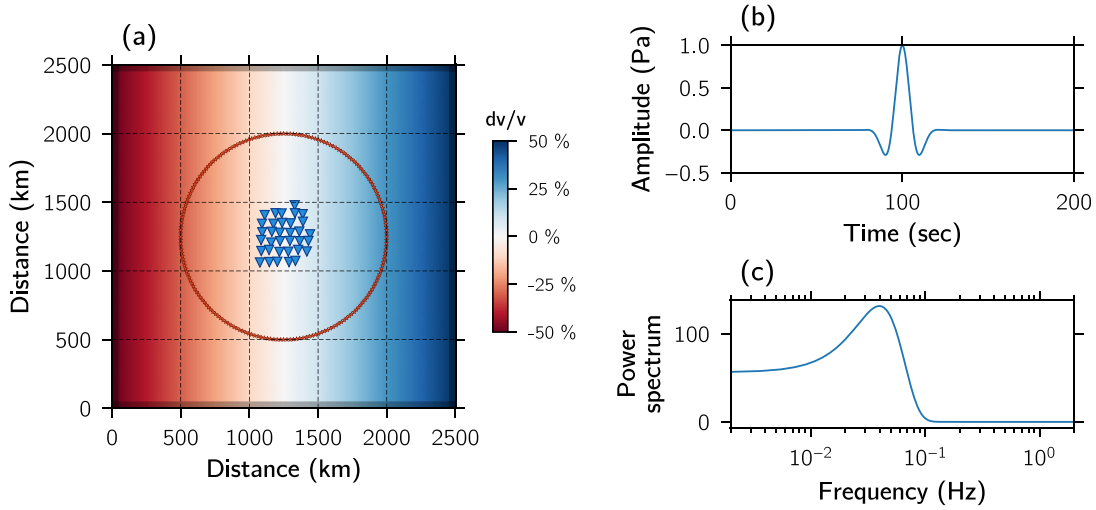
deployed circularly around the seismic array (shown as red stars in Fig. 3a). We consider this number of sources to ensure the sensor-to-sensor cross-correlations computed from numerical noise sources to have a correlation coefficient of more than 90 per cent with the direct Green's functions when assuming sufficient averaging and equal source amplitude (see for instance Derode *et al.* 2003).

### 4.2.2 Protocol: numerical Green's functions retrieval from ambient noise cross-correlation

In order to estimate the numerical Green's function  $G_{ab}(t)$  of the medium between two grid nodes  $a$  and  $b$ , we emit a pulse-like signal  $r(t)$  at a location  $a$  and consider the record of this pulse at any location  $b$  to consist in the Green's function. The pulse-like signal is chosen to be a Ricker wavelet (Ricker 1953):

$$r(t) = (1 - (t - t_0)^2 \beta^2) e^{-(t - t_0)^2 \beta^2} \Leftrightarrow R(f) = R_0 f^2 e^{-f^2 / \beta^2} \quad (13)$$

where  $\beta$  is the frequency bandwidth,  $t_0$  is the time shift of the wavelet, and  $R_0$  is a constant. Eq. (13) indicates that the both temporal and frequency content of the pulse are second derivative of a Gaussian function, which is a symmetric function with zero-mean amplitude. Its half-height width is of  $1/\beta$  in the temporal domain, and of  $\beta$  in the frequency domain. We performed the analysis between 100 and 10 s of period (0.01 and 0.1 Hz), so we chose  $\beta = 0.1 \text{ Hz}$ . The temporal amplitude of the pulse is shown in



**Figure 3.** Numerical setup of the 2-D acoustic finite-differences simulation. (a) Locations of the seismic stations (blue triangles) and of the seismic sources (red stars). The velocity model is represented in colour scale with the reference velocity  $v_0 = 4 \text{ km s}^{-1}$ . (b) Ricker pulse  $r(t)$  used to obtain the source-to-receiver numerical Green's function between the sources and the receivers (see eq. 13) (c) Ricker pulse Fourier transform  $R(f)$  of the pulse signal  $r(t)$  shown in (b).

Fig. 3(b) where  $t_0 = 100 \text{ s}$ , with its corresponding power spectrum in Fig. 3(c). We see that with our choice of parameters, the pulse spectrum is positive from 0.01 to 0.1 Hz, and its energy vanishes beneath  $-150 \text{ dB}$  at  $f > 0.2 \text{ Hz}$ .

We estimate all numerical Green's functions  $G_{is}(t)$  between each sensor  $i = 1 \dots N$  and sources  $s = 1 \dots S$ . Because of the reciprocity of the wave equation, we notice that  $G_{is}(t) = G_{si}(t)$ . Therefore, we compute  $G_{si}(t)$  with emitting the pulse from each seismic stations and record the wavefield at the sources in order to save computational time. (This allows to perform only 34 simulations against 200 if the pulse were sent from the seismic sources.) The numerical cross-correlation  $R_{ij}(t)$  of the signals collected at sensors  $i$  and  $j$  can be derived when the wavefield is generated by  $S$  seismic sources with known source-to-receiver Green's functions  $G_{is}(t)$  and when the sources are emitting signals  $n_s(t)$ :

$$R_{ij}(t) = \mathbf{E} \left[ \sum_{s,s'} (G_{is}(t) \otimes n_s(t)) \otimes (G_{js'}(-t) \otimes n_{s'}(-t)) \right], \quad (14)$$

where  $\otimes$  is the temporal convolution product. The isotropy of the numerical seismic wavefield can be controlled by the amplitude of the signals  $n_s(t)$  emitted by each sensors as detailed in the following.

#### 4.2.3 Numerical Green's functions retrieval from isotropic ambient seismic noise

An isotropic seismic wavefield can be obtained if we impose the seismic sources to emit uncorrelated noise  $n_s(t)$  of equal variances  $v$  from one source to another such as

$$\mathbf{E}[n_s(t) \otimes n_{s'}(-t)] = v^2 \delta_{ss'}, \quad (15)$$

where  $\delta_{ss'}$  is the Kronecker delta, and if the sources are evenly distributed on a surrounding curve around the seismic array, then the expected value of eq. (14) reduces to (omitting the amplitude normalization factor)

$$R_{ij}(t) = \sum_{s=1}^S G_{is}(t) \otimes G_{js}(-t) = -G_{ij}(t) + G_{ij}(-t), \quad (16)$$

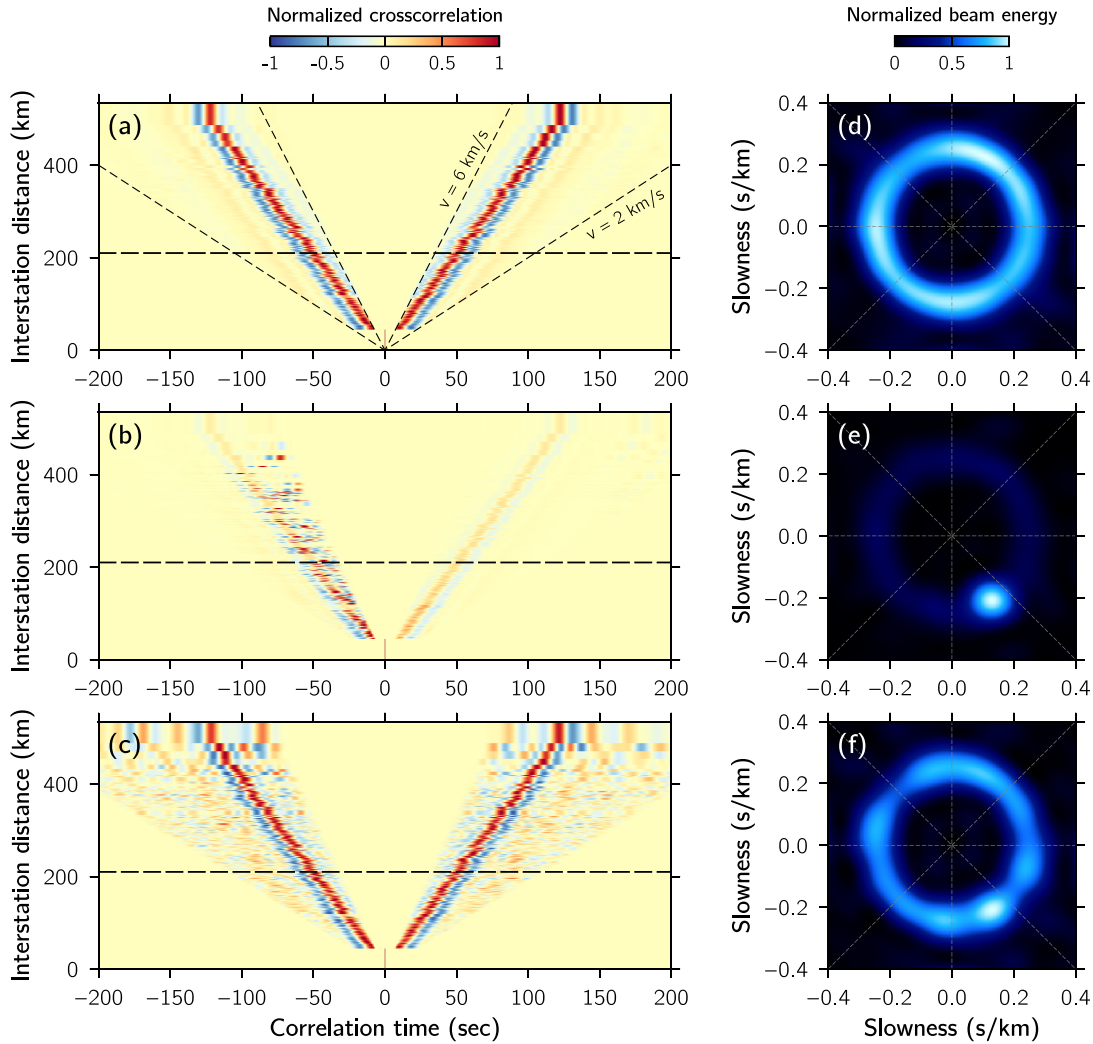
indicating that we retrieved eq. (1). In order to save computational time, we compute the cross-correlation functions directly from eq. (16), implicitly assuming uncorrelated noise sources and sufficient averaging.

The cross-correlations  $R_{ij}(\tau)$  computed between all  $34 \times (34 - 1)/2 = 561$  pairwise combinations of seismic stations are presented in colour scale in Fig. 4(a) sorted with the interstation spacing and against time. In order to automatically look for physically possible arrival times in the next section, we masked the correlation functions for arrival times with velocities higher than  $6 \text{ km s}^{-1}$  and smaller than  $2 \text{ km s}^{-1}$ , which correspond to the extreme values of the velocity model (indicated with dashed lines in Fig. 4a).

In this first experiment, we impose the wavefield to be isotropic. We applied a narrow bandpass-filter between 12.5 and 50 s (0.02–0.08 Hz) to the cross-correlation functions in order to allow for comparison of the results with beamforming analysis that is computed at a single frequency. We observe the 'V' shape formed by the cross-correlation functions, which is a typical expression of the good azimuthal coverage of the seismic sources with equal amplitudes because it induce symmetric interstation cross-correlations functions.

We perform the plane-wave beamforming analysis of the covariance matrix obtained from the inverse Fourier transform of the cross-correlation functions. The beam power obtained at  $T = 54.6 \text{ s}$  (which approximatively corresponds to the centre of the frequency band used to filter the cross-correlations in Fig. 4a) is presented in Fig. 4(d). We notice that the spectral maximum actually is 33 s. However, we chose 54.6 s in order to mitigate aliasing artefacts that contaminate the visibility of the results, without any loss of generality. We observe the typical circular shape, also synonymous with evenly distributed equal-amplitude seismic sources (i.e. it indicates the energy incomes from all directions with the same amplitude). Most of the beam power is nearby the average slowness circle ( $0.25 \text{ s km}^{-1}$ ).

Note that because we set the space-dependant velocity in the medium with a constant volumic mass, the waves have different amplitude with respect to the local velocity. This is naturally resulting from the energy conservation principle. In order to avoid



**Figure 4.** Numerical interstation cross-correlation functions (left) and corresponding plane-wave beamforming (right). The cross-correlation functions are computed from eq. (14), and are bandpass-filtered between 12.5 and 50 s. The beamforming analysis is performed at  $T = 54.6$  s. (a) Isotropic case, where the seismic sources are evenly distributed with respect to the azimuth and have equal amplitude. The limits of the amplitude mask are indicated with black dashed lines. (b) Anisotropic case, where all sources have the same amplitude except one located at azimuth  $158^\circ$  (corresponding to the Maule earthquake azimuth, see Fig. 8). This case simulates a strong source embedded in ambient and isotropic seismic noise. (c) Cross-correlations obtained after equalization of the covariance matrix illustrated in (b) with using eq. (9). (d–f) Beamforming analyses corresponding to cases (a)–(c), respectively.

any amplitude problems relative to the changing velocity in the medium, we normalized the broad-band source-to-receivers Green's functions  $G_{is}(t)$  to their maximal amplitude before performing the experiment.

#### 4.2.4 Numerical Green's functions retrieval from anisotropic seismic wavefield

We perform the same experiment than in the isotropic case, but we rewrite eq. (15) as following

$$\mathbf{E}[n_s(t) \otimes n_{s'}(-t)] = \nu \delta_{ss'} + \nu_0 \delta_{ss_0}, \quad (17)$$

where  $s_0$  is the index of the strong seismic source of amplitude  $\nu_0 \gg \nu$ . This new relationship correspond to a strong seismic source of amplitude  $\nu_0$  embedded in isotropic seismic noise of variance  $\nu$ . If the sources are evenly distributed, which is the case in the present simulation, the source index  $s_0$  is simply related to a given azimuth  $\theta_0$ .

We compute the interstation cross-correlation functions with a source located at azimuth  $\theta_0 = 158^\circ$  (corresponding to the one of the 2010 M8.8 Maule earthquake observed from the USArray, see Fig 8b), and with an amplitude ten times higher than the background isotropic noise (i.e.  $\nu_0 = 10\nu$ ). The results are presented in Fig. 4(b), in colour scale, sorted with the interstation spacing and with respect to time. We see that the 'V' shape related to the background isotropic noise is contaminated with many spurious waves arrivals with smaller traveltimes. These spurious peaks are caused by the strong source, and their traveltimes are shifted relatively to the azimuth between the interstation axis and the direction-of-arrival of the wave  $\theta_0$ .

The corresponding beamforming analysis is presented in Fig. 4(e). We clearly observe the presence of the strong seismic source because the energy is incoming from azimuth  $158^\circ$ . The isotropic seismic noise is still visible in the background, with the typical circular shape with amplitude 10 times smaller than the main spot.

#### 4.2.5 Numerical Green's functions retrieval from the equalized anisotropic seismic waveform

We finally apply our equalization technique to this last covariance matrix built up from isotropic noise plus a strong source. We perform the eigendecomposition at each frequency, and reconstruct the covariance matrix from the 2-D frequency-dependent number of equalized eigenvectors as described by eq. (9). We used  $\gamma_0$  equal to the average velocity in the medium  $0.25 \text{ s km}^{-1}$  in this approach.

The cross-correlations obtained from the equalized covariance matrix are presented in Fig. 4(c), and the corresponding beamforming in Fig. 4(f). We observe that the symmetry of the cross-correlation functions is strongly improved, suggesting that the effect of the coherent source has been substantially attenuated. We notice that the spatial equalization is less efficient for the cross-correlations with large interstation spacing (above 400 km) because their inter-spacing is large compared with the average one  $\bar{r}$ .

The corresponding beamforming shows a clear improvement in comparison with the raw result shown in Fig. 4(e). The circular shape is greatly improved, indicating that we reinforced the weak isotropic background noise. The presence of the seismic source is still visible at the azimuth  $\theta_0 = 158^\circ$ , meaning that the source effect has not been completely removed, in contrast with the method presented in Carrière *et al.* (2013), but its energy has been equalized.

#### 4.2.6 Traveltime measurements

A legitimate question remains: does this equalization process allow to recover the information relative to the velocity in the medium? In order to address this question, we propose to compare the waves traveltimes and to compute the inversion in order to have a look at the recovered velocity model.

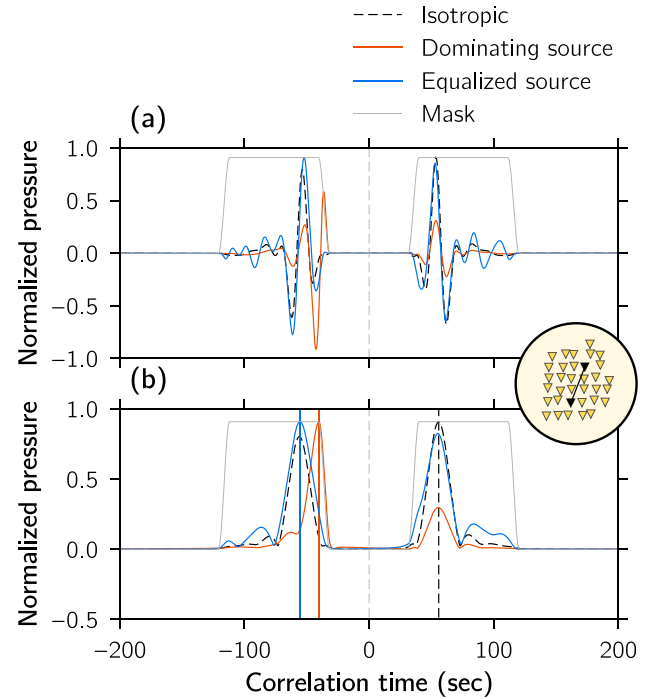
The arrival times are measured from the maximum of the envelope of the cross-correlation functions, which can be obtained from the Hilbert transform. Three selected cross-correlation functions are presented in Fig. 5(a), corresponding to the three cases investigated in Figs 4(a)–(c). The traveltime extraction is shown in Fig. 5(b). The selected sensor pair is highlighted with black triangle in the inset of Fig. 5.

The reference cross-correlation is obtained from the isotropic wavefield configuration. We see that the amplitude is almost symmetric, and that two separate wavepackets have emerged from the cross-correlation with symmetric traveltimes compared to the zero time lag, with almost no energy far from the wavepackets. The envelope maxima lead to the extraction of the waves traveltime between the two considered sensors.

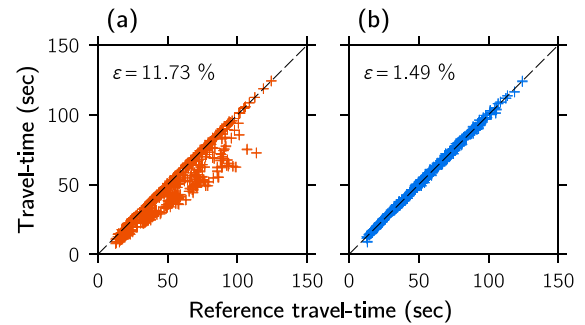
When a single source is dominating, we still observe two small wavepackets that correspond to the isotropic case, plus an additional strong-amplitude peak located near the zero time lag. This peak of amplitude is related to the source amplitude and azimuth, and lead to a biased traveltime estimation (abnormally close to 0).

The cross-correlation obtained from the equalized covariance matrix is closer to the isotropic case. Although the noise is still present, we can observe that the two wavepackets related to the isotropic wavefield have been recovered, whereas the strong-amplitude peak related to the coherent source has been attenuated. The traveltime extracted from the related envelope in Fig. 5(b) is now closer to the reference case, with a slight difference due to the imperfection of the reconstruction.

We systematically investigate the errors in traveltime measurements (Fig. 6) within two cases (1) when the wavefield is contami-

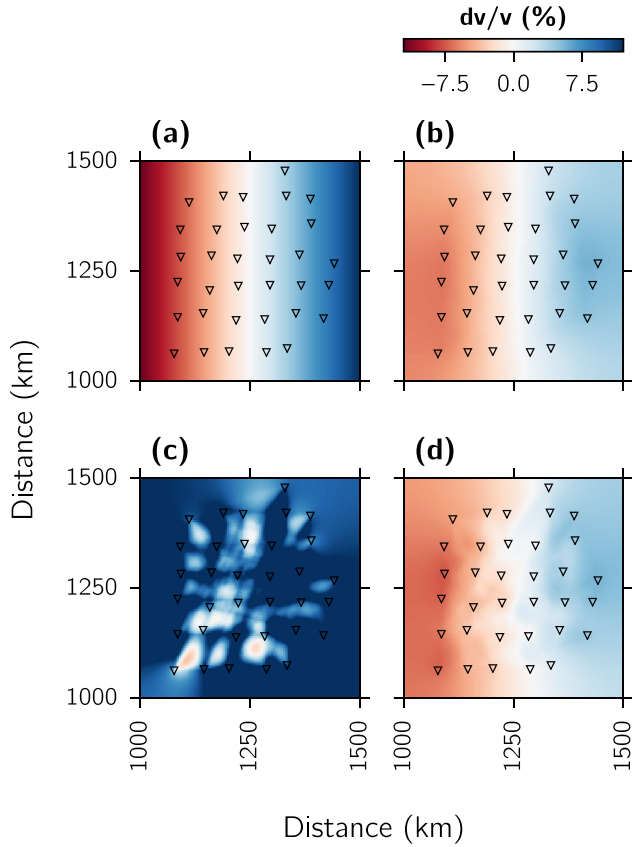


**Figure 5.** Numerical cross-correlation functions obtained at one station pair. The selected station pair is highlighted with black triangles in the inset. The amplitude mask used for this specific station pair is represented with a grey line. (a) Cross-correlation functions. The black dashed line shows cross-correlation obtained from an isotropic distribution of sources. The red line shows the cross-correlation from a distribution with one dominating source. The blue line shows the result of applying the equalization of the covariance matrix spectrum to the anisotropic distribution of sources. (b) Hilbert envelopes of the cross-correlation functions shown in (a). The traveltimes estimated from the envelope maxima are represented with vertical lines.



**Figure 6.** (a) Traveltimes measured from the cross-correlation functions presented in Fig. 4(b), where a strong coherent source at azimuth  $158^\circ$  has an amplitude 10 times higher than the ones compared with ‘reference’ traveltimes measured with an isotropic distribution of sources. (b) Traveltimes measured from the cross-correlation function presented in Fig. 4(d), where the covariance matrix spectrum has been equalized. The average error (eq. 18) in traveltime measurement is indicated in each case.

nated with the strong-amplitude source and (2) with the equalization of the eigenvalues applied to (1). The traveltime measurements are obtained from the time-lag corresponding to the maximum of the envelope, with taking the absolute value in order to get positive traveltimes. The reference traveltimes are obtained from the isotropic cross-correlation functions.



**Figure 7.** Inversion of the traveltimes measured from the numerical cross-correlation function presented in Figs 4(a)–(d). The input velocity model is presented in (a) with respect to the average velocity of the medium  $v_0 = 4 \text{ km s}^{-1}$ . The tomographic results are respectively obtained from (b) an isotropic source configuration, (c) with a strong source illuminating at  $\theta_0 = 158^\circ$  and (d) after the spatial equalization applied on (c).

The overall traveltime measurement error  $\epsilon$  is obtained in both cases from

$$\epsilon = 100 \times \frac{1}{N_c} \sum_{n=1}^{N_c} \frac{|t_{t_n} - t_{t_{ref,n}}|}{t_{t_{ref,n}}} \quad (18)$$

where  $t_{t_n}$  is the traveltime of the  $n$ th cross-correlation function,  $t_{t_{ref,n}}$  is the reference traveltime for the  $n$ th reference cross-correlation, and  $N_c = N(N - 1)/2$  is the total number of cross-correlation functions.

Fig. 6(a) shows that the strong source induces an overall error in the traveltime estimation of 11.73 per cent, with under-estimated traveltimes. This overall error reduces to 1.49 per cent after equalizing the covariance matrix spectrum in Fig. 6(b).

We present the results of the inversion of the traveltimes to recover the velocity model in Fig. 7. The true velocity model is shown in Fig. 7(a). Fig. 7(b) shows the velocity model obtained from the isotropic noise, Fig. 7(c) corresponds to the isotropic case with an additional strong seismic source and Fig. 7(d) is the case (c) where we equalized the covariance matrix spectrum. The inversion is performed using the traveltime measurements between all interstation cross-correlation function, using the open-source Python package *Fatiando* (Uieda *et al.* 2013). We performed the inversion with a 2-D straight-ray model on a  $131 \times 131$  grid, with a regulariza-

tion parameter of 50 km in order to smooth the bias in traveltime measurements from the cross-correlation functions.

The velocity model obtained in case of isotropic noise (Fig. 7a) is the best estimation since it is obtained from a best-case scenario. There are no path coverage outside the box between  $x, y \in [1100, 1400]$  which explains why the inversion leads underestimated velocities on the border of the grid. As expected, the case where a strong source is located in  $\theta_0 = 158^\circ$  lead to strong bias in the traveltime measurements, which induce strong anomalies in the velocity inversion. The colour scale is saturated between -10 per cent and +10 per cent relatively to the average velocity  $v_0 = 4 \text{ km s}^{-1}$ . Finally, the equalization of the covariance matrix spectrum strongly improves the velocity model exactness.

## 5 APPLICATION TO THE DATA COLLECTED BY THE USARRAY: EQUALIZATION OF THE 2010, M8.8 MAULE EARTHQUAKE

### 5.1 Standard cross-correlation in presence of earthquakes

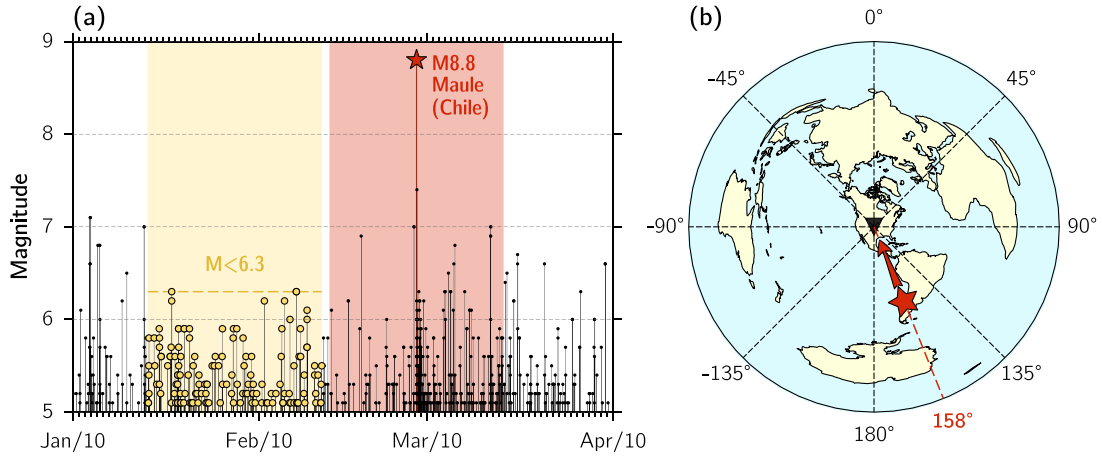
We present in Fig. 8 the magnitudes of the  $M \geq 5$  earthquakes that occurred worldwide between January 1 and April 1 of 2010 obtained from the U.S. Geological Survey search earthquake catalog. We narrowed the focus on two 30 d long time-segments of data indicated with coloured boxes in Fig. 8(a). The yellow box contains no earthquakes with magnitude higher than 6.3, with epicentre locations shown as yellow dots in the azimuthal world map in Fig. 8(b). The red window is centred around the M8.8 Maule, Chile megathrust earthquake (e.g. Delouis *et al.* 2010).

The covariance matrix  $C(f)$  is estimated from the USArray data using eq. (5). The duration  $\delta t$  of the sub-windows is 960 s, and we used a number of  $M = 20$  half-overlapping subwindows (*i.e.* the covariance matrix is obtained within  $\Delta t = 9600$  s long time-segments of data). Before computation, we pre-process the data over each 9600 s long windows with temporal and spectral smoothing techniques described in Bensen *et al.* (2007). We then stack it over 24 hr in order to obtain daily estimates  $C_d(f)$ . The effective number of sub-windows is then  $M_{\text{eff}} \approx 2 \times 20 \times 24 \times 3600/9600 = 360$ , the covariance matrix is thus robustly estimated on each day because  $M_{\text{eff}} \gg N$ , where  $N$  is the number of seismic stations (e.g. Menon *et al.* 2012).

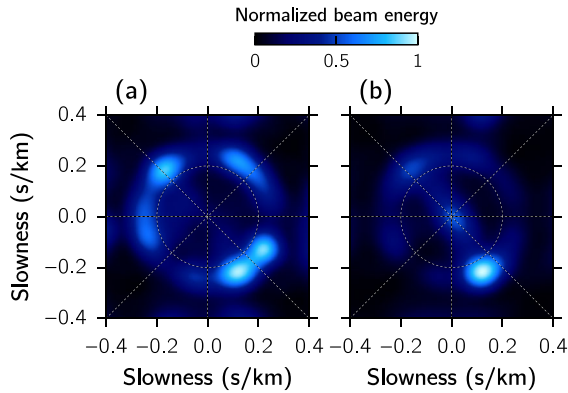
In order to investigate the wavefield isotropy of the two time-segments of data presented in Fig. 8, we stack the daily covariance matrices obtained in each period of time. The planewave beamforming analysis is then performed onto the stacked covariance matrices using eq. (10). We show in Fig. 9 the beamforming obtained within the yellow window (Fig. 9a) and red window (Fig. 9b).

The wavefield shown in Fig. 9(a) is mostly composed of Rayleigh waves (because most of the energy is located near the  $0.25 \text{ s km}^{-1}$  slowness circle) arriving from many different directions, representing a distribution of seismic sources suitable for noise-based tomography.

Fig. 9(b) is obtained from the time-segment when the Maule M8.8 earthquake occurred. Even if the covariance matrix is stacked over 30 d including 15 d before the earthquake nucleation, and even with the additional spectral and temporal pre-processings, we observe a strong dominance of Rayleigh waves generated by this earthquake and its aftershocks (the waves are incoming with an apparent slowness of  $0.25 \text{ s km}^{-1}$  and with about  $160^\circ$  of azimuth). We also



**Figure 8.** (a) Magnitude and date of  $M \geq 5$  earthquakes that occurred in 2010, obtained from the U.S. Geological Survey database (<http://earthquake.usgs.gov/earthquakes/search/>). The coloured boxes indicate two time segments (30 d long each) over which we analysed the wavefield isotropy. The yellow box is free from earthquakes with magnitudes higher than 6.3. The red box contains the Maule M8.8 megathrust earthquake which corresponding epicentre location is shown with red star in (b). (b) Azimuthal global map centred on the USArray barycentre (black triangle). We indicate the direction of arrival of Rayleigh waves incoming from the Maule earthquake.



**Figure 9.** Beamforming computed over the two windows of data shown in Fig. 8. The data were pre-processed with temporal and spectral equalizations. The beamforming output is averaged between 40 and 50 s of period. (a) Yellow window, where no earthquake of magnitude higher than 6.3 occurred (b) Red window, which contains the signal related to the M8.8 Maule earthquake.

observe the presence of body waves with apparent slownesses close to 0.

These results were obtained from the covariance matrix estimated from the data pre-processed with the classical temporal and spectral equalizations. We see that, in agreement with the results obtained in Seydoux *et al.* (2016b), the equalization process does not allow to fully isotropize the seismic wavefield and the presence of spatially localized coherent sources (Fig. 9b) remains clearly encoded in the covariance matrix.

The specific case of the Maule earthquake is herein further investigated because it represents a problematic case for accurate measurements of traveltimes from cross-correlations. In order to show the influence of the earthquake-related highly coherent signals on the cross-correlations, we compute the cross-correlation functions between all pairs of selected stations from the inverse Fourier transform of the covariance matrix, and present them in Fig. 10(a) in colour scale as a function of time and sorted with respect to the interstation spacing. The cross-correlation functions are bandpass-filtered between 12.5 and 50 s. The maximal

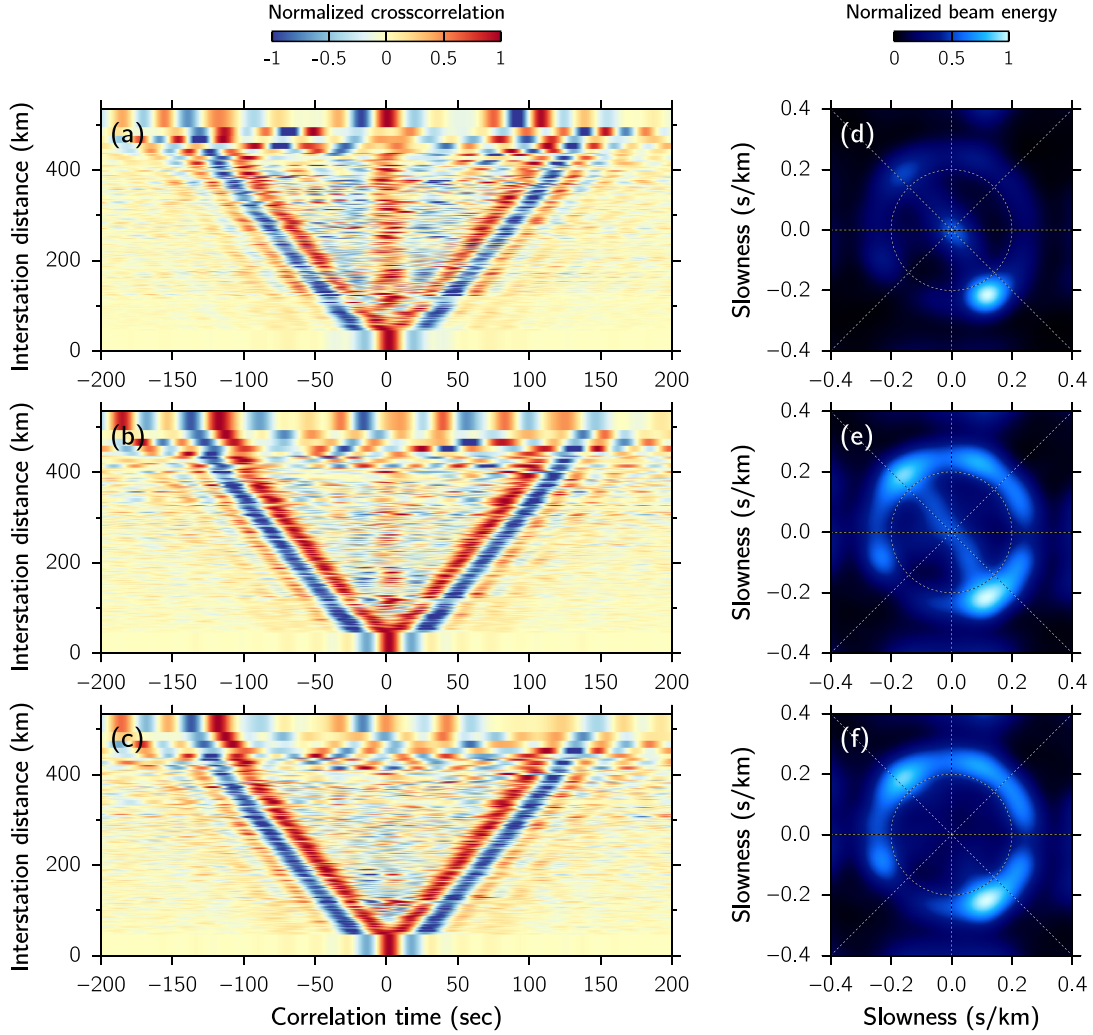
interstation spacing is about 525 km, and the correlation time is shown between  $-200$  and  $200$  s. Also, the corresponding beamforming is presented in Fig. 10(d), averaged between 40 and 50 s of period.

We can see several features in this first result. First, we observe in the background the typical ‘V’ shape formed by the Rayleigh waves due to the ambient seismic noise. The same observation can be done from the corresponding beamforming, with the low-amplitude circular shape at slowness  $0.25 \text{ s km}^{-1}$  in Fig. 10(d). This is an important observation because it allows to consider the seismic wavefield to be generated by seismic sources with a quasi-total azimuthal coverage, with different amplitudes, which is a required condition to apply our method. A second important feature is the strong peak present in many cross-correlations near the zero time-lag. This peak can be both related to (1) the spurious arrivals due to the Rayleigh waves incoming from the Maule earthquake at the pairs of stations which axis is orthogonal to the earthquake azimuth, and (2) to the body waves that income from the earthquake, also responsible for the low-energy peak visible in the beamforming, at apparent slownesses close to 0.

From these observations, we understand that the traveltimes measurements cannot be automatically evaluated from the cross-correlation functions because a lot of them will lead biased traveltimes, similarly than in Fig. 6(a).

## 5.2 Equalization of the M8.8 Maule earthquake

We equalize the eigenvectors of daily covariance matrix (eq. 7), and we then stack the filtered covariance matrices, allowing to increase the performances of the method. Even if the Rayleigh waves are known to be dispersive within the Earth, we use a constant slowness  $\gamma_0 = 0.25 \text{ s km}^{-1}$  (observed on the beamforming at  $T \in [40, 50]$  s of period) in order to define the eigenvalue cut-off  $L(f)$  from eq. (9). A more precise result should be obtained with considering the typical Rayleigh waves slowness with respect to the frequency (dispersive approach). As we want to retrieve the traveltimes of Rayleigh waves, we approximate the problem with a 2-D situation, and choose  $L(f) = L^{2D}(f)$  (eq. 9, Fig. A1b).



**Figure 10.** Spatial equalization of the 2010 M8.8 Maule earthquake. Cross-correlation functions obtained between all the 34 seismic station, with respect to the interstation distance. The cross-correlations are bandpass-filtered between 25 and 100 s. (a) Cross-correlation obtained with the temporal and the spectral normalization technique proposed by Bensen *et al.* (2007). (b) Same as (a) obtained from the equalized covariance matrix. (c) Same as (b) with the additional eigenvector selection. The eigenvectors with energy inside the  $0.15 \text{ km s}^{-1}$  slowness circle are systematically rejected. (d–f) Beamforming analyses of cases (a)–(c), respectively, averaged between 41 and 51.2 s of period.

The interstation cross-correlations obtained from the equalized covariance matrix is presented in Fig. 10(b), and the corresponding beamforming is shown in Fig. 10(e). We clearly observe that the equalization process has increased the energy of Rayleigh waves recovered from the background noise. The strong spot of energy related to the Maule earthquake Rayleigh wave is still visible in the beamforming. However, its amplitude have been quasi-equalized to waves coming from other directions. This is in accordance with the numerical result presented in Section 4 (Fig. 4c).

We also see in Fig. 10(e) a linear feature oriented along the direction corresponding to the Maule earthquake and filling the whole range of slownesses within the Rayleigh wave circle. Because the beamforming only gives access to an apparent slowness, this pattern most likely represents the body waves generated by the Maule earthquake. These body waves have been equalized by our method, because of its inability to distinguish between 2-D and 3-D wavefields. Indeed, we choose the eigenvalue  $L(f) = L^{2D}(f)$ , but as mentioned in Section 2,  $L^{3D} > L^{2D}(f)$ . We therefore see that from its fundamental definition (see eqs 9) the equalization of the covariance matrix spectrum can reinforce body waves if they are present in the

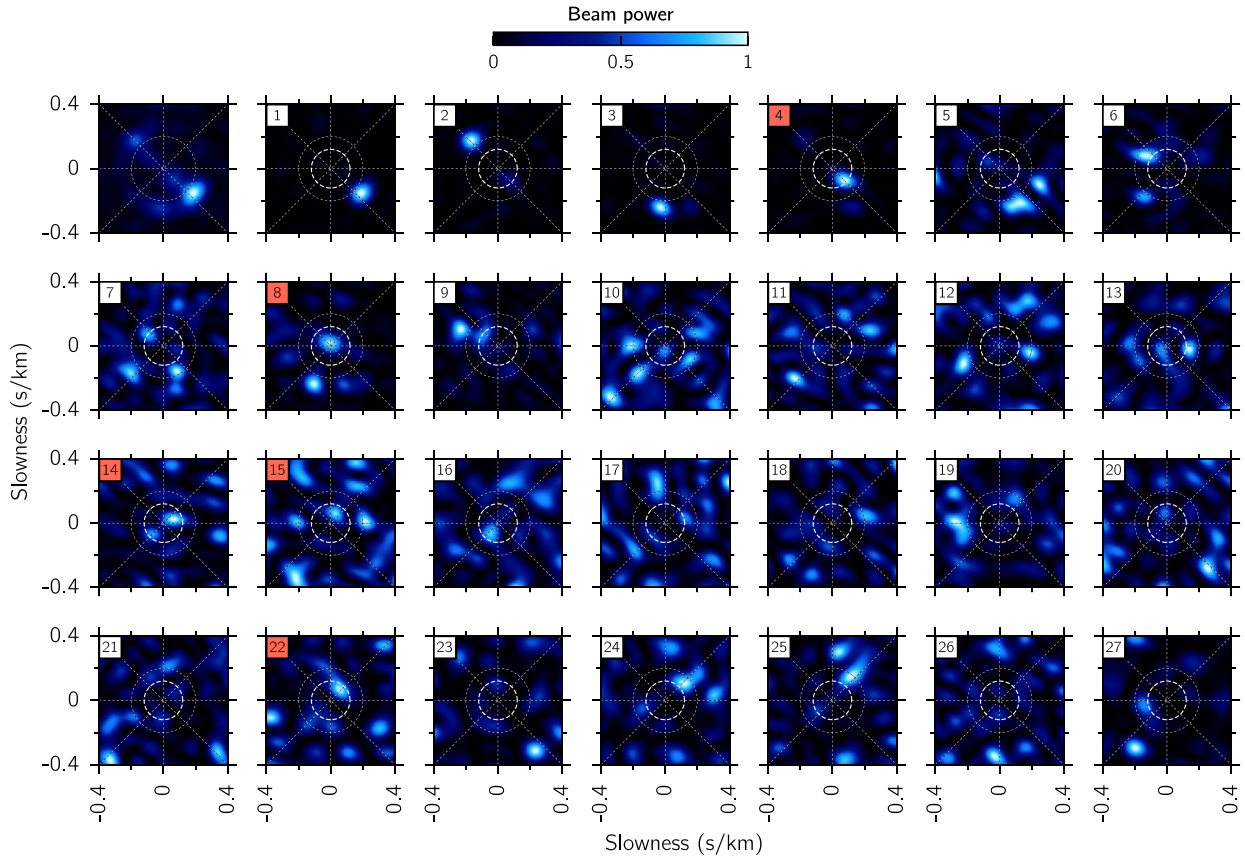
wavefield, and if they are spanned by eigenvectors  $\psi_i$  with index  $i < L^{2D}(f)$ .

### 5.3 Eigenvector-based body wave removal

In order to remove the influence of the unwanted body waves observed in Fig. 10(d), we propose to select surface waves from the covariance matrix eigenvectors before performing the eigenspectrum equalization. The selection can be done with the beamforming analysis, applied to the covariance matrix eigenvectors  $\psi_n$ , with rejecting those that contain energy inside an arbitrary slowness threshold  $\gamma_{\text{thr}}$ . The beamforming  $\mathbf{B}_n$  applied to the  $n$ th eigenvector where  $n = 1 \dots N$  is the eigenvector index is given by

$$\mathbf{B}_n = \mathbf{b}^\dagger \psi_n \psi_n^\dagger \mathbf{b}, \quad (19)$$

where  $\mathbf{b}$  is the beamformer defined in Section 2. The beamforming analyses of the first 27 eigenvectors of the array covariance matrix computed from the Maule earthquake (Fig. 10c) are presented in Fig. 11. Each eigenvector show a specific beam pattern, and some of them clearly contain body waves (for instance, eigenvectors #4,



**Figure 11.** Selection of the covariance matrix eigenvectors that are related to Rayleigh waves. Each plot shows the beamforming performed on one eigenvector of the covariance matrix computed from the M8.8 Maule earthquake. The full-rank beamforming is presented in the top-left plot. In each case, the number corresponds to the eigenvector index. The dashed circle corresponds to the slowness circle of rejection  $\gamma_{thr}$  and the rejected eigenvectors are indicated with red labels.

8, 14, 15, and 22), while the other ones does not. Therefore, we design a rejection criterion upon which the eigenvectors that contain energy inside an arbitrary circle are not taken into account in the equalization procedure.

In the present case, we rejected all the eigenvectors that contain energy inside the slowness circle  $\gamma_{thr} = 0.15 \text{ s km}^{-1}$ . To that end we compare the maximal value of the beamforming outside and inside the slowness circle. If the maximal value inside the slowness circle is greater than 85 per cent of the outside maximal value, we rejected the corresponding eigenvector. The rejected eigenvectors are visible with a red label in Fig. 11. This selection is performed at each frequency.

We perform the equalization of the covariance matrix with the additional eigenvector selection, and present the corresponding cross-correlation functions and beamforming in Figs 10(c) and (f) respectively. We observe that the beamforming pattern related to the body waves that were incoming from the Maule earthquake have been removed. The corresponding peaks visible near the zero time-lag in the cross-correlations have been attenuated as well.

In order to better understand the effects of this equalization process, we present two particular cases in Fig. 12, where the stations pairs have been selected to be parallel and orthogonal to the earthquake azimuth.

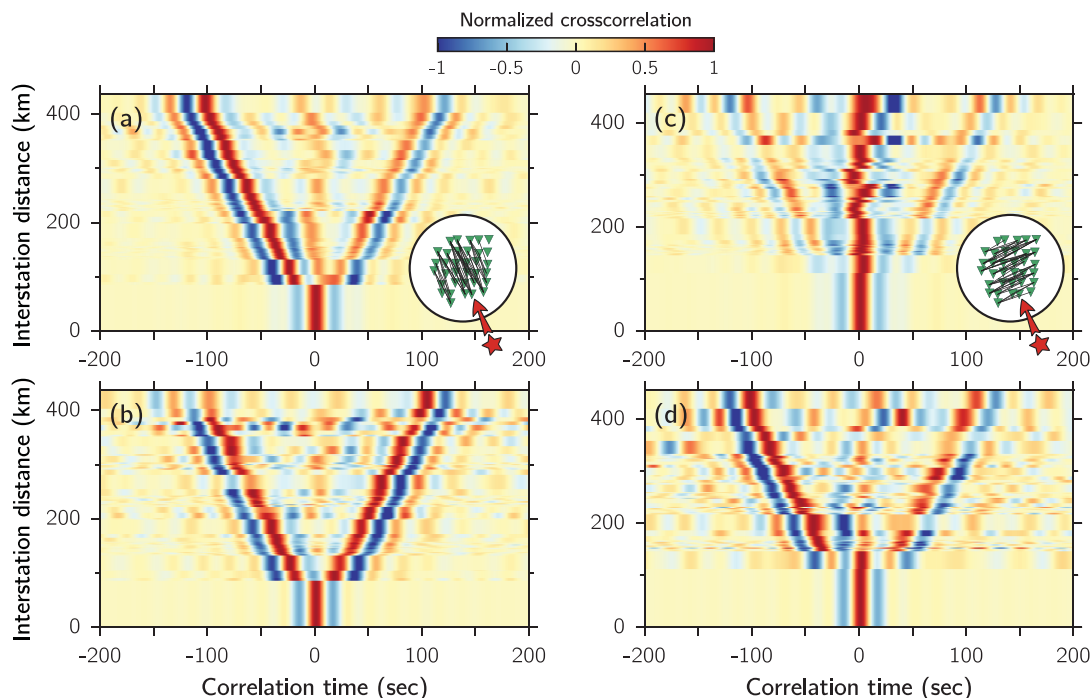
The standard cross-correlations aligned with the earthquake azimuth are presented in Fig. 12(a). As expected, the Rayleigh waves incoming from the earthquake do not induce spurious arrivals in the cross-correlations because they contribute to the construc-

tion of coherent wave fronts that propagate along the path between the two stations. Nevertheless, the amplitude of the cross-correlation functions are asymmetrical because most of the energy incomes from one side of the system. The same result after having applied the spatial equalization and eigenvector selection (in Fig. 12b) shows up to strongly improve the symmetry of these cross-correlations.

The station pairs which axis is orthogonal to the earthquake azimuth without equalization show cross-correlation functions with strong peak near the zero time-lag in Fig. 12(c). This can be simply explained geometrically: the teleseismic waves impinging on the stations from the earthquake arrive at the same moment to both stations of this orientation. This strong peak is almost completely cancelled from cross-correlation function obtained after the equalization and surface wave selection of the covariance matrix (Fig. 12d). We see that the symmetry of the cross-correlation functions is not as good as in the oriented case. However it is now suitable for automatic traveltimes measurements.

## 6 DISCUSSION

Many studies have highlighted the existing analogy between the spatial cross-correlation of a wavefield generated by uncorrelated seismic sources and the time-reversal operation (see for instance Derode *et al.* 2003; Larose 2004; Campillo 2006). Within this analogy, the sources used to perform the time-reversal operation are the



**Figure 12.** Comparison between the standard cross-correlations (a, c) and the cross-correlations obtained after equalizing the covariance matrices and selecting the eigenvectors (b, d). The selected station pairs are parallel (a, c) and orthogonal (c, d) to the earthquake azimuth, with a tolerance angle of  $10^\circ$ , as indicated in the inset in (a) and (b).

analogue of the noise seismic sources of the present study. Indeed, the cross-correlation function between a reference station and all other available stations can be seen as a sequence of (1) recording at the sources a pulse emitted by the reference station, (2) emitting the time-reversed records of the wavefield at each source, leading to (3) a focusing of the waves onto the reference station, followed by (4) a diverging wavefield from the reference station.

If the sources are homogeneously distributed in a surrounding path enclosing the seismic array, then the focused wavefield obtained in step (3) corresponds to the initial pulse emitted by the reference station in step (1) and, consequently, steps (3) and (4) are symmetrical with respect to time-reversion.

When the sources are not homogeneously distributed in the medium, this necessarily leads to asymmetrical cross-correlation functions because the converging (3) and diverging (4) wavefields are not anymore symmetrical with respect to time. In that case, the focusing of the wavefield is degraded, leading to spurious zones where the wavefield focuses outside the reference station location. Consequently, cross-correlation (or similarly time-reversal operation) cannot be directly used to infer the properties of the medium when the sources are not homogeneously deployed around the studied zone.

In such ill-conditioned context, Tanter *et al.* (2000) and Aubry *et al.* (2001) have proposed to improve the focusing of ultrasonic beams in heterogeneous media with replacing the time-reversal operation by an inverse-filter technique that imposes the wavefield to be 1 at the precise location of the reference station and 0 elsewhere. This approach, that can be applied when the sources are under control, leads to selecting proper eigenvectors of the matrix formed by the signals received at the time-reversal sources. Indeed, only a portion of these eigenvectors contain information about the propagating waves (forming the signal subspace), whereas the other ones are related to uncorrelated noise (i.e. the noise subspace) which is likely

to produce strong artefacts in this ill-conditioned time-reversal operation. By selecting only the signal-related eigenvectors, Tanter *et al.* (2000) and Aubry *et al.* (2001) have shown that the focusing is improved and the converging and diverging wavefield are symmetrized. This operation can therefore be interpreted as a projection of the recorded wavefield onto the signal subspace (Prada & Thomas 2003). This signal subspace is useful in many applications such as high-resolution beamforming techniques (Wax *et al.* 1984; Goldstein & Archuleta 1987).

The application of the inverse filter to a passive case (i.e. where the sources are not under control) has been first done in Gallot *et al.* (2012) with seismic data. Instead of applying the eigenvalue decomposition onto the matrix formed by the signals received at the sources (consequently unknown), the authors assume the ambient seismic wavefield to be generated by a great number of overlapping Green's functions which are likely to be captured in several arbitrary-defined windows of data. The authors apply the singular-value decomposition directly onto the matrix formed by the signals received at the seismic stations (data matrix). Despite showing similar results with our technique, their method involves to keep the whole  $N \times M$  data matrix where the number of time segments  $M$  is often large, and to perform the singular-value decomposition on it, inducing time- and memory-consuming computation. Neither can this technique be applied to pre-existing cross-correlation databases.

The method presented in our study is equivalent to the passive-inverse filter proposed in Gallot *et al.* (2012). In our case, we directly select the signal-related eigenvectors from the  $N \times N$  array covariance matrix, inducing a non-negligible gain in computation compared to Gallot *et al.* (2012), and making the method applicable to pre-existing cross-correlation databases. The equalization of the covariance matrix eigenspectrum therefore follows the idea of improving the focusing of the cross-correlated wavefield, and acts at every station of the seismic array simultaneously.

The equalization of the covariance matrix eigenspectrum can also be seen as the extension of the spectral equalization process (Bensen *et al.* 2007)—which operates on the Fourier spectrum of the individual seismic traces—to the spatial domain. Indeed, each row of the covariance matrix represents the correlated wavefield amplitude as a function of space. If we consider an infinite array with regularly spaced elements, the covariance matrix becomes block-Toeplitz (Gerstoft *et al.* 2012), meaning that each row of the covariance matrix is a shifted version of the preceding one. In that case, the eigenvalues of the covariance matrix are given by the Fourier transform of one single row, which can be interpreted as a spatial Fourier transform of the correlated wavefield.

We also presented a theoretical derivation of the eigenvalue threshold (Section 2.4, eq. 9 and Appendix A). Our result is similar to the one obtained in Carrière *et al.* (2013) for a volume noise geometry, where the authors considered a planar array with regularly spaced sensors. In the present study, we additionally derived the eigenvalue cut-off as a function of the frequency for a surface noise geometry. Another theoretical derivation of the eigenvalue cut-off was also derived from the random matrix theory in Menon *et al.* (2012) under regularly spaced linear array assumption, where the number of eigenvalues related to the seismic noise is obtained from an asymptotic density of probability.

Finally, our method can also be compared with the one presented in Menon *et al.* (2012), which is designed to reduce the influence of highly coherent sources in ocean acoustics such as passing ships. These authors developed an approach using the random matrix theory to infer the ship-related eigenvalues, and rejected the eigenvalues according to a sequence of hypothesis testing. The azimuthal zones where the ships are located therefore suffer from a lack of energy after filtering, involving the needs of records of the wavefield before and after the ship passes. Our method presents the advantage to keep the energy related to anomalously strong sources, and just reduces their respective contributions by equalizing them with respect to average background sources.

## 7 CONCLUSIONS

We have developed an approach useful to enhance Green's function recovering from ambient seismic noise in presence of strong seismic sources. Our approach is inspired by the passive inverse filter presented in Tanter *et al.* (2001), Aubry *et al.* (2001) and Gallot *et al.* (2012) and by the spatial filter presented in Carrière *et al.* (2014), and can be seen as an extent of the spectral equalization usually applied to individual seismograms (Bensen *et al.* 2007) to the spatial domain. Theoretical expression for the eigenvalue cut-off which is involved in our method as well as in Carrière *et al.* (2014) is derived, in both 2-D and 3-D noise configurations. We automatically define the eigenvalue threshold of our method as a function of the frequency from this theoretical derivation, without any prior knowledge on the spatial distribution of the noise seismic sources. The technique herein developed has the advantage to be directly performed onto the array covariance matrix, and therefore, it can be applied onto pre-existing cross-correlation databases. Finally, we propose an eigenvector-based spatial-filter to be combined with the inverse-filter, which consider the selection of the covariance matrix eigenvectors in order to remove unwanted seismic waves such as body waves.

We performed a numerical simulation with using the geometry of the USArray in order to investigate the performance of our method. We simulated the effect of a strong source embedded in isotropic

ambient noise, where the source has an amplitude 10 times higher than the background noise. Strong bias in the traveltime measurements with an average error of 11.73 per cent are observed on the standard cross-correlation computed from this highly anisotropic wavefield configuration. The equalization of the covariance matrix spectrum applied to this last case shows to be a powerful tool to attenuate the effect of the strong source. The source amplitude is strongly attenuated while the background isotropic noise is reinforced, leading to significantly better traveltime measurements, with overall errors of about 1.5 per cent.

Also, we see that our method requires the knowledge of the medium slowness. Even if the medium is inhomogeneous, we herein considered the average slowness. This is likely to produce artefacts in the filtered cross-correlations, because the reconstruction from too many Bessel functions implies undesired high-frequency fluctuations. A good improvement of this approach would be obtained with using a frequency-dependent slowness model, in order to improve the accuracy of the cut-off. Also, the definition of the average radius  $\bar{r}$  is important. We here decided to take the average interstation spacing of all possible station couples (eq. 8). This is likely to mainly improve the cross-correlation symmetry for the station pairs with interstation spacing close to  $\bar{r}$ , whereas the ones with large interstation spacing could be partially degraded.

We have shown that the cross-correlation functions obtained from the data pre-processed with temporal and spectral equalization (Bensen *et al.* 2007) still suffers from the presence of strong earthquakes with both spurious coherent arrivals and asymmetry of the causal and anti-causal parts. We equalized the covariance matrices eigenspectra of 30 d of data recorded by the USArray, centred around the date of the 2010 M8.8 Maule earthquake. This shown up to strongly reduce both spurious coherent arrivals and the asymmetry of the cross-correlation.

We also proposed a beamforming-based selection of eigenvectors in order to cancel out body waves. Using beamforming analysis on eigenvectors, we reject the eigenvectors that contain strong energy with slownesses corresponding to body waves. This eigenvector selection, combined together with the equalization of the covariance matrix spectrum strongly improve the cross-correlation symmetry and reduces the spurious arrivals caused by strong earthquakes.

We believe this approach is strongly attractive because it allows to perform the cross-correlation analysis even when the wavefield does not present the required conditions that ensure the Green's function retrieval. The covariance matrix eigenspectrum equalization method presented in our paper helps to extend the applicability of the methods based on correlations of ambient seismic noise to cases with strongly heterogeneous distribution of seismic sources. It might be of particular interest when applied to monitoring seismically active regions, where earthquakes and seismic tremors might act as strong and seismic sources whose influence can be strongly diminished after applying the covariance matrix based equalization.

## ACKNOWLEDGEMENTS

This work was supported by the City of Paris under the program Emergence and by the Russian Science Foundation (grant 14-47-00002) and by LABEX WIFI (Laboratory of Excellence within the French Program Investments for the Future) under references ANR-10-LABX-24 and ANR-10IDEX-0001-02 PSL. Computations in this study were performed using the High-Performance Computing infrastructure S-CAPAD at the Institut de Physique du Globe de Paris. Data from the TA network were made freely available as part of the EarthScope USArray facility, operated

by Incorporated Research Institutions for Seismology (IRIS) and supported by the National Science Foundation, under Cooperative Agreements EAR-1261681 and were downloaded from the IRIS Data Management Center (<http://ds.iris.edu/ds/nodes/dmc/>). The authors would also like to thank Gregor Hillers for the useful comments and discussions.

## REFERENCES

- Aubry, J.-F., Tanter, M., Gerber, J., Thomas, J.-L. & Fink, M., 2001. Optimal focusing by spatio-temporal inverse filter. II. Experiments. Application to focusing through absorbing and reverberating media, *J. acoust. Soc. Am.*, **110**(1), 48–58.
- Behm, M., Nakata, N. & Bokelmann, G., 2016. Regional ambient noise tomography in the eastern alps of Europe, *Pure appl. Geophys.*, **173**, 2813–2840.
- Bensen, G.D., Ritzwoller, M.H., Barmin, M.P., Levshin, A.L., Lin, F., Moschetti, M.P., Shapiro, N.M. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, **169**(3), 1239–1260.
- Bienvenu, G. & Kopp, L., 1980. Adaptivity to background noise spatial coherence for high resolution passive methods, in *ICASSP '80. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 307–310, Institute of Electrical & Electronics Engineers (IEEE).
- Boué, P., Poli, P., Campillo, M., Pedersen, H., Briand, X. & Roux, P., 2013. Teleseismic correlations of ambient seismic noise for deep global imaging of the earth, *Geophys. J. Int.*, **194**(2), 844–848.
- Brenguier, F., Shapiro, N.M., Campillo, M., Nercessian, A. & Ferrazzini, V., 2007. 3-D surface wave tomography of the Piton de la Fournaise volcano using seismic noise correlations, *Geophys. Res. Lett.*, **34**(2), L02305, doi:10.1029/2006GL028586.
- Brenguier, F., Campillo, M., Hadziioannou, C., Shapiro, N.M., Nadeau, R.M. & Larose, E., 2008a. Postseismic relaxation along the San Andreas fault at Parkfield from continuous seismological observations, *Science*, **321**(5895), 1478–1481.
- Brenguier, F., Shapiro, N.M., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O. & Nercessian, A., 2008b. Towards forecasting volcanic eruptions using seismic noise, *Nat. Geosci.*, **1**(2), 126–130.
- Brenguier, F., Campillo, M., Takeda, T., Aoki, Y., Shapiro, N.M., Briand, X., Emoto, K. & Miyake, H., 2014. Mapping pressurized volcanic fluids from induced crustal seismic velocity drops, *Science*, **345**(6192), 80–82.
- Brenguier, F. *et al.*, 2015. Toward 4D noise-based seismic probing of volcanoes: Perspectives from a large-*n* experiment on Piton de la Fournaise volcano, *Seismol. Res. Lett.*, **87**(1), 15–25.
- Bromirski, P.D., Stephen, R.A. & Gerstoft, P., 2013. Are deep-ocean-generated surface-wave microseisms observed on land?, *J. geophys. Res.*, **118**(7), 3610–3629.
- Campillo, M., 2003. Long-range correlations in the diffuse seismic coda, *Science*, **299**(5606), 547–549.
- Campillo, M., 2006. Phase and correlation in ‘random’ seismic fields and the reconstruction of the Green function, *Pure appl. Geophys.*, **163**(2–3), 475–502.
- Carrière, O., Gerstoft, P. & Hodgkiss, W., 2013. Spatial filtering in ambient noise crosscorrelation, in *ICA 2013 Montreal*, pp. 070043–070043, ASA.
- Carrière, O., Gerstoft, P. & Hodgkiss, W.S., 2014. Spatial filtering in ambient noise interferometry, *J. acoust. Soc. Am.*, **135**(3), 1186–1196.
- Cox, H., 1974. Spatial correlation in arbitrary noise fields, *J. acoust. Soc. Am.*, **55**(2), 1289–1301.
- Delouis, B., Nocquet, J.-M. & Vallée, M., 2010. Slip distribution of the February 27, 2010  $M_w = 8.8$  Maule Earthquake, central Chile, from static and high-rate GPS, InSAR, and broadband teleseismic data, *Geophys. Res. Lett.*, **37**, L17305, doi:10.1029/2010GL043899.
- de Ridder, S. A.L., Biondi, B.L. & Clapp, R.G., 2014. Time-lapse seismic noise correlation tomography at Valhall, *Geophys. Res. Lett.*, **41**(17), 6116–6122.
- Derode, A., Larose, E., Tanter, M., de Rosny, J., Tourin, A., Campillo, M. & Fink, M., 2003. Recovering the Green’s function from field-field correlations in an open scattering medium (I), *J. acoust. Soc. Am.*, **113**(6), 2973–2976.
- Duputel, Z., Ferrazzini, V., Brenguier, F., Shapiro, N., Campillo, M. & Nercessian, A., 2009. Real time monitoring of relative velocity changes using ambient seismic noise at the Piton de la Fournaise volcano (la Réunion) from January 2006 to June 2007, *J. Volcanol. Geotherm. Res.*, **184**(1–2), 164–173.
- Ermert, L., Villaseñor, A. & Fichtner, A., 2015. Cross-correlation imaging of ambient noise sources, *Geophys. J. Int.*, **204**(1), 347–364.
- Fichtner, A., Stehly, L., Ermert, L. & Boehm, C., 2016. Generalised interferometry - I. Theory for inter-station correlations, *Geophys. J. Int.*, **208**, 603–638.
- Froment, B., Campillo, M., Roux, P., Gouédard, P., Verdel, A. & Weaver, R.L., 2010. Estimation of the effect of nonisotropically distributed energy on the apparent arrival time in correlations, *Geophysics*, **75**(5), SA85–SA93.
- Gallot, T., Catheline, S., Roux, P. & Campillo, M., 2012. A passive inverse filter for Green’s function retrieval, *J. acoust. Soc. Am.*, **131**(1), EL21–7.
- Gerstoft, P., Menon, R., Hodgkiss, W.S. & Mecklenbräuker, C., 2012. Eigenvalues of the sample covariance matrix for a towed array, *J. acoust. Soc. Am.*, **132**(4), 2388–2396.
- Goldstein, P. & Archuleta, R.J., 1987. Array analysis of seismic signals, *Geophys. Res. Lett.*, **14**(1), 13–16.
- Gouédard, P. *et al.*, 2008. Cross-correlation of random fields: mathematical approach and applications, *Geophys. Prospect.*, **56**(3), 375–393.
- Haned, A., Stutzmann, E., Schimmel, M., Kiselev, S., Davaille, A. & Yelles-Chaouche, A., 2015. Global tomography using seismic hum, *Geophys. J. Int.*, **204**(2), 1222–1236.
- Hillers, G., Campillo, M., Ben-Zion, Y. & Roux, P., 2014. Seismic fault zone trapped noise, *J. geophys. Res.*, **119**(7), 5786–5799.
- Hillers, G., Roux, P., Campillo, M. & Ben-Zion, Y., 2016. Focal spot imaging based on zero lag cross correlation amplitude fields: application to dense array data at the San Jacinto fault zone, *J. geophys. Res.*, **121**, 8048–8067.
- Holcomb, L.G., 1980. Microseisms: A twenty-six-second spectral line in long-period earth motion, *Bull. seism. Soc. Am.*, **70**(4), 1055–1070.
- Jaxybulatov, K., Shapiro, N.M., Koulakov, I., Mordret, A., Landes, M. & Sens-Schönfelder, C., 2014. A large magmatic sill complex beneath the Toba Caldera, *Science*, **346**(6209), 617–619.
- Kimman, W.P. & Trampert, J., 2010. Approximations in seismic interferometry and their effects on surface waves, *Geophys. J. Int.*, **182**(1), 461–476.
- Landès, M., Hubans, F., Shapiro, N.M., Paul, A. & Campillo, M., 2010. Origin of deep ocean microseisms by using teleseismic body waves, *J. geophys. Res.*, **115**, B05302, doi:10.1029/2009JB006918.
- Lanza, F., Kenyon, L.M. & Waite, G.P., 2016. Near-surface velocity structure of Pacaya volcano, Guatemala, derived from small-aperture array analysis of seismic tremor, *Bull. seism. Soc. Am.*, **106**(4), 1438–1445.
- Larose, E., 2004. Imaging from one-bit correlations of wideband diffuse wave fields, *J. appl. Phys.*, **95**(12), 8393–8399.
- Leroy, C., Lani, S., Sabra, K.G., Hodgkiss, W.S., Kuperman, W.A. & Roux, P., 2012. Enhancing the emergence rate of coherent wavefronts from ocean ambient noise correlations using spatio-temporal filters, *J. acoust. Soc. Am.*, **132**(2), 883–893.
- Lin, F.-C., Tsai, V.C., Schmandt, B., Duputel, Z. & Zhan, Z., 2013. Extracting seismic core phases with array interferometry, *Geophys. Res. Lett.*, **40**(6), 1049–1053.
- Liu, Q.-H., 1997. The perfectly matched layer for acoustic waves in absorptive media, *J. acoust. Soc. Am.*, **102**(4), 2072–2082.
- Longuet-Higgins, M.S., 1950. A Theory of the Origin of Microseisms, *Phil. Trans. R. Soc. A*, **243**(857), 1–35.
- Menon, R., Gerstoft, P. & Hodgkiss, W.S., 2012. Asymptotic eigenvalue density of noise covariance matrices, *IEEE Trans. Signal Process.*, **60**(7), 3415–3424.
- Moiola, A., Hiptmair, R. & Perugia, I., 2011. Vekua theory for the Helmholtz operator, *Z. Angew. Math. Phys.*, **62**(5), 779–807.
- 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.

- Mordret, A., Landes, M., Shapiro, N.M., Singh, S.C. & Roux, P., 2014. Ambient noise surface wave tomography to determine the shallow shear velocity structure at Valhall: depth inversion with a neighbourhood algorithm, *Geophys. J. Int.*, **198**(3), 1514–1525.
- Nakata, N., Chang, J.P., Lawrence, J.F. & Boué, P., 2015. Body wave extraction and tomography at long beach, California, with ambient-noise interferometry, *J. geophys. Res.*, **120**(2), 1159–1173.
- Nishida, K., Montagner, J.-P. & Kawakatsu, H., 2009. Global surface wave tomography using seismic hum, *Science*, **326**(5949), 112–112.
- Prada, C. & Thomas, J.-L., 2003. Experimental subwavelength localization of scatterers by decomposition of the time reversal operator interpreted as a covariance matrix, *J. acoust. Soc. Am.*, **114**(1), 235–243.
- Rhie, J. & Romanowicz, B., 2004. Excitation of Earth's continuous free oscillations by atmosphere-ocean-seafloor coupling, *Nature*, **431**(7008), 552–556.
- Ricker, N., 1953. The form and laws of propagation of seismic wavelets, *Geophysics*, **18**(1), 10–40.
- Rost, S., 2002. Array seismology: Methods and applications, *Rev. Geophys.*, **40**(3), 1008.
- Roux, P., 2009. Passive seismic imaging with directive ambient noise: application to surface waves and the San Andreas fault in Parkfield, CA, *Geophys. J. Int.*, **179**(1), 367–373.
- Roux, P., Kuperman, W.A. & the NPAL Group, 2004. Extracting coherent wave fronts from acoustic ambient noise in the ocean, *J. acoust. Soc. Am.*, **116**(4), 1995–1999.
- Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A. & Fehler, M.C., 2005a. Surface wave tomography from microseisms in Southern California, *Geophys. Res. Lett.*, **32**, L14311, doi:10.1029/2005GL023155.
- Sabra, K.G., Roux, P. & Kuperman, W.A., 2005b. Emergence rate of the time-domain Green's function from the ambient noise cross-correlation function, *J. acoust. Soc. Am.*, **118**(6), 3524–3531.
- Seats, K.J., Lawrence, J.F. & Prieto, G.A., 2011. Improved ambient noise correlation functions using Welch's method, *Geophys. J. Int.*, **188**(2), 513–523.
- Sens-Schönfelder, C. & Wegler, U., 2006. Passive image interferometry and seasonal variations of seismic velocities at Merapi volcano, Indonesia, *Geophys. Res. Lett.*, **33**, L21302, doi:10.1029/2006GL027797.
- Seydoux, L., Shapiro, N., de Rosny, J., Brenguier, F. & Landès, M., 2016a. Detecting seismic activity with a covariance matrix analysis of data recorded on seismic arrays, *Geophys. J. Int.*, **204**(3), 1430–1442.
- Seydoux, L., Shapiro, N.M., de Rosny, J. & Landès, M., 2016b. Spatial coherence of the seismic wavefield continuously recorded by the USArray, *Geophys. Res. Lett.*, **43**(18), 9644–9652.
- Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, **31**, L07614, doi:10.1029/2004GL019491.
- Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**(5715), 1615–1618.
- Shapiro, N.M., Ritzwoller, M.H. & Bensen, G.D., 2006. Source location of the 26 sec microseism from cross-correlations of ambient seismic noise, *Geophys. Res. Lett.*, **33**, L18310, doi:10.1029/2006GL027010.
- Stehly, L., Campillo, M. & Shapiro, N.M., 2006. A study of the seismic noise from its long-range correlation properties, *J. geophys. Res.*, **111**, B10306, doi:10.1029/2005JB004237.
- Tanter, M., Thomas, J.-L. & Fink, M., 2000. Time reversal and the inverse filter, *J. acoust. Soc. Am.*, **108**(1), 223–234.
- Tanter, M., Aubry, J.-F., Gerber, J., Thomas, J.-L. & Fink, M., 2001. Optimal focusing by spatio-temporal inverse filter. I. Basic principles, *J. acoust. Soc. Am.*, **110**(1), 37–47.
- Tian, Y. & Ritzwoller, M.H., 2015. Directionality of ambient noise on the Juan de Fuca Plate: implications for source locations of the primary and secondary microseisms, *Geophys. J. Int.*, **201**(1), 429–443.
- Traer, J., Gerstoft, P., Bromirski, P.D. & Shearer, P.M., 2012. Microseisms and hum from ocean surface gravity waves, *J. geophys. Res.*, **117**, B11307, doi:10.1029/2012JB009550.
- Tsai, V.C., 2009. On establishing the accuracy of noise tomography travel-time measurements in a realistic medium, *Geophys. J. Int.*, **178**(3), 1555–1564.
- Uieda, L., Oliveira Jr, V.C. & Barbosa, V.C.F., 2013. Modeling the Earth with Fatiando a Terra, in *Proceedings of the 12th Python in Science Conference (SciPy 2013)*, pp. 96–103, eds van der Walt, S., Millman, J. & Huff, K.
- Veen, B.V. & Buckley, K., 1988. Beamforming: a versatile approach to spatial filtering, *IEEE ASSP Mag.*, **5**(2), 4–24.
- Walsh, J. & Vekua, I.N., 1969. New methods for solving elliptic equations, *Math. Gaz.*, **53**(384), 210.
- Wapenaar, K., 2004. Retrieving the elastodynamic Green's function of an arbitrary inhomogeneous medium by cross correlation, *Phys. Rev. Lett.*, **93**(25), 254301.
- Wax, M. & Kailath, T., 1985. Detection of signals by information theoretic criteria, *IEEE Trans. Acoust. Speech Signal Process.*, **33**(2), 387–392.
- Wax, M., Shan, T.-J. & Kailath, T., 1984. Spatio-temporal spectral analysis by eigenstructure methods, *IEEE Trans. Acoust. Speech Signal Process.*, **32**(4), 817–827.
- Weaver, R.L. & Lobkis, O.I., 2001. On the emergence of the Green's function in the correlations of a diffuse field, *J. acoust. Soc. Am.*, **109**(5), 2410–2410.
- Weaver, R., Froment, B. & Campillo, M., 2009. On the correlation of non-isotropically distributed ballistic scalar diffuse waves, *J. acoust. Soc. Am.*, **126**(4), 1817–1826.
- Wegler, U. & Sens-Schönfelder, C., 2007. Fault zone monitoring with passive image interferometry, *Geophys. J. Int.*, **168**(3), 1029–1033.
- Wu, C., Delorey, A., Brenguier, F., Hadziioannou, C., Daub, E.G. & Johnson, P., 2016. Constraining depth range of S-wave velocity decrease after large earthquakes near Parkfield, California, *Geophys. Res. Lett.*, **43**, 6129–6136.
- Xia, Y., Ni, S. & Zeng, X., 2013. Twin enigmatic microseismic sources in the gulf of Guinea observed on intercontinental seismic stations, *Geophys. J. Int.*, **194**(1), 362–366.

## APPENDIX A: THEORETICAL DERIVATION OF THE EIGENVALUE CUT-OFF

### A1 Eigenvalue cut-off with infinitely dense seismic array

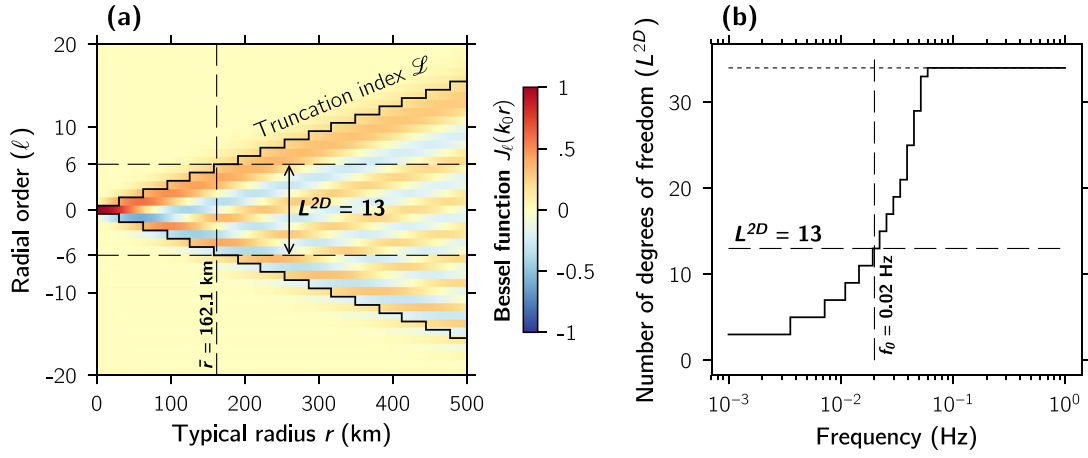
Let us consider the frequency-domain homogeneous wave equation (Helmholtz equation) given by  $(\nabla^2 + k^2)u = 0$ , where  $\nabla^2$  is the Laplace operator and  $k = 2\pi f\gamma$  is the wavenumber that depends on the frequency  $f$  and the slowness  $\gamma$ . In our case, the seismic wavefield is sampled on a typical surface  $\Omega$  covered by the seismic array. Note that we implicitly assume the seismic array to be infinitely dense, the artefacts related to the sampling quality being detailed in the next section. It can be shown (Walsh & Vekua 1969) with considering simple geometry for  $\Omega$ , that only a few Bessel functions contribute to an adequate description of the seismic wavefield. Indeed, we can locally approximate the wavefield  $u_i(f)$  at any point  $\mathbf{r}_i$  with cylindrical coordinates  $(r_i, \theta_i)$  of the medium in the 2-D case (i.e. that considers surface waves) with the following truncated summation

$$u_i(f) \approx \sum_{\ell=-\mathcal{L}}^{\mathcal{L}} \alpha_{\ell} J_{\ell}(2\pi f \gamma r_i) e^{i\ell\theta_i}, \quad (\text{A1})$$

where  $\alpha_{\ell}$  is the  $\ell$ -order coefficient of the wavefield decomposition and  $J_{\ell}(x)$  is the  $\ell$ -order Bessel function of the first kind. A similar approximation is obtained in the 3-D case (i.e. volume noise where a point  $\mathbf{r}_i$  is described with spherical coordinates  $(r_i, \theta_i, \phi_i)$ ) with

$$u_i(f) \approx \sum_{\ell=0}^{\mathcal{L}} \sum_{m=-\ell}^{\ell} \alpha_{\ell m} j_{\ell}(2\pi f \gamma r_i) Y_{\ell m}(\theta_i, \phi_i), \quad (\text{A2})$$

where  $\alpha_{\ell m}$  is the  $\ell m$ -order coefficient of the wavefield decomposition,  $j_{\ell}(x)$  is the  $\ell$ -order spherical Bessel function of the first kind, and  $Y_{\ell m}(\theta, \phi)$  is the spherical harmonic function. The value of the truncation index  $\mathcal{L}$  is explicitly obtained in Moiola *et al.* (2011)



**Figure A1.** Degrees of freedom of the seismic wavefield illustrated with the Fourier–Bessel decomposition (2-D). (a) Bessel functions (in colour scale) of the first kind  $J_\ell(k_0 r)$ , where the wavenumber  $k_0 = 2\pi f_0 \gamma_0$  is obtained with the frequency  $f_0 = 0.02$  Hz and the slowness  $\gamma_0 = 0.25$  s km $^{-1}$ . The truncation index  $\mathcal{L}$  is shown with the black solid line. The vertical dashed line indicates the typical radius  $\bar{r}$  of the USArray geometry investigated in Sections 3–5. The horizontal dashed lines show the upper (lower) limit above (below) which the Bessel functions do not contribute to the wavefield reconstruction. The number of contributing functions  $L^{2D}$  defines the number of degrees of freedom with respect to the given  $k_0$  wavenumber. (b) Frequency dependence of the truncation index  $L^{2D}$  with taking the typical radius  $\bar{r}$ . The dashed lines indicate the case presented in (a). The truncation index is the total number of sensors of the considered seismic array (34 in our case, see the horizontal dotted line) when the frequency is higher than 0.01 Hz.

and is related to the frequency  $f$ , the slowness  $\gamma$  of the medium and the typical radius  $\bar{r}$  of the domain  $\Omega$  as

$$\mathcal{L} = \lceil k\bar{r} \rceil = \lceil 2\pi f \gamma \bar{r} \rceil \quad (\text{A3})$$

where  $\lceil x \rceil$  is the next integer greater than  $x$  and  $\bar{r}$  is the average interstation distance defined in eq. (8).

We see from eqs. (A1) and (A2) that the number of degrees of freedom  $L$  of the wavefield differs in the 2-D and 3-D cases and are given by

$$\begin{cases} L^{2D} = 2\mathcal{L} + 1 \\ L^{3D} = (\mathcal{L} + 1)^2 \end{cases} \quad (\text{A4})$$

We note that the number of degrees of freedom of a volumic wavefield  $L^{3D}$  is greater than the one of a surface noise  $L^{2D}$ . As  $\mathcal{L}$  is frequency-dependent, the truncation index can be explicitly written with respect to the frequency

$$\begin{cases} L^{2D}(f) = 2\lceil 2\pi f \gamma \bar{r} \rceil + 1 \\ L^{3D}(f) = (\lceil 2\pi f \gamma \bar{r} \rceil + 1)^2. \end{cases} \quad (\text{A5})$$

The number of degrees of freedom can be understood more intuitively from the following: if the wavefield is recorded on a limited space region  $\Omega$ , the Bessel functions of rank  $\ell > \mathcal{L}$  do not contribute to the description of the seismic wavefield because  $J_\ell(kr) \approx 0$  when  $r < \bar{r}$ . The corresponding coefficients  $\alpha_\ell$  will therefore be close to 0.

This is illustrated in Fig. A1(a), where the Bessel functions of order  $-20 < \ell < 20$  are shown in colour scale as a function of the radial order  $\ell$  in vertical axis, with respect to the typical radius  $r$  of the domain  $\Omega$ . In the present case, we set  $f = 0.02$  Hz and  $\gamma = 0.25$  s km $^{-1}$ , and we vary the typical radius  $r$  from 0 to 500 km in order to observe the behaviour of the truncation index  $\mathcal{L}$  as a function of the spatial extent of the seismic array (in black solid line). We clearly see that  $\mathcal{L}$  delimits the region where the Bessels functions start to be nonzero anymore.

The number of degrees of freedom  $L^{2D}$  then corresponds to the number of Bessel functions to take into account at the given

frequency, slowness and typical radius. We can examine the particular case of the USArray geometry presented in Section 3, which has a typical radius  $\bar{r} = 162.1$  km. We see that in This case,  $\mathcal{L} = 6$  as  $f_0 = 0.02$  Hz and  $\gamma = 0.25$  s km $^{-1}$ . Using eq. (A4), this leads to  $L^{2D} = 13$ .

We can also observe the variation of  $L^{2D}$  with respect to the frequency in Fig. A1(b) while  $\gamma = 0.25$  s km $^{-1}$  and with the typical radius  $\bar{r} = 161.2$  km of the USArray. We observe that the degrees of freedom of the ambient wavefield is increasing with the increasing frequency, which is an observed behaviour in Gerstoft *et al.* (2012), Menon *et al.* (2012) and Seydoux *et al.* (2016a). We also note that  $L^{2D}$  saturates to  $N$  at frequencies higher than 0.1 Hz, because the number of array elements is limiting the number of degrees of freedom that we can investigate.

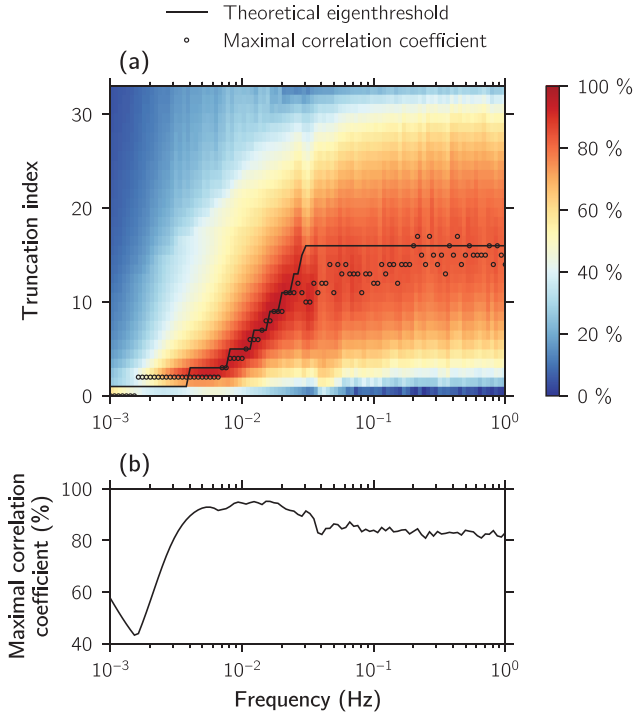
## A2 Eigenvalue cut-off with discrete seismic array

The theoretical rank  $L$  introduced in eq. (A5) tends to infinity with respect to the increasing frequency. This is a natural consequence of the underlying assumption of infinitely dense seismic array, which leads to infinitely large covariance matrix. A discrete seismic array with finite number  $N$  of seismic stations leads to finite-sized  $N \times N$  covariance matrix that can be obviously equalized at maximum with an eigenvalue cut-off  $L = N$ . Yet, taking  $L = N$  induces  $\hat{\mathbf{A}} = \mathbf{I}$ , leading to the following equalized covariance matrix:

$$\hat{\mathbf{C}} = \Psi \hat{\mathbf{A}} \Psi^\dagger = \Psi \Psi^\dagger. \quad (\text{A6})$$

The fundamental property of the eigenvalue decomposition implies that  $\Psi \Psi^\dagger = \mathbf{I}$  because  $\Psi$  is a normal matrix. Saturating  $L$  to  $N$  therefore induces the equalized covariance matrix to be equal to the identity matrix and, as a consequence, to miss out any relevant information spanned by the off-diagonal terms.

In this high-rank regime, we can then wonder if any optimal eigenvalue cut-off can be found. In order to infer it, we investigate the reconstruction quality of the idealist covariance matrix from the equalized anisotropic covariance matrix for various eigenvalue cut-offs  $L$  and with respect to the frequency  $f$ . We use



**Figure A2.** Analysis of the maximal eigenthreshold. (a) Similarity coefficient (in colour scale) between the off-diagonal terms of the ideal covariance matrix  $\mathbf{C}^{\text{ideal}}(f)$  and the equalized anisotropic covariance matrix  $\mathbf{C}^{\text{equalized}}(L, f)$  for various eigenvalue thresholds  $L^{2D}$ . The theoretical eigenthreshold derived from eq. (A9) is indicated in dashed line, and the points for which the maximal coefficient value is obtained are indicated with black solid line. (b) Similarity coefficient for the maximal values extracted in (a).

the synthetic test presented in Section 4.2 where we consider the theoretical expression of the isotropic covariance matrix  $\mathbf{C}^{\text{ideal}}(f)$  in a homogeneous medium (eq. 11), and the anisotropic covariance matrix  $\mathbf{C}^{\text{source}}(f)$  which consist in the isotropic seismic noise plus a planewave with large amplitude (eq. 12). We then equalize the anisotropic covariance matrix at each frequencies with all eigenvalue cut-offs  $L$ , namely  $\mathbf{C}^{\text{equalized}}(L, f)$ .

As a reconstruction quality measurement, we define the similarity coefficient  $s(L, f)$  between the equalized anisotropic covariance matrix and the isotropic covariance matrix such as

$$s(L, f) = \frac{\left| \sum_{i \neq j} \mathbf{C}_{ij}^{\text{ideal}}(f) \mathbf{C}_{ij}^{*\text{equalized}}(L, f) \right|}{\left( \sum_{i \neq j} |\mathbf{C}_{ij}^{\text{ideal}}(f)|^2 \cdot \sum_{i \neq j} |\mathbf{C}_{ij}^{\text{equalized}}(L, f)|^2 \right)^{1/2}}. \quad (\text{A7})$$

Note that we focused onto the off-diagonal terms because they are related to the propagating wavefield, which are of main interested for tomography.

We perform the synthetic test with a slowness  $\gamma_0 = 0.25 \text{ s km}^{-1}$ , and we vary the eigenvalue threshold  $L$  from 0 and  $N = 34$  for the frequencies between 0.001 Hz to 1 Hz. The source has an amplitude 10 times higher than the background noise. The similarity coefficient  $s(L, f)$  is presented in colour scale in Fig. A2(a).

First, we observe that the theoretical eigenthreshold derived from Appendix A1 (shown in black solid line) correlates well with the maximal correlation coefficient (black circles), contributing to support our theory at low frequencies. The similarity then reaches a plateau at high frequencies, close to  $N/2$ . As an indicator, the maximal reachable quality of reconstruction obtained at each maxima shown in Fig. A2(a) is indicated as a function of the frequency in Fig. A2(b).

From these observations, we can finally reformulate eq. (A5) for the 2-D case so that:

$$L^{2D}(f) = \min \{2\lceil 2\pi f \gamma \bar{r} \rceil + 1, N/2\}, \quad (\text{A8})$$

and similarly,

$$L^{3D}(f) = \min \{2\lceil 2\pi f \gamma \bar{r} \rceil + 1, N/2\}. \quad (\text{A9})$$

We consider the frequency-dependent  $L$  as the eigenvalue cut-off in our method. The use of  $L^{2D}$  or  $L^{3D}$  depends on the desired analysis that aims at being performed on the seismic records. We also notice that  $L$  is expressed with the slowness of the medium, which is not precisely known. A rough estimation of the average slowness  $\gamma_0$  is nevertheless sufficient to extend our approach to the heterogeneous case.