

# Application of seismic interferometry by multidimensional deconvolution to ambient seismic noise recorded in Malargüe, Argentina

Cornelis Weemstra,<sup>1,3</sup> Deyan Draganov,<sup>1</sup> Elmer N. Ruigrok,<sup>2,3</sup> Jürg Hunziker,<sup>4</sup> Martin Gomez<sup>5</sup> and Kees Wapenaar<sup>1</sup>

<sup>1</sup>*Department of Geoscience and Engineering, Delft University of Technology, Delft, The Netherlands. E-mail: kweemstra@gmail.com*

<sup>2</sup>*Royal Netherlands Meteorological Institute, De Bilt, The Netherlands*

<sup>3</sup>*Department of Earth Sciences, Utrecht University, Utrecht, The Netherlands*

<sup>4</sup>*Department of Earth Sciences, University of Lausanne, Lausanne, Switzerland*

<sup>5</sup>*International Center for Earth Sciences, Comision Nacional de Energia Atomica, Buenos Aires, Argentina*

Accepted 2016 November 9. Received 2016 November 7; in original form 2016 March 22

## SUMMARY

Obtaining new seismic responses from existing recordings is generally referred to as seismic interferometry (SI). Conventionally, the SI responses are retrieved by simple crosscorrelation of recordings made by separate receivers: one of the receivers acts as a ‘virtual source’ whose response is retrieved at the other receivers. When SI is applied to recordings of ambient seismic noise, mostly surface waves are retrieved. The newly retrieved surface wave responses can be used to extract receiver-receiver phase velocities. These phase velocities often serve as input parameters for tomographic inverse problems. Another application of SI exploits the temporal stability of the multiply scattered arrivals of the newly retrieved surface wave responses. Temporal variations in the stability and/or arrival time of these multiply scattered arrivals can often be linked to temporally varying parameters such as hydrocarbon production and precipitation. For all applications, however, the accuracy of the retrieved responses is paramount. Correct response retrieval relies on a uniform illumination of the receivers: irregularities in the illumination pattern degrade the accuracy of the newly retrieved responses. In practice, the illumination pattern is often far from uniform. In that case, simple crosscorrelation of separate receiver recordings only yields an estimate of the actual, correct virtual-source response. Reformulating the theory underlying SI by crosscorrelation as a multidimensional deconvolution (MDD) process, allows this estimate to be improved. SI by MDD corrects for the non-uniform illumination pattern by means of a so-called point-spread function (PSF), which captures the irregularities in the illumination pattern. Deconvolution by this PSF removes the imprint of the irregularities on the responses obtained through simple crosscorrelation. We apply SI by MDD to surface wave data recorded by the Malargüe seismic array in western Argentina. The aperture of the array is approximately 60 km and it is located on a plateau just east of the Andean mountain range. The array has a T-shape: the receivers along one of the two lines act as virtual sources whose responses are recorded by the receivers along the other (perpendicular) line. We select time windows dominated by surface wave noise travelling in a favourable direction, that is, traversing the line of virtual sources before arriving at the receivers at which we aim to retrieve the virtual-source responses. These time windows are selected through a frequency-dependent slowness analysis along the two receiver lines. From the selected time windows, estimates of virtual-source responses are retrieved by means of crosscorrelations. Similarly, crosscorrelations between the positions of the virtual sources are computed to build

the PSF. We use the PSF to deconvolve the effect of illumination irregularities and the source function from the virtual-source responses retrieved by crosscorrelation. The combined effect of time-window selection and MDD results in more accurate and temporally stable surface wave responses.

**Key words:** Broad-band seismometers; Seismic monitoring and test-ban treaty verification; Surface waves and free oscillations; Interface waves; Seismic attenuation; Seismic tomography.

## 1 INTRODUCTION

Seismic interferometry (SI) refers to the principle of generating new seismic responses from existing recordings. In its simplest form, two receivers are used of which one receiver is turned into a so-called ‘virtual source’ whose response is retrieved at the second receiver. Often, these new responses are obtained by simple crosscorrelation of the seismic observations at the two receiver locations (e.g. Campillo & Paul 2003; Draganov *et al.* 2009; Zhan *et al.* 2010). In case of controlled sources, the process involves an additional summation of crosscorrelations over the available controlled-source positions (e.g. Schuster *et al.* 2004; Bakulin & Calvert 2006). When passive wavefields are exploited, this explicit summation is not required if the simultaneously acting sources are uncorrelated (e.g. Shapiro & Campillo 2004; Wapenaar & Fokkema 2006). We will refer to this, by now conventional way of response retrieval, as ‘SI by crosscorrelation’.

Under specific conditions, responses obtained through SI by crosscorrelation can be related to the Green’s function of the medium. Over the last decade, a myriad of applications have emerged that rely on this explicit relation. The most notable probably being the extraction of receiver-receiver phase and/or group velocities from ambient seismic surface wave noise. The extracted phase velocities often serve as input parameters for tomographic inverse problems (e.g. Shapiro *et al.* 2005). Another application involves the exploitation of the multiply scattered waves present in the later-arriving part of the retrieved responses. These multiply scattered waves are more sensitive to structural changes in the medium and are therefore well suited for seismic monitoring studies (Sens-Schönfelder & Wegler 2006). Seismic monitoring has been applied successfully on various scales (e.g. Obermann *et al.* 2014; Salvermoser *et al.* 2015).

In order to accurately retrieve the Green’s function by means of SI by crosscorrelation, the medium has to be lossless and illuminated with equal power from all directions (Malcolm *et al.* 2004; Wapenaar & Fokkema 2006). The accuracy and quality of the aforementioned applications rely on these conditions being fulfilled. Reformulating the theory underlying SI by crosscorrelation in terms of a multidimensional deconvolution (MDD) process relaxes these conditions (Wapenaar & van der Neut 2010; Wapenaar *et al.* 2011b). Most notably, MDD accounts for directional variations of the wavefield by means of a so-called ‘point-spread function’ (PSF). This PSF captures irregularities in the illumination pattern (van der Neut *et al.* 2011). Multidimensionally deconvolving the responses obtained through SI by crosscorrelation by the PSF corrects for non-uniformities in the illumination pattern. That is, MDD has the ability to remove artefacts from those responses. This implies that it will improve the quality of tomographic images and will lead to seismic-monitoring results with a higher temporal resolution. We will refer to the MDD process as ‘SI by MDD’ in the remainder of this work.

Recently, van Dalen *et al.* (2015) applied SI by MDD to recordings of ambient seismic noise collected as part of the Batholiths experiment. They found that MDD of the responses retrieved through SI by crosscorrelation focuses them in time and increases their signal-to-noise ratio. Furthermore, they found that SI by MDD

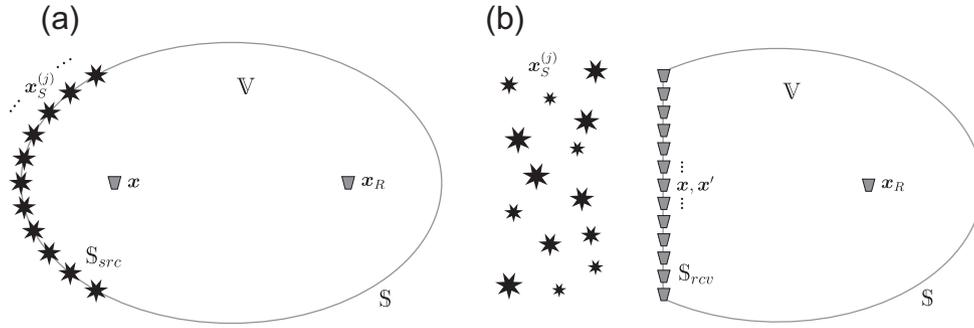
slightly corrects the phase of the retrieved responses at short offsets, resulting in more accurate phase-velocity estimates. The improvements found corroborate earlier numerical experiments considering a subset of USAarray stations illuminated by fundamental-mode surface wave noise excited along the eastern coast of the USA (Wapenaar *et al.* 2011a).

In this study, we use an array of stations located east of Malargüe (Argentina) to test SI by MDD on ambient seismic surface wave noise. The array has an aperture of approximately 60 km and has partly been designed for the application of SI by MDD (Ruigrok *et al.* 2012). The denser receiver spacing of this array with respect to the deployments considered in Wapenaar *et al.* (2011a) and van Dalen *et al.* (2015), allows application of SI by MDD over a larger frequency range. The T-shape of the array lends itself well for the application of SI by MDD: it allows the construction of a PSF along one of the two receiver lines. These receivers act as the virtual sources and their responses are retrieved at the receivers along the other (perpendicular) line of receivers. We utilize recordings of ambient seismic noise between 0.10 and 0.50 Hz; the ambient vibrations in this frequency band are generally referred to as ‘secondary microseisms’. They originate from the interaction of ocean gravity waves with coinciding periods travelling in opposite directions (Longuet-Higgins 1950). The secondary microseisms peak between 0.1 and 0.25 Hz (Gualtieri *et al.* 2013) and are generally the most energetic ambient vibrations measured. Apart from the generation of surface waves, Landès *et al.* (2010) found that the Southern Pacific Ocean is also an effective source of body-wave energy in the frequency band associated with the secondary microseisms.

In the following section, we summarize the theory underlying SI by crosscorrelation and SI by MDD and highlight the differences between the two methods. We furthermore describe how phase velocity may be estimated from the retrieved virtual-source responses. Subsequently, we briefly discuss the geological setting, introduce the Malargüe seismic array (MalARRgüe), and describe the characteristics of its recordings. In Section 4, we combine the configuration of our seismic array with a simple subsurface model. This allows us (i) to introduce our methodology for the selection of time windows during which the wavefield possesses a favourable directionality for the application of SI by MDD, (ii) to test the methodology and, (iii) to exemplify the merits of using SI by MDD. In Section 5, we then apply SI by MDD to the field noise recordings and find that it corrects the phase of the virtual-source responses retrieved through SI by crosscorrelation and show how this leads to improved estimates of phase velocity. Additionally, we compare the temporal stability of the responses retrieved through SI by crosscorrelation with the responses retrieved through SI by MDD. Finally, in Section 6, we discuss our results and draw conclusions.

## 2 THEORY

In this section, we briefly revisit the theory underlying SI by crosscorrelation and SI by MDD. Both techniques allow the retrieval of a medium’s virtual-source response, although SI by crosscorrelation



**Figure 1.** Configuration associated with seismic interferometry by (a) crosscorrelation and (b) MDD. Sources in (a) are regularly placed along  $\mathbb{S}_{src}$  with separation  $\Delta x_S$  and have equal power. Sources in (b) are randomly placed on the appropriate side of  $\mathbb{S}_{rcv}$  and may vary in power.

is more restrictive in terms of medium properties and illumination. SI by MDD is not bounded by most of these restrictions, but is technically more involved and has greater requirements concerning station geometry. Provided these requirements are met, SI by MDD results in a more accurate retrieval of the medium's response.

## 2.1 SI by crosscorrelation

Consider the vertical particle velocity associated with a single surface wave mode propagating along the surface of a laterally homogeneous Earth. In Appendix A we show that in that case the representation theorem of the correlation type can be simplified to

$$\hat{G}(\mathbf{x}_R, \mathbf{x}) + \hat{G}^*(\mathbf{x}_R, \mathbf{x}) \approx \hat{B} \oint_{\mathbb{S}} \hat{G}(\mathbf{x}_R, \mathbf{x}_S) \hat{G}^*(\mathbf{x}, \mathbf{x}_S) d\mathbf{x}_S, \quad (1)$$

where a (frequency-dependent) Green's function  $\hat{G}(\mathbf{x}_R, \mathbf{x})$  represents vertical particle velocity at  $\mathbf{x}_R$  due to a vertical point source at  $\mathbf{x}$ ; both located at the Earth's surface. The asterisk denotes complex conjugation and hence the products on the right-hand side correspond to crosscorrelations in the time domain.  $\hat{B}$  is a simple (frequency-dependent) scale factor.

In practice, one does not record Green's functions due to impulsive sources distributed along a closed contour. Let us therefore consider the vertical (surface wave) particle velocity at  $\mathbf{x}_R$  due to a vertically polarized source at  $\mathbf{x}_S$  with source spectrum  $\hat{s}$ , that is,

$$\hat{v}(\mathbf{x}_R, \mathbf{x}_S) \equiv \hat{G}(\mathbf{x}_R, \mathbf{x}_S) \hat{s}(\mathbf{x}_S), \quad (2)$$

and similarly for  $\hat{v}(\mathbf{x}, \mathbf{x}_S)$ . For comparison with the theory underlying SI by MDD and to comply with the station configuration of the Malargüe seismic array (see Section 3), we consider the one-sided illumination shown in Fig. 1(a). Substituting the particle velocities in eq. (1), and assuming that sources are distributed uniformly along  $\mathbb{S}_{src}$  and have equal power, we find,

$$\hat{G}(\mathbf{x}_R, \mathbf{x}) \hat{S} \approx \hat{B} \int_{\mathbb{S}_{src}} \hat{v}(\mathbf{x}_R, \mathbf{x}_S) \hat{v}^*(\mathbf{x}, \mathbf{x}_S) d\mathbf{x}_S, \quad (3)$$

where  $\hat{S} \equiv \hat{s}(\mathbf{x}_S) \hat{s}^*(\mathbf{x}_S)$ . The one-sided illumination allows retrieval of the so-called 'causal' Green's function only. Because single-mode surface waves travelling along the surface of laterally homogeneous medium are assumed, the responses in eq. (3) represent direct surface wave arrivals only (*cf.* Halliday & Curtis 2008). That is to say, eq. (3) does not hold in the presence of scattering.

The integral in eq. (3) can be discretized considering discrete sources at  $\mathbf{x}_S^{(j)}$  along  $\mathbb{S}_{src}$  with their spectra denoted by  $\hat{s}^{(j)}$  (see

Fig. 1a). Denoting the separation between adjacent  $\mathbf{x}_S^{(j)}$  by  $\Delta \mathbf{x}_S$ , we find,

$$\hat{G}(\mathbf{x}_R, \mathbf{x}) \hat{S} \approx \hat{B} \sum_j \hat{v}(\mathbf{x}_R, \mathbf{x}_S^{(j)}) \hat{v}^*(\mathbf{x}, \mathbf{x}_S^{(j)}) \Delta \mathbf{x}_S. \quad (4)$$

Eq. (4) states that the response at  $\mathbf{x}_R$ , due to a virtual source at  $\mathbf{x}$ , can be retrieved by summing over crosscorrelations of signal emitted by equally powerful sources distributed uniformly along  $\mathbb{S}_{src}$ .

For simultaneously acting (noise) sources, the vertical component of the particle velocities at  $\mathbf{x}_R$  and  $\mathbf{x}$  read

$$\hat{v}(\mathbf{x}_R) \equiv \sum_i \hat{G}(\mathbf{x}_R, \mathbf{x}_S^{(i)}) \hat{s}^{(i)}, \quad (5)$$

and,

$$\hat{v}(\mathbf{x}) \equiv \sum_j \hat{G}(\mathbf{x}, \mathbf{x}_S^{(j)}) \hat{s}^{(j)}, \quad (6)$$

respectively. Under the assumption that these (noise) sources are uncorrelated and their power spectra coincide,  $\langle \hat{s}^{(i)} \hat{s}^{(j)*} \rangle = \hat{S} \delta_{ij}$ , where  $\langle \cdot \rangle$  denotes the ensemble average,  $\delta_{ij}$  is the Kronecker delta, and  $\hat{S} \equiv |\hat{s}^{(i)}|^2$ . By substituting expressions (5) and (6) in

$$\hat{G}(\mathbf{x}_R, \mathbf{x}) \hat{S} \approx \hat{B} (\hat{v}(\mathbf{x}_R) \hat{v}^*(\mathbf{x})) \Delta \mathbf{x}_S, \quad (7)$$

and exploiting the uncorrelatedness of the sources, one finds that this equation coincides with eq. (4).

In practice, ensemble averaging is replaced by averaging over long periods of time and/or large numbers of time windows (e.g. Weemstra *et al.* 2013). At low frequencies (<1 Hz), ambient seismic-field recordings generally consist of a continuous stream of noise associated with ocean-seafloor interaction, interrupted by occasional earthquake events (e.g. McNamara & Buland 2004). In practice, therefore, both eq. (4) and eq. (7) apply when averaging crosscorrelations over a large number of time windows.

Eq. (4) (or eq. (7), in case of simultaneously acting, equally powerful noise sources) is quite accurate in case the medium inside  $\mathbb{V}$  is lossless and homogeneous, and the boundary  $\mathbb{S}_{src}$  is far away (Wapenaar & Fokkema 2006). In practice, however, the (one-sided) illumination is often non-uniform (e.g. Stehly *et al.* 2006; Mulargia 2012) and the medium inside  $\mathbb{V}$  heterogeneous and lossy. This may significantly degrade the accuracy of the retrieved Green's function and may result in erroneous estimates of phase velocity (e.g. Tsai 2009; Weaver *et al.* 2009). SI by MDD permits the correction of deviations of the crosscorrelation from the medium's actual Green's function. It corrects for non-uniformities in the source distribution and/or differences

in the source signatures and accounts for dissipation and scattering inside  $\mathbb{V}$ . Additionally, it allows some degree of correlation between the different sources (Wapenaar *et al.* 2012). SI by MDD, therefore, results in a more accurate response retrieval.

## 2.2 SI by MDD

In Appendix A, we show that under the same assumptions and for the same definitions as the ones required to arrive at eq. (1), the representation theorem of the convolution type can be simplified to

$$\hat{G}(\mathbf{x}_R, \mathbf{x}_S) = \hat{D} \oint_{\mathbb{S}} \hat{G}(\mathbf{x}_R, \mathbf{x}) \hat{G}^{(\text{in})}(\mathbf{x}, \mathbf{x}_S) \, \text{d}\mathbf{x}, \quad (8)$$

where  $\hat{D}$  is another scale factor. The locations  $\mathbf{x}_R$ ,  $\mathbf{x}_S$ , as well as the contour  $\mathbb{S}$ , are all confined to the surface. The absence of a conjugate sign on the right-hand side, implies that the products correspond to convolutions in the time domain. Just like eq. (1), eq. (8) is strictly valid only for a single surface wave mode. In reality, multiple surface wave modes may exist (e.g. Kimman *et al.* 2012). Throughout this work, however, we shall assume that a single surface wave mode dominates the ambient seismic surface wave wavefield.

We note that eq. (8) requires a slightly different interpretation in case the medium is heterogeneous. First, the reference Green's function, which is the Green's function that we aim to retrieve through SI by MDD, is associated with a medium that coincides with the actual (heterogeneous) medium inside  $\mathbb{V}$ , but which is homogeneous on and outside of  $\mathbb{S}$ . This implies that the retrieved Green's function does not include events scattered back into  $\mathbb{V}$ . Second, the amplitude of the Green's function that is retrieved through SI by MDD is scaled by the absolute amplitude of the cosine of the angle between the normal to the surface  $\mathbb{S}_{\text{rcv}}$  and the spatial derivative of  $\hat{G}(\mathbf{x}_R, \mathbf{x})$  (Wapenaar *et al.* 2011b). This scaling coincides with one for the far-field approximation used to derive eq. (A5) (van Dalen *et al.* 2014). In case scatterers exist inside  $\mathbb{V}$ , however,  $\hat{G}(\mathbf{x}_R, \mathbf{x}, t)$  contains a direct arrival as well as scattered events. The angle-dependent scaling of the response retrieved through SI by MDD implies that the direct arrival and each of the scattered events of the retrieved Green's function may be scaled with different scaling factors with respect to their actual amplitude in  $\hat{G}(\mathbf{x}_R, \mathbf{x}, t)$ . In this work we will be mainly concerned with the (accuracy of the) direct surface wave arrival.

We consider the vertical (surface wave) particle velocities at  $\mathbf{x}_R$  and  $\mathbf{x}$  due to a multitude of vertically polarized surface sources at  $\mathbf{x}_S^{(j)}$ . eq. (8) only holds for the inward-propagating part of  $\hat{G}$  and we hence assume the  $\mathbf{x}_S^{(j)}$  to exist on the appropriate side of  $\mathbb{S}$  (see Fig. 1b). Substituting the particle velocities in eq. (8) gives for each source,

$$\hat{v}(\mathbf{x}_R, \mathbf{x}_S^{(j)}) = \hat{D} \int_{\mathbb{S}_{\text{rcv}}} \hat{G}(\mathbf{x}_R, \mathbf{x}) \hat{v}^{(\text{in})}(\mathbf{x}, \mathbf{x}_S^{(j)}) \, \text{d}\mathbf{x}, \quad (9)$$

where we have replaced the closed contour  $\mathbb{S}$  by an open (receiver) contour  $\mathbb{S}_{\text{rcv}}$ , and where  $\hat{v}(\mathbf{x}_R, \mathbf{x}_S^{(j)})$  is defined by eq. (2) (with  $\mathbf{x}_S$  replaced by  $\mathbf{x}_S^{(j)}$ ). Similarly,  $\hat{v}^{(\text{in})}(\mathbf{x}, \mathbf{x}_S^{(j)})$  is defined as the product of  $\hat{G}^{(\text{in})}(\mathbf{x}, \mathbf{x}_S^{(j)})$  and  $\hat{s}(\mathbf{x}_S^{(j)})$ . SI by MDD involves the retrieval of  $\hat{G}(\mathbf{x}_R, \mathbf{x})$  by solving eq. (9).

In practice, multiple sources may be acting simultaneously and their power spectra may differ. We comply with this more general case by considering the total responses instead of the individual

source responses. In this case eq. (9) becomes,

$$\hat{v}(\mathbf{x}_R) = \hat{D} \int_{\mathbb{S}_{\text{rcv}}} \hat{G}(\mathbf{x}_R, \mathbf{x}) \hat{v}^{(\text{in})}(\mathbf{x}) \, \text{d}\mathbf{x}, \quad (10)$$

where  $\hat{v}(\mathbf{x}_R)$  is given by eq. (5) and where  $\hat{v}^{(\text{in})}(\mathbf{x})$  is given by the right-hand side of eq. (6), but with  $\hat{G}(\mathbf{x}, \mathbf{x}_S^{(j)})$  replaced by  $\hat{G}^{(\text{in})}(\mathbf{x}, \mathbf{x}_S^{(j)})$ .

The surface wave response  $\hat{G}(\mathbf{x}_R, \mathbf{x})$  can be retrieved by solving eq. (10) in a least-squares sense, for example. Introducing the auxiliary location variable  $\mathbf{x}'$  along  $\mathbb{S}_{\text{rcv}}$  (see Fig. 1b) and multiplying both sides of eq. (10) with  $\hat{v}^{(\text{in})*}(\mathbf{x}')$ , we obtain the normal equation (van der Neut *et al.* 2011; Wapenaar *et al.* 2011b),

$$\hat{C}(\mathbf{x}_R, \mathbf{x}') = \hat{D} \int_{\mathbb{S}_{\text{rcv}}} \hat{G}(\mathbf{x}_R, \mathbf{x}) \hat{\Gamma}(\mathbf{x}, \mathbf{x}') \, \text{d}\mathbf{x}, \quad (11)$$

where,

$$\hat{C}(\mathbf{x}_R, \mathbf{x}') \equiv \langle \hat{v}(\mathbf{x}_R) \hat{v}^{(\text{in})*}(\mathbf{x}') \rangle \quad (12)$$

and

$$\hat{\Gamma}(\mathbf{x}, \mathbf{x}') \equiv \langle \hat{v}^{(\text{in})}(\mathbf{x}) \hat{v}^{(\text{in})*}(\mathbf{x}') \rangle. \quad (13)$$

We refer to  $\hat{C}(\mathbf{x}_R, \mathbf{x}')$  and  $\hat{\Gamma}(\mathbf{x}, \mathbf{x}')$  as the crosscorrelation function (CCF) and the PSF, respectively. Eq. (11) states that the crosscorrelation function coincides with the sought-for Green's function convolved with the point-spread function. The PSF quantifies the smearing of the virtual source in space and time.

Just as eq. (7) reduces to a single summation if the different sources are uncorrelated (i.e. to eq. 4), so too do eqs (12) and (13). Assuming  $\langle \hat{s}^{(i)} \hat{s}^{(j)*} \rangle = \hat{s}^{(i)} \delta_{ij}$ , we find,

$$\hat{C}(\mathbf{x}_R, \mathbf{x}') \equiv \sum_j \hat{v}(\mathbf{x}_R, \mathbf{x}_S^{(j)}) \hat{v}^{(\text{in})*}(\mathbf{x}', \mathbf{x}_S^{(j)}), \quad (14)$$

and,

$$\hat{\Gamma}(\mathbf{x}, \mathbf{x}') \equiv \sum_j \hat{v}^{(\text{in})}(\mathbf{x}, \mathbf{x}_S^{(j)}) \hat{v}^{(\text{in})*}(\mathbf{x}', \mathbf{x}_S^{(j)}), \quad (15)$$

respectively. Note that we did not assume the different sources to be equally strong as  $\hat{s}^{(i)}$  represents the power spectrum of a specific source (at  $\mathbf{x}_S^{(i)}$ ). Furthermore, contrary to SI by crosscorrelation, the assumption that simultaneously acting sources are uncorrelated is not required: eq. (11) still holds in case the sources (or scatterers) illuminating  $\mathbb{S}_{\text{rcv}}$  exhibit some degree of correlation. The resulting spurious arrivals in  $C(\mathbf{x}_R, \mathbf{x}', t)$  are removed in the deconvolution process, because these arrivals are captured by  $\Gamma(\mathbf{x}, \mathbf{x}', t)$  as well. For details we refer to Wapenaar *et al.* (2012).

In Section 5, we employ slowness analysis, which allows the selection of time windows during which (the bulk of the) energy traverses  $\mathbb{S}_{\text{rcv}}$  before arriving at  $\mathbf{x}_R$ . This permits the application of SI by MDD. The effect of the 'blurring' of  $\hat{G}(\mathbf{x}_R, \mathbf{x})$  by the PSF can be removed by MDD. Moreover, since the ensemble of time windows over which crosscorrelations are averaged is in practice not infinite, spurious arrivals associated with correlations between signal from different noise sources remain after averaging. Deconvolution by the PSF suppresses these spurious arrivals (Wapenaar *et al.* 2011a).

Practical implementation of MDD requires discretization of the integral along  $\mathbb{S}_{\text{rcv}}$ . In matrix notation, eq. (11) can be written as

$$\hat{C} = \hat{G} \hat{\Gamma}, \quad (16)$$

where the rows and columns of  $\hat{C}$  correspond to different  $\mathbf{x}_R$  and  $\mathbf{x}'$ , respectively;  $\hat{G}$  and  $\hat{\Gamma}$  are organized accordingly. Eq. (16) is solved

for each discrete frequency individually. The (frequency-dependent) constant  $\hat{D}$  is implicitly included in the obtained solutions for  $\hat{G}$ . Right multiplying eq. (16) by the stabilized inverse of  $\mathbf{\Gamma}$  gives

$$\hat{G} = \hat{C} (\hat{\mathbf{r}} + \epsilon^2 \mathbf{I})^{-1}, \quad (17)$$

where  $\mathbf{I}$  denotes the identity matrix and  $\epsilon$  is a small number. The degree to which the multidimensional deconvolution has been able to remove the imprint of the PSF can be diagnosed by the virtual-source function (VSF; van Dalen *et al.* 2015):

$$\hat{\mathbf{Y}} = \hat{\mathbf{r}} (\hat{\mathbf{r}} + \epsilon^2 \mathbf{I})^{-1}. \quad (18)$$

The VSF coincides with an identity matrix in case of perfect deconvolution. The inverse Fourier transform of  $\hat{\mathbf{Y}}$ , that is  $\Upsilon(\mathbf{x}, \mathbf{x}', t)$ , represents the space-time source function of the virtual-source response retrieved through SI by MDD. If the deconvolution has been successful for all frequencies, it coincides with a band-limited spike centred at  $\mathbf{x} = \mathbf{x}'$ .

Apart from the fact that only inward-propagating energy is considered at  $\mathbb{S}_{\text{rcv}}$ , the right-hand sides of eqs (12) and (7) coincide. Assuming the one-sided illumination depicted in Fig. 1 therefore facilitates a direct comparison between SI by crosscorrelation and SI by MDD. First, accurate response retrieval through SI by crosscorrelation relies critically on the condition that the sources along  $\mathbb{S}_{\text{src}}$  illuminate the receivers with equal power. The accuracy of the estimate of  $\hat{G}(\mathbf{x}_R, \mathbf{x})\hat{S}$  by the crosscorrelation function degrades if this condition is not met. Additionally, for the configuration considered in Fig. 1(a), the medium is required to be homogeneous and lossless in order to be able to accurately retrieve  $\hat{G}(\mathbf{x}_R, \mathbf{x})\hat{S}$ . Finally, in case the sources along  $\mathbb{S}_{\text{src}}$  are simultaneously acting, they are required to be uncorrelated. SI by MDD does not suffer from any of these shortcomings: multidimensional deconvolution by the PSF corrects the CCF for the effects of (i) irregularities in the distribution of sources, (ii) differences in the power spectra between the different sources, and (iii) correlation between different sources. Furthermore, the one-sided illumination due to sources on  $\mathbb{S}_{\text{rcv}}$  implies that the integral along  $\mathbb{S}$  in eq. (1) is only partly evaluated. In the presence of scattering, this absence of equally powerful sources enclosing  $\mathbb{V}$ , that is, along  $\mathbb{S}$ , results in spurious arrivals in the crosscorrelation function due to incomplete cancellation of contributions from different stationary-phase points (Snieder *et al.* 2008; Snieder & Fleury 2010). These spurious arrivals may erroneously be interpreted as being part of  $\hat{G}(\mathbf{x}_R, \mathbf{x})$ . Virtual-source responses obtained through SI by MDD do not contain such spurious arrivals, because the deconvolution by the PSF removes them.

Assuming the medium inside  $\mathbb{V}$  is homogeneous, the phase of the correlation function coincides with the phase of the actual medium's response in case the illumination by the sources along  $\mathbb{S}_{\text{src}}$  is uniform. Such illumination also implies that the phase of the response retrieved through SI by MDD coincides with the phase of the response retrieved through SI by crosscorrelation. This becomes apparent from eq. (11): uniform illumination of  $\mathbb{S}_{\text{rcv}}$  implies  $\hat{\mathbf{r}}(\mathbf{x}, \mathbf{x}')$  is purely real and hence the phases of  $\hat{C}(\mathbf{x}_R, \mathbf{x}')$  and  $\hat{G}(\mathbf{x}_R, \mathbf{x})$  coincide. Even for that ideal case, however, multidimensional deconvolution will still deconvolve the (average) autocorrelation of the source-time function, that is  $S(t)$ , from the crosscorrelation function and correctly account for the losses in the medium.

### 2.3 Estimating phase velocity

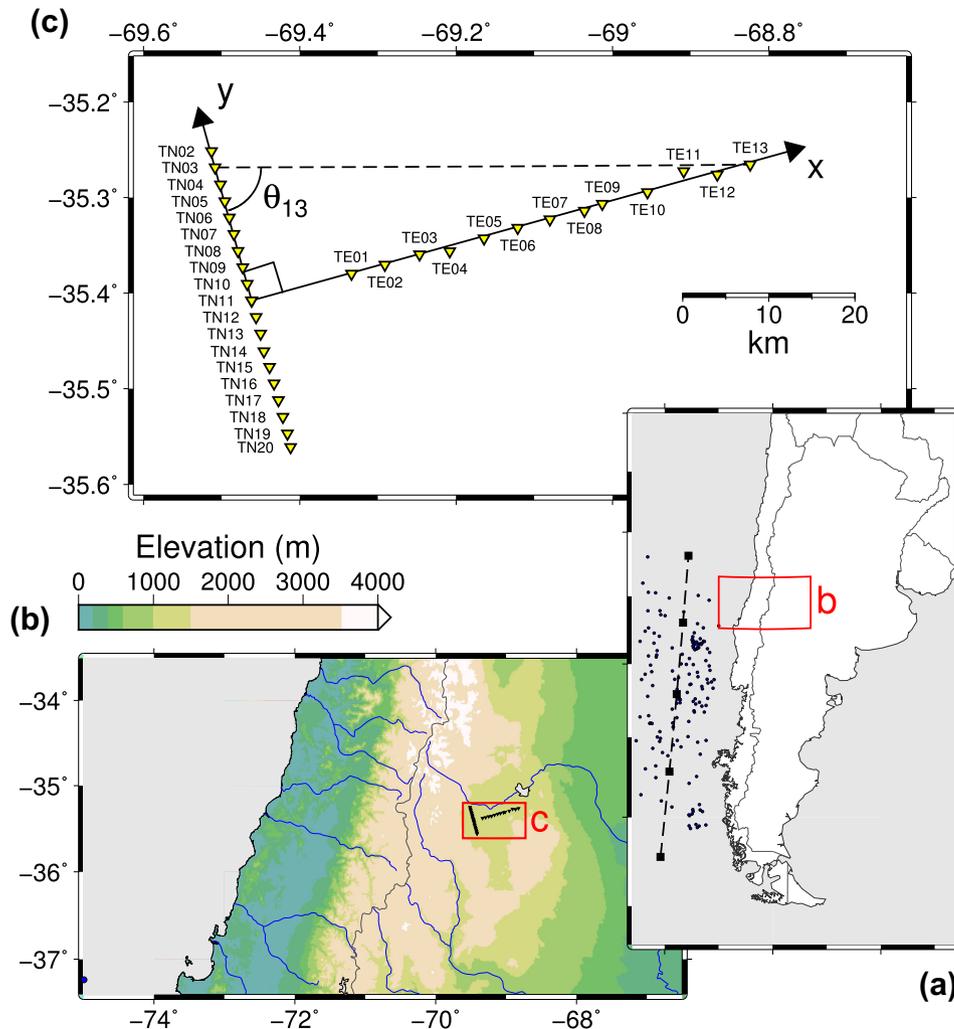
For a single surface wave mode propagating along the surface of a laterally homogeneous and isotropic Earth, the particle-velocity Green's functions in eqs (1) and (8) are proportional to zeroth-order Hankel functions of the second kind (Aki & Richards 2002; van Dalen *et al.* 2015),

$$\hat{G}(kr) \propto H_0^{(2)}(kr), \quad (19)$$

where  $r$  is the horizontal distance between the source and receiver coordinate. The wavenumber  $k$  coincides with  $\omega/c(\omega)$  for a lossless medium, where  $c(\omega)$  is the frequency-dependent phase velocity. Since  $H_0^{(2)}(kr) \equiv J_0(kr) - iY_0(kr)$ , where  $J_0$  and  $Y_0$  are zeroth-order Bessel functions of the first and second kind, respectively, inserting  $H_0^{(2)}(kr)$  in eq. (1) yields  $2J_0(kr)$  at the left-hand side. This proportionality of the crosscorrelation to a zeroth-order Bessel function has previously been recognized (e.g. Aki 1957; Yokoi & Margaryan 2008) and has been used to estimate both surface wave phase velocity (e.g. Ekström *et al.* 2009) and surface wave attenuation (Prieto *et al.* 2009; Weemstra *et al.* 2013).

The one-sided illumination considered in this study implies that only estimates of causal Green's functions are retrieved. A surface wave velocity estimate can therefore be obtained by equating the zeros of the real and imaginary part of a retrieved  $\hat{C}$  (or  $\hat{G}$ ) to the zeros of  $J_0(kr)$  and  $-Y_0(kr)$ , respectively. By only considering the zeros, we discard amplitude information. This implies that values  $\omega_i$  ( $i = 1, 2, \dots$ ) of frequency are identified for which the real part of the retrieved  $\hat{C}$  (or  $\hat{G}$ ) is zero. For each of these frequencies, a set of phase velocities  $c_n(\omega_i)$  ( $n = 1, 2, \dots$ ) exist for which  $\mathcal{J}[\omega_i r/c_n(\omega_i)] = 0$ . Similarly, denoting the frequencies for which the imaginary part of the retrieved response coincide with zero by  $\omega_j$  ( $j = 1, 2, \dots$ ), a set of phase velocities  $c_m(\omega_j)$  ( $m = 1, 2, \dots$ ) can be found for each  $\omega_j$  for which  $-Y_0[\omega_j r/c_m(\omega_j)] = 0$ . Selecting a single phase velocity  $c_n$  per  $\omega_i$  as well as a single  $c_m$  per  $\omega_j$ , an array of phase velocities representing a single dispersion curve is obtained. Of all the different possible combinations of phase velocities, the one that gives the seismologically most plausible dispersion curve is chosen (e.g. Boschi *et al.* 2013).

A caveat should be made regarding the measurement of phase velocity between stations whose separation is small with respect to the wavelength. Halliday & Curtis (2008) and van Dalen *et al.* (2015) employ a stationary-phase approximation while evaluating the integrals in eqs (A2) and (A4), respectively. This implies essentially that an infinite-frequency assumption is made. In fact, Tsai (2009) shows that the right-hand side of eq. (1) coincides with the sum of  $J_0(kr) - iH_0(kr)$  and its complex conjugate, where  $H_0(kr)$  is a Struve function of order zero. Even for a uniformly illuminated 2D medium, therefore, the coincidence between the crosscorrelation and the Green's function deteriorates at distances small with respect to the wavelength. For example, at a distance of one wavelength, that is  $kr = 2\pi$ , the maximum phase difference between  $J_0(kr) - iH_0(kr)$  and  $J_0(kr) - iY_0(kr)$  is approximately 0.3 rad, which corresponds to a traveltime error of  $0.3/(2\pi) \times 100$  per cent  $\approx 4.8$  per cent. At distances larger than a few wavelengths, however,  $H_0(kr)$  and  $Y_0(kr)$  behave very similar (at a distance of three wavelengths, the maximum traveltime error has already reduced to 1.8 per cent). Crosscorrelations computed between stations separated by less than a few wavelengths are therefore generally discarded (e.g. Yao *et al.* 2006). Note that the real part of the infinite-frequency approximation of the crosscorrelation and the real part of the exact crosscorrelation (and hence their zeros) still coincide at short distances (Tsai 2009).



**Figure 2.** Location (a,b) and configuration (c) of the T-array. The inverted triangles depict the station locations. The arrows define the local coordinate system employed in Section 4. In that section, we also use the angle  $\theta$  to select time windows during which the wavefield possesses a favourable energy flux (see Section 4). The blue dots in a depict the locations of sources used in the numerical modelling. The black dashed line indicates the cosine probability function whose peak, zeros, and inflection points are depicted by black squares.

We finally note that in case the medium is dissipative,  $k$  is complex-valued with its real and imaginary part coinciding with  $\omega/c(\omega)$  and  $\alpha$  respectively, that is,  $k = \omega/c - i\alpha$ , where  $\alpha$  denotes the attenuation coefficient. For values of  $\alpha$  that are small in the sense that  $\alpha \ll \omega/c$ , the complex wavenumber can be approximated by  $(\omega/c)\sqrt{1 - (2i\alpha c/\omega)}$  (Weemstra *et al.* 2015). The Hankel function may in that case be approximated by  $H_0^{(2)}(\omega r/c(\omega))e^{-\alpha r}$  (Tsai 2011). The procedure for estimating the phase velocity described above therefore applies equally well to dissipative media.

### 3 THE MALARGÜE SEISMIC ARRAY

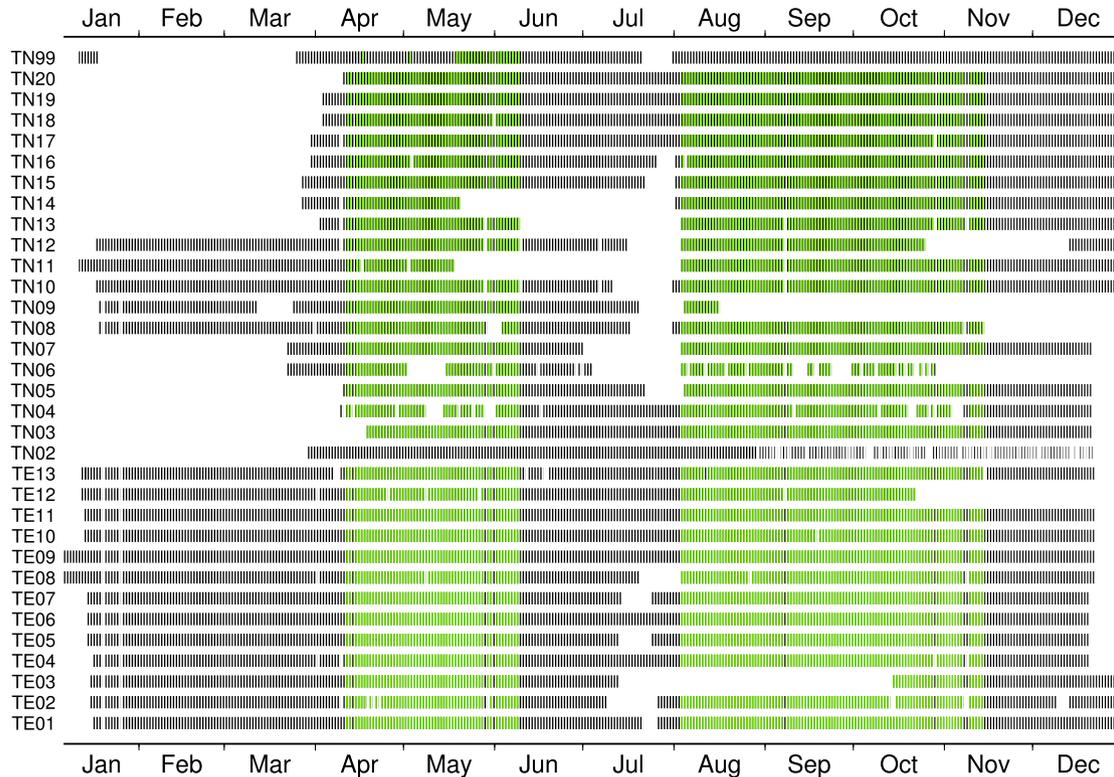
The MalARRgüe was installed in western Argentina at the beginning of 2012. It consisted of two subarrays, a ‘P-array’ and a ‘T-array’, separated by approximately 100 km. In this work, we only consider the data recorded by the T-array. More detailed information about MalARRgüe can be found in Ruigrok *et al.* (2012). Figs 2(a) and (b) present the location of the T-array, which was centred at about  $35^\circ 24'S$ ,  $69^\circ 27'W$ . It was situated on a  $\sim 1400$  m high plateau along the eastern flanks of the immense Andean mountain range. The area is part of the Mendoza administrative unit.

#### 3.1 Geological setting

As a result of the Andean orogeny the north-south trending Malargüe anticline developed in the area of the MalARRgüe (Kraemer *et al.* 2011). The fold axis of this anticline is located approximately 10 km west of the most western stations of the T-array. East of this anticline, a syntectonic depocenter developed, which is generally referred to as the Malargüe basin. The formations that developed in this basin underlie most of the T-array stations. At its deepest point the Malargüe basin accumulated over 3 km of Tertiary sediments. These sediments rest unconformably over the older Cretaceous strata and are mainly of Pliocene age (5–1 Ma). They vary in thickness due to the differential basin subsidence associated with the Andean orogeny; seismic sections presented in Kraemer *et al.* (2011) show how the formations thin progressively towards the east.

#### 3.2 Array and data characteristics

During almost a year, the T-array continuously sampled (excluding intermittent gaps in recording at some stations) the wavefield at 32 distinct locations with a sampling rate of 100 samples per second.



**Figure 3.** Data availability during 2012 for each of the stations of the T-array (black bars). Green coloured bars (overlain) represent periods during which the data quality was sufficiently high for the application of MDD. The diurnal gaps are due to the spiky nature of the daytime recordings (Appendix B).

We distinguish between a line of stations aligned NNW–SSE and a line of stations oriented perpendicular to that line; We refer to these two lines as the ‘TN-line’ and the ‘TE-line’, respectively (Fig. 2c). All sensors recorded three orthogonal components of particle velocity. The deployed sensors were short-period with a corner frequency of 2 Hz (Sercel L-22) and were borrowed from the Incorporated Research Institutions for Seismology (IRIS) under the conditions of the PASSCAL program (PASSCAL is an acronym for ‘Portable Array Seismic Studies of the Continental Lithosphere’). Co-located with short-period sensor TN11 was a single broad-band sensor (Güralp CMG-40T), dubbed TN99. This broad-band sensor had a corner frequency of 0.033 Hz and served as a reference for the recordings by the short-period sensors. In Appendix B, we show that, despite the corner frequency of 2 Hz, the recordings of the short-period sensors can be used for the application of SI down to about 0.1 Hz.

In part, the T-array has been designed for the application of SI by MDD (Ruigrok *et al.* 2012): surface waves induced by ocean-seafloor interaction in the Southern Pacific Ocean will traverse the TN-line before arriving at a TE station. Consequently, this energy meets the requirements associated with SI by MDD concerning the direction of propagation through the ‘boundary’ of the ‘volume’; the TN-line being part of the boundary and the area in which the TE stations reside being (part of) the volume in this case (compare Fig. 2c with Fig. 1). The nominal station separation of only 2 km along the TN-line allows for MDD of surface waves with wavelengths longer than 4 km. A surface wave phase velocity of  $2 \text{ km s}^{-1}$  would therefore allow frequencies up to 0.5 Hz to be eligible for the application of SI by MDD. Of course, lower/higher phase velocities translate to lower/higher upper frequency limits. The same criterion applies for

the estimation of the (frequency-dependent) slowness of an incident wave along the TN-line (see Section 4). This allows separation of waves travelling in a northward or southward direction and waves travelling in an eastward or westward direction. Together with a slowness analysis along the TE-line this enables us to select time windows during which (most of) the energy is propagating into the appropriate direction for the application of SI by MDD (i.e. predominantly eastward). This implies that virtual sources are created at the locations of the TN stations, whereas we retrieve their responses at the locations of the TE stations. The nominal station separation along the TE-line is 4 km.

In the previous section, we have shown that SI by MDD involves deconvolving the virtual-source responses by a PSF. We build the PSF from the time-averaged noise correlations between the TN stations. The latter implies that noise recordings are required to be available over the same period of time for all TN stations and the TE station at which we want to retrieve the virtual-source response. Fig. 3 presents the temporal data coverage of all the stations of the T-array. Data represented by green bars are selected for the application of SI by MDD. The recordings by sensor TN02 are discarded altogether for the application of SI by MDD because of poor data quality. During the periods selected for the application of SI by MDD, the data set meets the following criteria: (i) only four or less TN stations are offline, (ii) none of these offline TN stations are adjacent and, (iii) not a single other TN station is offline in case station TN03 or station TN20 is. These criteria allow proper interpolation of the point-spread function if needed (or extrapolation in case of station TN03 or TN20). We finally note that we have used the data recorded by station TN99 in case station TN11 was offline.

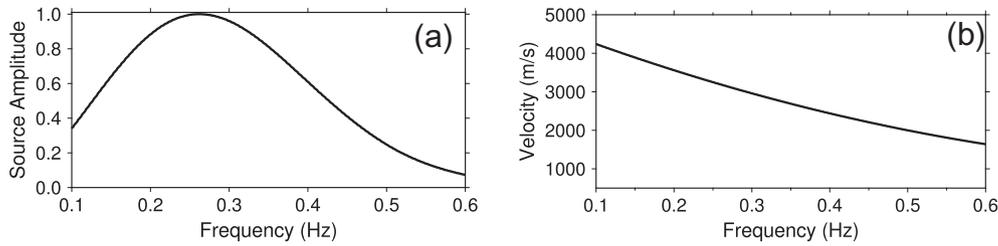


Figure 4. Amplitude spectrum (a) and phase-velocity dispersion (b) employed in our model.

#### 4 PROOF OF CONCEPT: A SIMPLE NUMERICAL MODEL

In this section, we showcase the benefits of using SI by MDD instead of using SI by crosscorrelation for the station configuration of the T-array. We modelled synthetic data due to excitations along the west coast of South America. These synthetic recordings were processed with the same software as was used for the field data later on. The modelled data hence served as an excellent test case for our slowness analysis, crosscorrelation, and MDD codes. The considered source distribution is presented in Fig. 2(a) and rather arbitrary: source latitudes are drawn from an approximately north-south aligned cosine probability function (dashed line) whose peak, zeros, and inflection points are depicted by black squares; source longitudes are drawn from a uniform probability function with an east-west extent of a few hundred kilometres. Two additional clusters of sources are created, representative of increased ocean wave activity (e.g. due to regional storms). These sources act sequentially, which implies that eq. (14) applies. We assume a laterally homogeneous, non-dissipative Earth and vertical point sources (conform eqs 1 and 8; see Appendix A). The (frequency-domain) surface wave Green's functions are therefore simply given by the Hankel functions described in Section 2.3. The amplitude spectra of the sources coincide and the phase is zero for all frequencies. The fact that the amplitude spectra coincide, implies that the non-uniform illumination of the T-array is solely due to the spatially non-uniform distribution of sources. The source amplitude spectrum is shown in Fig. 4(a). An increase of phase velocity with depth is suggested by means of a frequency-dependent phase-velocity (Fig. 4b). The values in Figs 4(a) and (b) are representative of the amplitude of the Earth's ambient seismic field and its surface wave phase velocity, respectively (e.g. Ekström 2014).

##### 4.1 Slowness analysis

The theory in Section 2.2 relies on the assumption of a one-way wavefield at  $\mathbb{S}_{\text{rcv}}$ . In practice, this often implies that inward- and outward-propagating waves need to be separated along  $\mathbb{S}_{\text{rcv}}$  (e.g. Grobbe *et al.* 2013; van Dalen *et al.* 2015). The TN stations make up  $\mathbb{S}_{\text{rcv}}$  in our case, whereas their responses can be retrieved at the locations of the TE stations. Using the coordinate system defined in Fig. 2(c), we employ a frequency-dependent slowness analysis along (a subset of) the TN-line and TE-line to select time windows during which the wavefield is characterized by a favourable energy flux.

Let us define the crosscorrelation between a pair of stations (auxiliary numbered  $m$  and  $n$ ) along the TN-line by,

$$\hat{C}_{mn}^{(\text{TN})} \equiv \hat{v}_m \hat{v}_n^*, \quad (20)$$

where  $\hat{v}_m$  and  $\hat{v}_n$  are the Fourier-transformed recordings of vertical particle velocity at stations  $m$  and  $n$ , respectively. Because we model

sequential sources in this section, each recording contains signal due to a single source. In Section 5, however, the  $\hat{C}_{mn}^{(\text{TN})}$  are associated with crosscorrelations of individual time windows. We define the 'power' along (a subset of) this line by

$$\hat{P}^{(\text{TN})}(p_y) \equiv \frac{1}{N(N-1)/2} \sum_{m=1}^N \sum_{n=m+1}^N \frac{\hat{C}_{mn}^{(\text{TN})}}{|\hat{C}_{mn}^{(\text{TN})}|} e^{i\omega p_y(y_m - y_n)}, \quad (21)$$

where  $p_y$  denotes the slowness in the  $y$ -direction and  $y_m$  and  $y_n$  the  $y$ -coordinates of the stations associated with recordings  $\hat{v}_m$  and  $\hat{v}_n$ , respectively. The number of TN stations used to obtain  $\hat{P}^{(\text{TN})}$ , is denoted by  $N$ . In the application to the field data, this number may vary with frequency and time (see Section 5). For  $\hat{C}_{mn}^{(\text{TN})}$  computed from recordings of a plane wave whose slowness in the  $y$ -direction is denoted by  $\hat{p}_y^{(\text{TN})}$ , the elements of the double sum in eq. (21) will sum constructively along the real axis for  $\hat{p}_y = \hat{p}_y^{(\text{TN})}$ . In practice, we average  $\hat{P}^{(\text{TN})}(\omega, p_y)$  over a bandwidth of a quarter octave centred around  $\omega$  to obtain a more stable result.

The slowness along the TN-line at which the bulk of the energy propagates is obtained by,

$$\hat{p}_y^{(\text{TN})} = \arg \max_{p_y} \left( \Re \left[ \hat{P}^{(\text{TN})}(p_y) \right] \right), \quad (22)$$

where the operator  $\Re[\dots]$  maps its complex argument into its real part. Time windows for which  $|\hat{p}_y^{(\text{TN})}|$  is larger than a certain threshold are discarded. This threshold depends on the location of the TE station at which we aim to retrieve the virtual-source response and a reference phase velocity  $c_{\text{ref}}^{(\text{TN})}$  along the TN-line. The threshold for a station TE $k$  ( $k = 1, 2, \dots, 13$ ) is given by

$$\hat{p}_{\text{th}}^{(\text{TN})}(\text{TE}k) = \frac{\cos(\theta_k)}{c_{\text{ref}}^{(\text{TN})}}, \quad (23)$$

where the angle  $\theta_k$  between the  $y$ -axis and the location of station TE $k$  is measured at station TN03 (see Fig. 2c). This implies that for lower  $k$  (and hence higher  $\cos(\theta_k)$ ), a larger number of time windows will qualify for the application of SI by MDD.

An additional slowness analysis along the TE-line is performed to determine whether the energy is propagating in a positive or negative  $x$ -direction and at which speed. This second procedure is contingent on the outcome of the slowness analysis along the TN-line: it is only initiated for those time windows (and frequencies) that were not discarded. Similar to eq. (20), the crosscorrelation between a pair of stations (auxiliary numbered  $m$  and  $n$ ) along the TE-line is defined as,

$$\hat{C}_{mn}^{(\text{TE})} \equiv \hat{v}_m \hat{v}_n^*, \quad (24)$$

where  $\hat{v}_m$  and  $\hat{v}_n$  are the Fourier-transformed recordings of vertical particle velocity at stations  $m$  and  $n$ , respectively. We define,

$$\hat{P}^{(\text{TE})}(p_x) \equiv \frac{1}{N(N-1)/2} \sum_{m=1}^N \sum_{n=m+1}^N \frac{\hat{C}_{mn}^{(\text{TE})}}{|\hat{C}_{mn}^{(\text{TE})}|} e^{i\omega p_x(x_m - x_n)}, \quad (25)$$

**Table 1.** Meaning of the various acronyms and abbreviations used in this work.

Abbreviation	Details
MalARRgue	The Malargüe Seismic Array
PASSCAL	Portable Array Seismic Studies of the Continental Lithosphere
MDD	Multidimensional deconvolution (applied to CC1 responses)
CCF	The crosscorrelation function; defined in eq. (12)
PSF	The point-spread function; defined in eq. (13)
VSF	The virtual-source function; defined in eq. (18)
CC1	CCF computed from recordings selected by means of slowness analysis
CC3	CCF computed from all available recordings (green coloured bars in Fig. 3)
DM	Directly modelled virtual-source responses

where  $p_x$  denotes the slowness in the  $x$ -direction and  $x_m$  and  $x_n$  the  $x$ -coordinates of the stations associated with recordings  $\hat{v}_m$  and  $\hat{v}_n$ , respectively. The number of TE stations used to obtain  $\hat{P}^{(\text{TE})}$ , is denoted by  $N$ . Just as for  $\hat{P}^{(\text{TN})}$ , this number may vary with frequency and time in the application to the field data (see Section 5). For  $\hat{C}_{mn}^{(\text{TE})}$  computed from recordings of a plane wave whose slowness in the  $x$ -direction is denoted by  $\hat{p}_x^{(\text{TE})}$ , the elements of the double sum in eq. (25) will sum constructively along the real axis for  $\hat{p}_x = \hat{p}_x^{(\text{TE})}$ . Just as for  $\hat{P}^{(\text{TN})}(p_y)$ , we average  $\hat{P}^{(\text{TE})}(\omega, p_x)$  over a bandwidth of a quarter octave centred around  $\omega$ . The slowness at which most energy propagates along the TE-line is given by,

$$\hat{p}_x^{(\text{TE})} = \arg \max_{p_x} \left( \Re \left[ \hat{P}^{(\text{TE})}(p_x) \right] \right). \quad (26)$$

Time windows for which  $\hat{p}_x^{(\text{TE})}$  is smaller than a certain threshold are neglected. Of course, also this threshold depends on the location of the TE station at which we aim to retrieve the virtual-source response and another reference phase velocity, denoted by  $c_{\text{ref}}^{(\text{TE})}$ . The threshold for a station TE $k$  ( $k = 1, 2, \dots, 13$ ) is given by

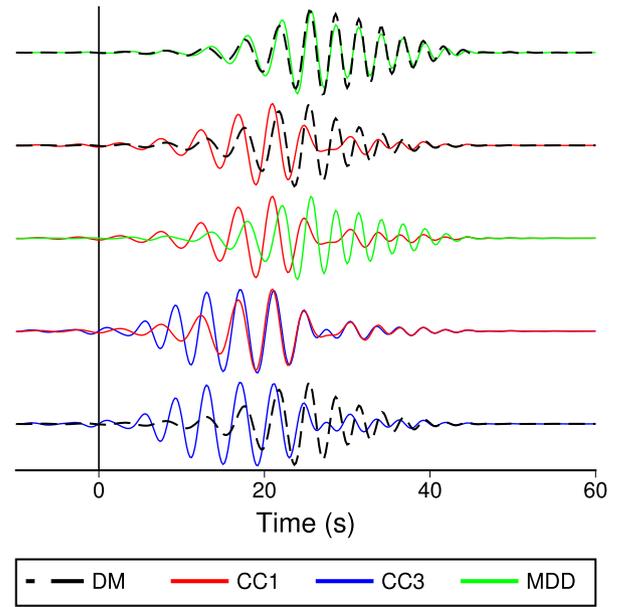
$$\hat{p}_{\text{th}}^{(\text{TE})}(\text{TE}k) = \frac{\sin(\theta_k)}{c_{\text{ref}}^{(\text{TE})}}. \quad (27)$$

The thresholds defined in eqs (23) and (27) depend on the reference phase velocities  $c_{\text{ref}}^{(\text{TN})}$  and  $c_{\text{ref}}^{(\text{TE})}$ , respectively. While evaluating the slownesses of our synthetic recordings, we simply use the model's dispersion curve, that is, the curve in Fig. 4(b), to compute these thresholds. For our field data, this is of course not possible. In that case, we estimate  $c_{\text{ref}}^{(\text{TN})}$  and  $c_{\text{ref}}^{(\text{TE})}$  separately, using responses obtained through SI by crosscorrelation of all available recordings (black bars in Fig. 3; see Section 5).

## 4.2 Results

We compare responses obtained using SI by crosscorrelation with responses obtained through SI by MDD. Results obtained from the application of SI by crosscorrelation using all time windows (and hence all sources) are referred to as CC3. Application of SI by crosscorrelation to that subset of the data for which  $|\hat{p}_y^{(\text{TN})}|$  is smaller than  $\hat{p}_{\text{th}}^{(\text{TN})}$  and  $\hat{p}_x^{(\text{TE})}$  is larger than  $\hat{p}_{\text{th}}^{(\text{TE})}$ , yields responses which we refer to as CC1. The chosen abbreviations are consistent with the study of van Dalen *et al.* (2015). Multidimensionally deconvolving the CC1 responses by the PSF subsequently results in the improved MDD responses. We additionally benchmark the retrieved responses against the directly modelled (DM) responses. A summary of these (and other) abbreviations is given in Table 1.

The non-uniform illumination causes the virtual-source responses to deviate significantly from the actual source responses. This is exemplified in Fig. 5 using the response at station TE10 due to a virtual source at the location of station TN07: the CC3 response

**VS response of TN07 at TE10 (0.20 – 0.40 Hz)**

**Figure 5.** Comparison of the different methods and/or data sets (abbreviations explained in the main text) using the synthetic data. Responses are retrieved at station TE10 and are due to a virtual source at station TN07. The responses are filtered between 0.2 and 0.4 Hz.

is rather dissimilar from the correct medium response (CC3 versus DM; bottom). The selection of time windows with an energy flux fit for the application of SI by MDD yields some improvement (CC1 versus CC3), but there remains a discrepancy (CC1 versus DM). Applying SI by MDD to the CC1 results (CC1 versus MDD) mostly corrects the crosscorrelation results for the non-uniform illumination pattern and results in a good fit to the directly modelled response (MDD versus DM). Fig. S1 (Supporting Information) makes the same comparison for most of the virtual sources. Especially for the TN stations with low numbers, application of MDD results in a good fit to the directly modelled responses (top left), whereas the fit of the CC1 responses is significantly poorer (top right). The difference between the two methods can be assessed, qualitatively at least, at the bottom right.

A similar numerical example presented by Wapenaar *et al.* (2011a) considered simultaneously acting, uncorrelated noise sources instead of sequential sources. Station-station crosscorrelations computed from such individual noise recordings contain spurious arrivals due to the simultaneously acting sources. Averaging crosscorrelations over a large number of time windows reduces the relative amplitude of such spurious signal (eq. 12). For a finite number of time windows, however, such signal will never disappear

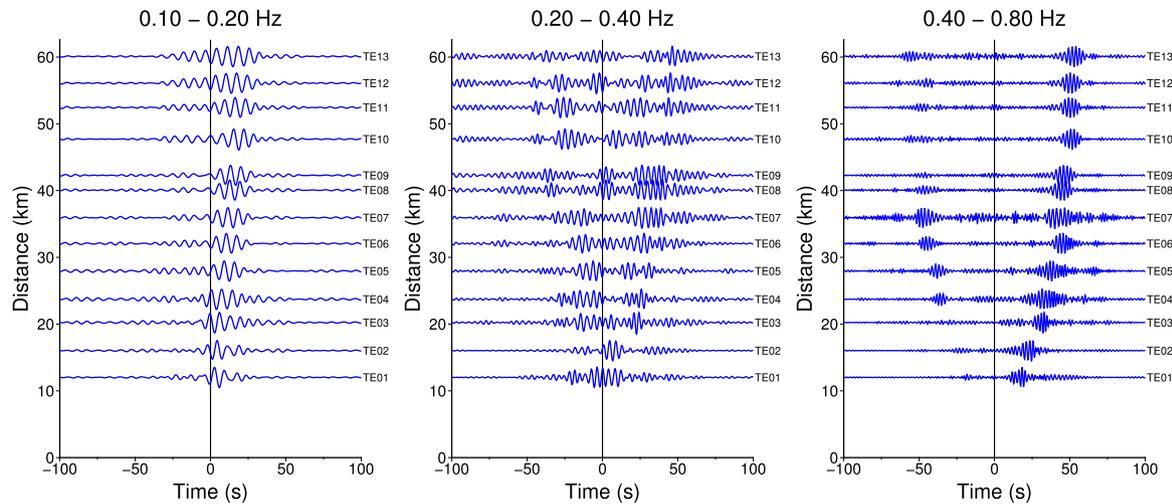


Figure 6. Virtual-source responses along the TE-line obtained from SI by crosscorrelation; the virtual source is located at station TN11.

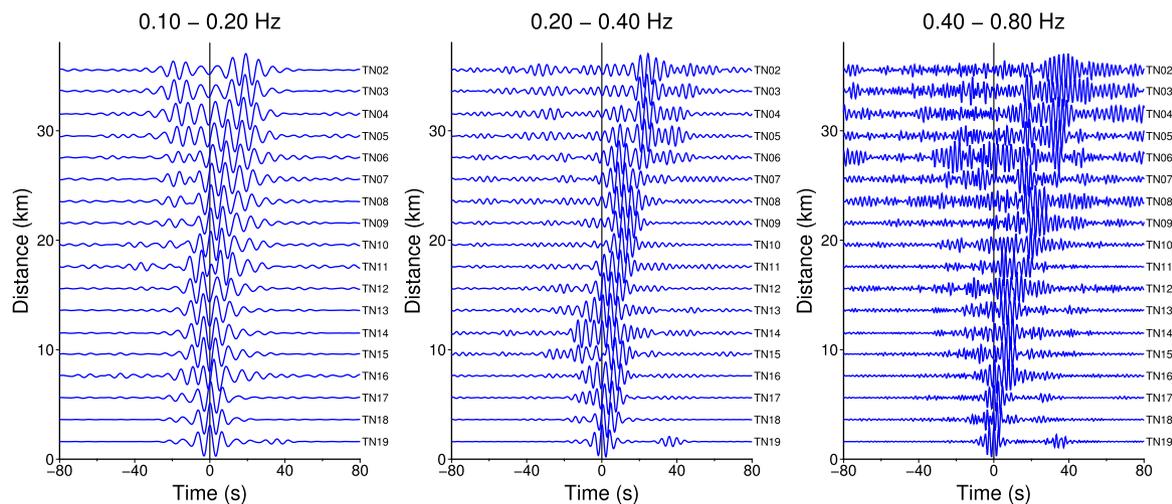


Figure 7. Virtual-source responses along the TN-line obtained from SI by crosscorrelation; the virtual source is located at station TN20.

completely. Wapenaar *et al.* (2011a) show that SI by MDD also has the ability to suppress these spurious arrivals.

## 5 APPLICATION TO THE MALARRGUE DATA SET

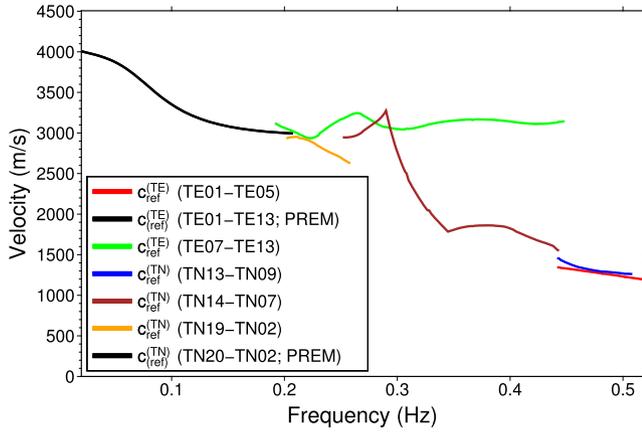
In this section, we present the results of the application of SI to the noise recordings of the T-array. Throughout this section, the length of the time window employed is 10 min and an overlap of 50 per cent between consecutive windows is used. We first describe the preprocessing and present the results of the slowness analysis; subsequently we present the results of the application of SI by MDD to the recordings of the selected time windows. For brevity, we only present the MDD results for frequencies lower than 0.2 Hz in this section. In Appendix C, results for frequencies higher than 0.2 Hz are presented.

### 5.1 Preprocessing and slowness analysis

We employ the frequency-dependent slowness analysis described in Section 4.1 to select time windows during which the wavefield possesses a favourable energy flux. This implies that we require estimates of the surface wave phase velocities along the TN-line and

TE-line to determine for each station TE $k$  the thresholds  $\hat{p}_{\text{trh}}^{(\text{TN})}$  (TE $k$ ) and  $\hat{p}_{\text{trh}}^{(\text{TE})}$  (TE $k$ ), respectively. For frequencies below 0.2 Hz we use Rayleigh wave fundamental-mode phase velocities for the preliminary reference Earth model (PREM; Dziewonski & Anderson 1981) with the oceanic layer removed. At these frequencies, the wavelengths of the surface waves are large with respect to the average station separation. The PREM does not accurately describe the near surface, however, and can therefore not be used to predict phase velocities of surface waves sensitive to shallower structure. For frequencies higher than 0.2 Hz we therefore estimate  $c_{\text{ref}}^{(\text{TN})}$  and  $c_{\text{ref}}^{(\text{TE})}$  from the surface wave responses obtained from SI by crosscorrelation using the entire data set (black coloured data in Fig. 3).

The time-averaged crosscorrelations between station TN11 and all the TE stations, and station TN20 and all the other TN stations are depicted in Figs 6 and 7, respectively. Energy at positive time (generally referred to as the ‘causal part’) is associated with signal travelling from the virtual source to the receiver, whereas energy at negative time is associated with energy travelling in the opposite direction (generally referred to as the ‘acausal part’). Along the TE-line, as well as the TN-line, the causal parts exhibit the largest amplitudes for all frequency bands. This implies that, on average, most ambient surface waves are travelling eastward and northward, respectively. The virtual-source responses of TN11 and



**Figure 8.** Dispersion curves derived from the frequency-wavenumber spectra presented in the Supporting Information (Figs S2–S5). Note that below 0.2 Hz Rayleigh-wave fundamental-mode phase velocities for the preliminary reference Earth model were used. Consequently,  $c_{\text{ref}}^{(\text{TN})}$  and  $c_{\text{ref}}^{(\text{TE})}$  coincide below 0.2 Hz.

TE20 (i.e. the causal parts in Figs 6 and 7, respectively) were used to estimate  $c_{\text{ref}}^{(\text{TE})}$  and  $c_{\text{ref}}^{(\text{TN})}$ , respectively. To that end, we Fourier transformed the responses to the frequency-wavenumber ( $f, k$ ) domain (e.g. Park *et al.* 1998). Estimating these velocities from the ( $f, k$ )-spectra implies that the medium is assumed to be invariant along the two station lines. Especially at higher frequencies, however, this assumption turns out to be severely violated. For example, between 0.4 and 0.8 Hz the moveout along the TE-line shows a clear bend around station TE06 (Fig. 6). We have therefore selected different subsets of stations for different frequency bands and computed frequency-wavenumber spectra for these subsets. Of course,  $\hat{P}^{(\text{TN})}$  and  $\hat{P}^{(\text{TE})}$  are later computed from the recordings by the same subsets of stations.

The frequency-wavenumber spectra from which the dispersion curves are extracted can be found in the Supporting Information (Figs S2–S5). Wavenumber curves are obtained by picking for each frequency the wavenumber associated with the maximum amplitude of the frequency-wavenumber spectrum; wavenumbers are only picked for those frequencies for which the frequency-wavenumber spectra exhibit a clear maximum. The picked wavenumbers are translated to velocities and the obtained curves are subsequently smoothed over a frequency band of a quarter octave to yield the dispersion curves depicted in Fig. 8. Around 0.2 Hz, the retrieved phase velocities are relatively close to the dispersion curve derived from the PREM model. At higher frequencies, however, a clear disparity between the surface wave phase velocity in the eastern part of the array (dispersion curve extracted from the virtual-source responses at stations TE07–TE13) and the western part of the array can be observed (other dispersion curves). The higher velocities in the east can most probably be attributed to the limited depth of the Malargüe basin below that part of the array (Kraemer *et al.* 2011; Nishitsuji *et al.* 2014).

The dispersion curves depicted in Fig. 8 are used to determine for each station TE*k* the thresholds  $\hat{p}_{\text{trh}}^{(\text{TN})}$  (TE*k*) and  $\hat{p}_{\text{trh}}^{(\text{TE})}$  (TE*k*). As pointed out in Section 4.1, completely accurate  $\hat{P}^{(\text{TN})}$  and  $\hat{P}^{(\text{TE})}$  will only be obtained in case of a homogeneous subsurface and the presence of a single source. In practice, however, signal originating from multiple (potentially simultaneously acting) sources may arrive in a single time window. Moreover, despite the use of subsets of stations, spatial variations of the surface wave phase velocity may still degrade the accuracy of the estimates  $\hat{p}_y^{(\text{TN})}$  and  $\hat{p}_x^{(\text{TE})}$ . To

account for such inaccuracies, we decrease  $c_{\text{ref}}^{(\text{TN})}$  by 10 per cent and increase  $c_{\text{ref}}^{(\text{TE})}$  by 10 per cent while computing the thresholds in eqs (23) and (27), respectively. Both  $\hat{P}^{(\text{TN})}$  and  $\hat{P}^{(\text{TE})}$  are averaged over a bandwidth of a quarter octave and  $\hat{P}^{(\text{TE})}$  is only computed for time windows for which  $|\hat{p}_y^{(\text{TN})}| < \hat{p}_{\text{trh}}^{(\text{TN})}$ . Furthermore, they are computed at increments of 0.01 Hz and hence  $\hat{p}_y^{(\text{TN})}$  and  $\hat{p}_x^{(\text{TE})}$  are compared against their respective thresholds at these increments only.

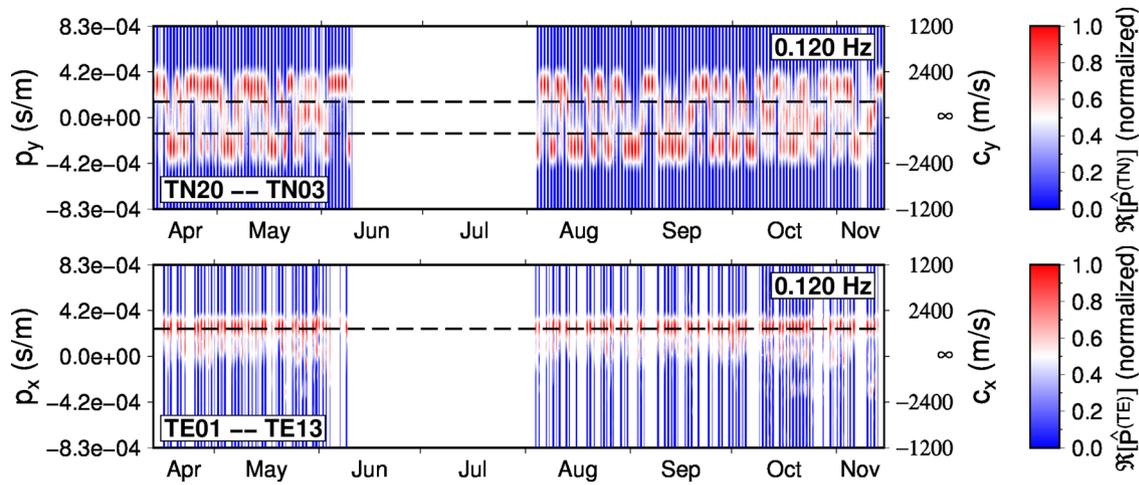
Fig. 9 presents the result of the slowness analysis at 0.12 Hz. The power along the TE-line ( $\Re[\hat{P}^{(\text{TE})}]$ ; bottom) is only shown for time windows during which the energy flux is favourable for the retrieval of a virtual-source response at station TE07. Most of the energy is propagating (sub)parallel to the TN-line, but in both directions. This appears to contradict the higher amplitude causal part of the virtual-source responses along the TN-line (Fig. 7). However, whereas the northward propagating energy hits the TE-line dominantly at a right angle (i.e.  $\hat{p}_x^{(\text{TE})} \approx 0$ ), the southward propagating energy is hitting this line mostly at smaller angles. Consequently, this energy is not coming from a stationary-phase direction and hence gives a subdominant contribution to the time-averaged crosscorrelation (Snieder 2004). Note that this is not immediately clear from Fig. 9, because  $\Re[\hat{P}^{(\text{TE})}]$  (bottom) is only displayed for windows for which  $|\hat{p}_y^{(\text{TN})}| < \hat{p}_{\text{trh}}^{(\text{TN})}$  and  $\hat{p}_x^{(\text{TE})} > \hat{p}_{\text{trh}}^{(\text{TE})}$ . A number of conclusions can be drawn from Fig. 9. First, it appears that there are sufficient time windows during which the wavefield possesses a favourable energy flux for the application of SI by MDD. Second, these time windows are (fairly) uniformly distributed over time. Finally, occasional body-wave energy ( $p_x \approx 0$ ) can be observed, which may hinder the successful application of MDD. Results of the slowness analysis at 0.32 Hz are presented in Fig. C1 of Appendix C.

Fig. 10 shows how the surface waves associated with the secondary microseisms illuminate the T-array more frequently from a favourable direction than the surface waves at higher frequencies (0.2–0.5 Hz) do. This results in an overall higher number of time windows that qualify for the application of SI by MDD. In the Supporting Information (Fig. S6), a histogram is presented showing the cumulative number of selected time windows for the MDD eligible period (green coloured data in Fig. 3). It is useful to note that although the absolute number of selected time windows changes from one TE station to another due to the changing thresholds, the pattern of the distribution of the number of selected time windows with frequency and time remains approximately the same.

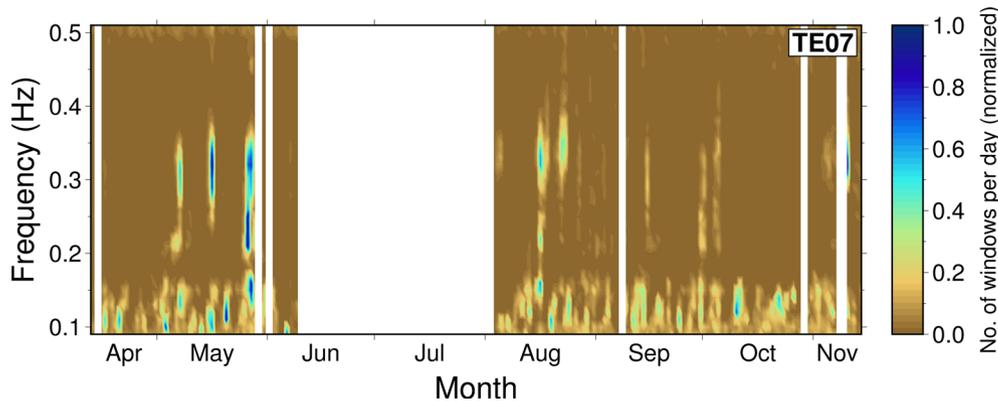
Fig. 11 presents the average  $\Re[\hat{P}^{(\text{TN})}]$  (solid line) and  $\Re[\hat{P}^{(\text{TE})}]$  (dashed line) at 0.12 Hz. It reveals that the selected time windows are, on average, associated with recordings of ambient surface wave noise with a positive  $y$  component; this statement holds for both frequencies. The average of  $\Re[\hat{P}^{(\text{TE})}]$  peaks of course just above  $\hat{p}_x^{(\text{TE})} = \hat{p}_{\text{trh}}^{(\text{TE})}$ , but does also show some energy at lower  $\hat{p}_x^{(\text{TE})}$ , which most likely can be attributed to body waves hitting the T-array from below. The average  $\Re[\hat{P}^{(\text{TN})}]$  and  $\Re[\hat{P}^{(\text{TE})}]$  at 0.32 Hz are presented in Fig. C2 of Appendix C.

## 5.2 Results

We apply SI by MDD to the crosscorrelations averaged over the selected time windows. Prior to crosscorrelation, and hence for each window individually (and also for each station individually), recordings are normalized with respect to the root-mean-square of the spectral amplitudes between 0.1 and 0.5 Hz. This reduces the impact of high amplitude, transient events on the time-averaged crosscorrelations. The PSF is built from the crosscorrelations



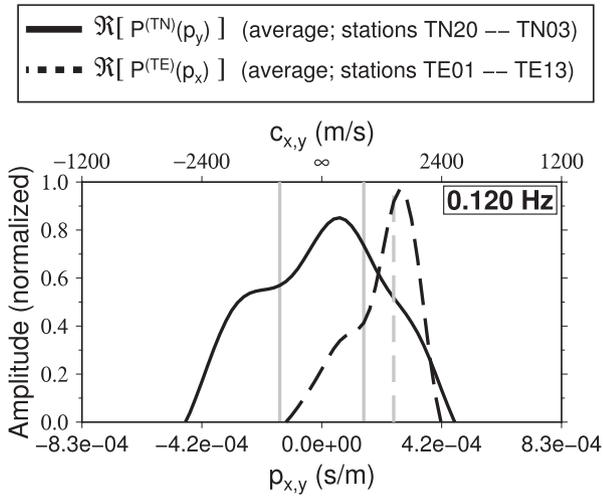
**Figure 9.**  $\Re[\hat{P}^{(TN)}]$  (top) and  $\Re[\hat{P}^{(TE)}]$  (bottom) at 0.12 Hz and normalized with respect to their maximum value, that is, with respect to  $\Re[\hat{P}^{(TN)}(\hat{p}_y^{(TN)})]$  and  $\Re[\hat{P}^{(TE)}(\hat{p}_x^{(TE)})]$ , respectively. Whereas  $\Re[\hat{P}^{(TN)}]$  are computed and depicted for all time windows,  $\Re[\hat{P}^{(TE)}]$  are only computed and shown for time windows during which  $|\hat{p}_y^{(TN)}| < \hat{p}_{th}^{(TN)}(\text{TE07})$  and  $\hat{p}_x^{(TE)} > \hat{p}_{th}^{(TE)}(\text{TE07})$ . Furthermore, although  $\hat{p}_y^{(TN)}$  and  $\hat{p}_x^{(TE)}$  were determined for each (10 min) time window individually,  $\Re[\hat{P}^{(TN)}]$  and  $\Re[\hat{P}^{(TE)}]$  are averaged over a period of one hour and half an hour, respectively. This was done for practical reasons: the size of the figure does not allow interpretation at higher resolution, whereas temporal averaging reduces the file size significantly. The dashed black lines indicate  $+\hat{p}_{th}^{(TN)}$  and  $-\hat{p}_{th}^{(TN)}$  (top) and  $\hat{p}_{th}^{(TE)}$  (bottom) for station TE07.



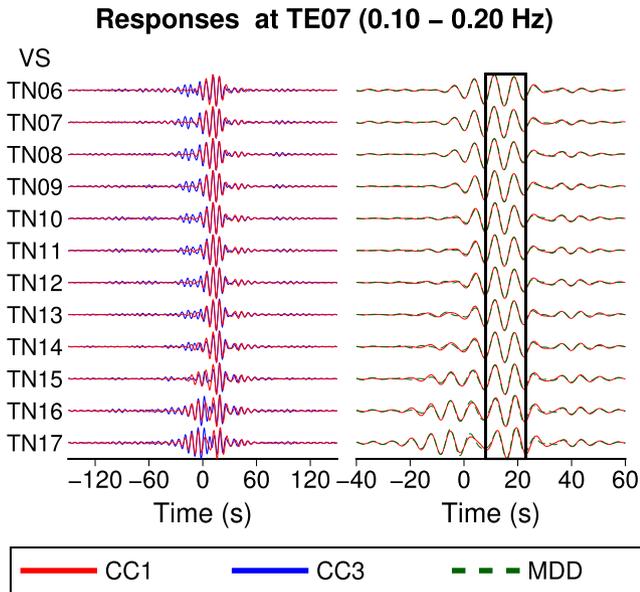
**Figure 10.** Distribution with time and frequency of time windows selected for the application of SI by MDD for receiver TE07. Time windows are counted per day and normalized with respect to the overall maximum. Recall that  $\hat{P}^{(TN)}$  and  $\hat{P}^{(TE)}$  are computed at increments of 0.01 Hz and averaged over a bandwidth of a quarter octave.

between the TN stations, which act as the virtual sources. The PSF is linearly interpolated for offline TN stations. The crosscorrelation function is computed from the crosscorrelations between each of the virtual sources and the TE station at which we aim to retrieve the virtual-source responses. Recall that the frequency dependency of the slowness analysis implies that the CCF and PSF may be associated with different sets of time windows for different frequencies. Similarly, with increasing number of the TE station at which we aim to retrieve the virtual-source responses, the number of selected time windows decreases. Due to the fact that slownesses are not evaluated at each frequency, but at increments of 0.01 Hz only, we smooth the amplitude spectrum over a bandwidth of 0.01 Hz prior to inverse Fourier transformation. This prevents the occurrence of ringing effects, due to the stepwise increase or decrease of the number of selected time windows with frequency (see Supporting Information Fig. S6). All frequency-domain results presented in this section (and the next) are of course unaffected by this smoothing as the inversion in eq. (17) is performed for each frequency individually. In this and the next sections, we mainly present (results based on the) virtual-source responses retrieved at station TE07.

Fig. 12 shows the virtual-source responses retrieved at station TE07 filtered between 0.1 and 0.2 Hz. We again restrict ourselves to the virtual sources TN06–TN17, because the truncation of the PSF on the northern and southern ends of the TN line does not allow a fair comparison between SI by crosscorrelation and SI by MDD for the most northern and southern virtual sources. The absence of time windows associated with a predominant flux of energy towards the west has reduced and in many cases simply eliminated the amplitudes of the acausal parts (CC1 versus CC3). Also, the CC1 responses are significantly more condensed in time than the CC3 responses. Multidimensional deconvolution of the CC1 responses does not appear to have a large effect. Focusing on the main arrival, however, we observe a systematic time shift (Fig. 13). This time shift is virtual-source dependent: the MDD process decreases the arrival time of the main peak of the responses of the southern virtual sources, whereas it slightly delays the arrival of the main peak of the responses of the northern virtual sources. The source functions for virtual source TN15 are depicted in Fig. 14. The smearing of the source function of the crosscorrelation response, quantified by the PSF, is apparent. We also note that the PSF is not symmetric in space



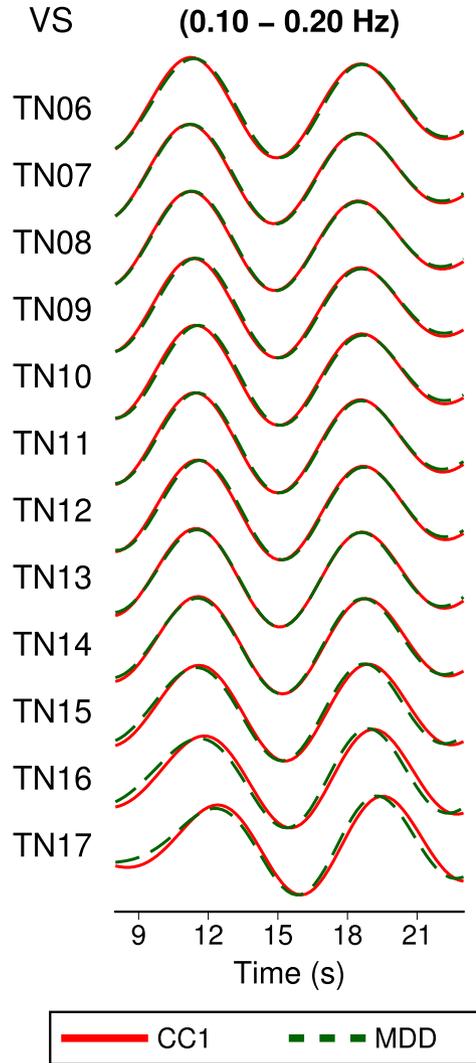
**Figure 11.** Average of  $\Re[\hat{P}^{(TN)}]$  (solid curve) and  $\Re[\hat{P}^{(TE)}]$  (dashed curve) over all time windows with a favourable energy flux, that is, time windows for which  $|\hat{\rho}_y^{(TN)}| < \hat{\rho}_{\text{trh}}^{(TN)}$  and  $\hat{\rho}_x^{(TE)} > \hat{\rho}_{\text{trh}}^{(TE)}$ . The solid grey lines and dashed grey line indicate  $\pm \hat{\rho}_{\text{trh}}^{(TN)}$  and  $\hat{\rho}_{\text{trh}}^{(TE)}$ , respectively.



**Figure 12.** Comparison between the different virtual-source responses. Responses are retrieved at station TE07 and filtered between 0.1 and 0.2 Hz. Note the difference in scale along the time axis between left and right. The area in the black box is shown at a higher resolution in Fig. 13. A comparison of the virtual-source responses filtered between 0.2 and 0.4 Hz is presented in Appendix C (Fig. C3).

with respect to the virtual source, which results in a time shift of the crosscorrelation response with respect to the actual, medium's Green's function (Wapenaar *et al.* 2011b). SI by MDD corrects for the smearing of the source function and hence the time shift, as witnessed by Fig. 13. After MDD, the focused source function (VSF in Fig. 14) is more condensed in space with values close to zero at locations other than the virtual source. We also observe that it is more condensed in time, although the improvement is relatively small.

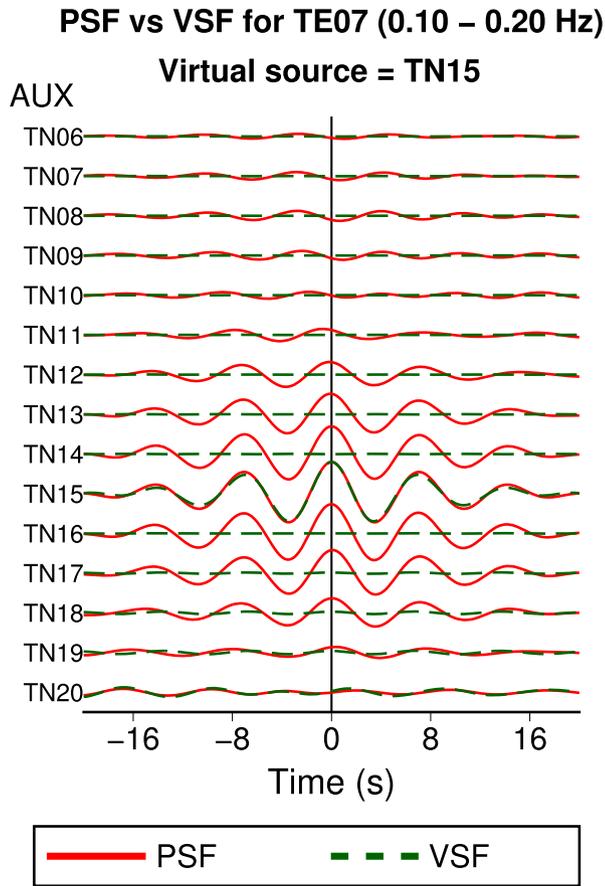
### Responses at TE07 (0.10 – 0.20 Hz)



**Figure 13.** Comparison between the main arrivals of the different virtual-source responses. Responses are retrieved at station TE07 and filtered between 0.1 and 0.2 Hz.

### 5.3 Illumination effects

The time shifts in Fig. 13 correspond to phase shifts in the frequency domain. Fig. 15 depicts the phases of the different virtual-source responses at a frequency of 0.12 Hz. For comparison, the phases of the virtual-source responses of a PREM Earth are depicted. The latter are simply the phases of  $H_0^{(2)}(\omega r/c)$ , where  $\omega = 2\pi \times 0.12$ ,  $r$  the distance between the virtual source and station TE07, and  $c$  the PREM phase velocity at 0.12 Hz (Fig. 4b). The fact that this phase is larger than most of the phases of the retrieved virtual-source responses (CC3, CC1 and MDD) implies that the signal at 0.12 Hz arrives slightly later in the actual medium than it would in a PREM Earth (note that phase increases counter-clockwise, by definition, whereas the Hankel function and the phases of our retrieved responses rotate clockwise with increasing time because of the Fourier transform convention adopted). Fig. 15 shows how the phases of the virtual-source responses change significantly by selecting only time windows during which the wavefield possesses a favourable energy flux (CC1 versus CC3). More explicitly than



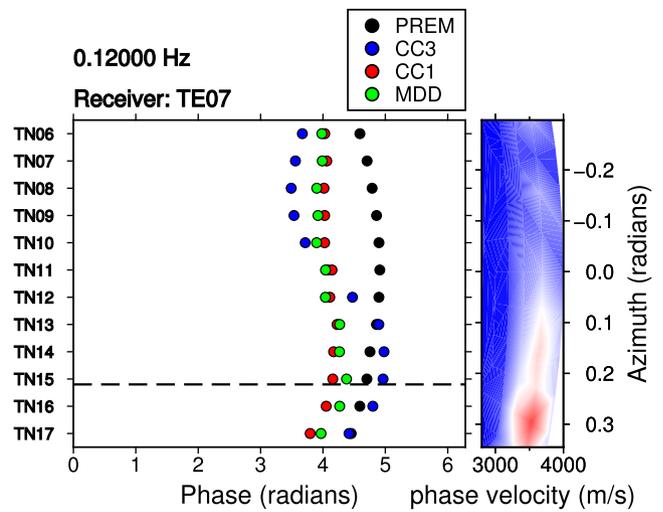
**Figure 14.** Comparison between source function of the crosscorrelation function, that is, the point-spread function, and the source function of the response obtained by SI by MDD, that is, the VSF for virtual source TN15. The PSF and VSF are not shown for stations TN03–TN05 because they did not show a significant signature at those locations.

Fig. 13, it also shows the merit of subsequently multidimensionally deconvolving those responses (MDD versus CC1). The virtual-source responses in line with the dominant noise direction (TN17–TN13) experience a positive phase shift, whereas the MDD enforces a negative phase shift for the responses from the virtual sources in the north. This indeed implies a decrease in traveltime (and hence an increase in phase velocity) for the responses of the southern virtual sources and an increase in traveltime (and hence a decrease in phase velocity) for the responses of the northern virtual sources.

The correction observed in Fig. 15 is the correction that one would expect based on the results of Weaver *et al.* (2009). These authors consider a pair of stations embedded in a homogeneous 2-D medium and illuminated by ballistic waves. The power of the illumination by the ballistic waves is denoted by  $B(\phi)$ , where  $\phi$  denotes the angle, measured positive counter-clockwise, with respect to the line pointing from the receiver at  $\mathbf{x}_R$  towards the virtual source at  $\mathbf{x}_S$ . For an arbitrary (but smooth)  $B(\phi)$  they find that the traveltime error due to the non-uniformity of the illumination pattern is given by

$$\delta t \approx \frac{B''(0)}{2t\omega^2 B(0)}, \quad (28)$$

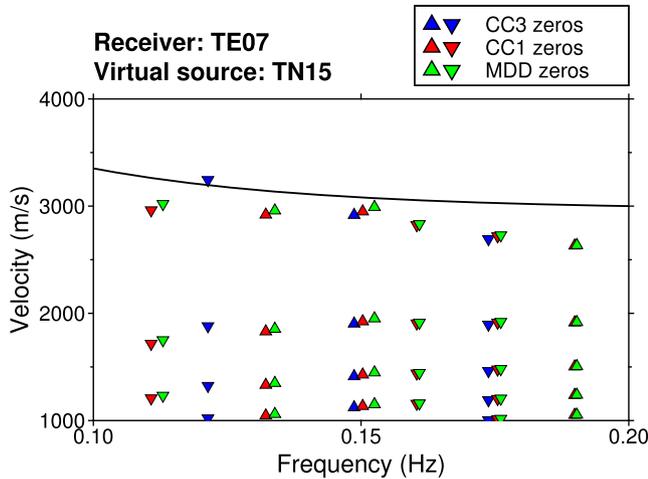
where  $\delta t$  denotes the traveltime error with respect to the traveltime associated with a uniform illumination pattern,  $t$  the measured traveltime and  $B''(0)$  the second derivative of  $B(\phi)$  evaluated at  $\phi = 0$ . A positive error means an arrival time earlier than the true arrival



**Figure 15.** Phase of the different virtual-source responses at TE07 at 0.12 Hz. Phases of the responses obtained by the crosscorrelation of all data (green coloured bars in Fig. 3) are depicted by the blue dots (CC3). Phases of the responses obtained by the crosscorrelation of the data selected by means of a slowness analysis along both station lines are depicted by the red dots (CC1; slownesses of these windows along TE-line are shown in Fig. 9). The phases of the responses obtained by multidimensionally deconvolving the latter responses are depicted by the green dots (MDD). The phases of a surface wave response of a PREM Earth are depicted by the black dots. The dashed black line indicates the location of a virtual source along the TN-line of which the azimuth of the line connecting it with station TE07 coincides with the azimuth of the slowness vector whose components are given by the maxima in Fig. 11. The contour on the right represents a histogram of the number of selected windows per slowness vector as a function of the azimuth and velocity associated with these slowness vectors. Values are normalized with respect to the overall maximum, which implies that dark blue colours represent values close to zero, whereas dark red represents a value of one. The azimuths along the vertical axis coincide with the azimuths of the lines connecting the virtual sources on the left with TE07. This implies that the azimuth is measured positive clockwise from the  $x$ -axis in Fig. 2. The phase of the different virtual-source responses at 0.32 Hz is presented in Appendix C (Fig. C4).

time by an amount of  $\delta t$ . Eq. (28) implies that a maximum of  $B(\phi)$  at  $\phi = 0$  (i.e. a positive second derivative) results in an arrival time later than the true arrival time, whereas a minimum of  $B(\phi)$  at  $\phi = 0$  (i.e. a negative second derivative) results in an arrival time earlier than the true arrival time. The contour plot on the right side of Fig. 15 serves as an estimate of the illumination pattern. The highest flux of energy appears to be parallel to the line connecting virtual source TN16 with stations TE07. According to eq. (28), this would result in a measured traveltime larger than the traveltime associated with the actual medium's Green's function for this virtual source. Fig. 15 shows how MDD decreases this erroneously large traveltime of the responses of the southern virtual sources, whereas it increases the erroneously small traveltime of the responses of the northern virtual sources. It therefore nicely demonstrates the spatial character of the MDD in eq. (17).

The fact that irregularities in the illumination pattern may not only lead to underestimations of the true arrival time but can also lead to overestimations of the true arrival may seem counterintuitive. It is therefore useful to note that this can be understood by considering the phase of the asymptotic approximation of the homogeneous medium's Green's function, that is, the phase of the asymptotic approximation of the Hankel function, and the limiting case of



**Figure 16.** Zeros of  $J_0(\omega_i r/c_n)$  and  $H_0(\omega_i r/c_n)$  for a set of phase velocities  $c_n(\omega_i)$  at those frequencies  $\omega_i$  for which the real and imaginary parts, respectively, of the retrieved virtual-source responses are zero. For comparison the phase-velocity dispersion curve associated with the PREM model is depicted (solid black line).

illumination of the virtual source by a single impulsive source aligned with the receiver and the virtual source, that is,  $B(\phi) = \delta(\phi)$ . At distances larger than about one wavelength, the phase of  $H_0^{(2)}(kr)$  can be approximated by  $-kr + \pi/4$ , which is the phase retrieved in case the medium would be illuminated uniformly from all angles. In the limiting case, however, the phase of the crosscorrelation between the signal recorded by the virtual source and that signal recorded by the receiver at  $\mathbf{x}_R$  is simply given by  $-kr$ , which implies a traveltime larger than the traveltime of the actual medium's Green's function. We finally note that, in general, eq. (28) may be more applicable to lower-frequency surface waves (primary and double frequency peak). Such long period seismic surface waves have limited sensitivity to the shallow, more heterogeneous Earth structure. Consequently, their behaviour is generally better approximated by the homogeneous medium's Green's function than the behaviour of higher-frequency surface waves (e.g. Ekström 2014).

#### 5.4 Phase velocity estimation

We employ the technique described in Section 2.3 to estimate phase-velocity dispersion at frequencies between 0.1 and 0.2 Hz for the response at station TE07 due to a virtual source at TN15. To that end, we determine the linear combination of third-order polynomials (cubic splines) that best fits the behaviour of the real and imaginary parts of the retrieved responses. The polynomials are equally spaced and selected such that their linear combination is sufficiently smooth. Fig. 16 depicts the real and imaginary parts of the different virtual-source responses between 0.1 and 0.2 Hz and their fit. We equate the zeros of the real parts to the zeros of  $J_0(\omega r/c)$ . Additionally, the imaginary parts of the virtual-source responses associated with a one-sided illumination are used (CC1 and MDD). Because the aperture of the T-array is short with respect to the wavelength of the surface waves at those frequencies, we equate the zeros of the imaginary parts to the zeros of  $H_0(\omega r/c)$ , instead of  $Y_0(\omega r/c)$ . The phase-velocity curve that gives the seismologically most plausible dispersion curve is easily determined in this case: it is obtained by connecting the zeros  $c_n$  closest to the PREM dispersion curve with line segments. Fig. 16 shows how the

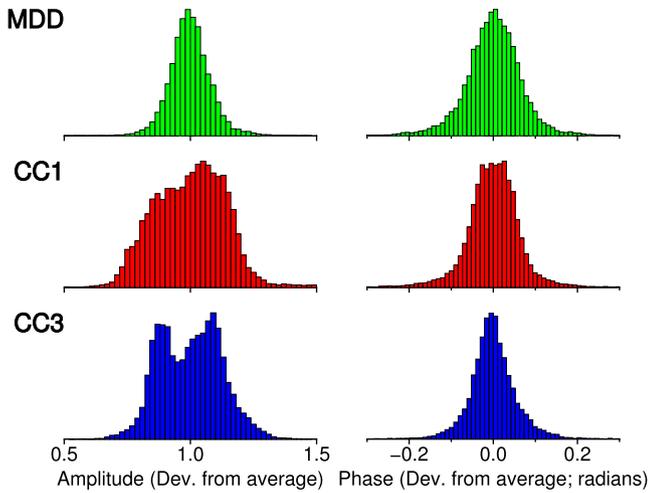
application of SI by MDD gives rise to different zeros and hence an improved estimate of phase-velocity dispersion. Often, realistic velocity values are determined at the longest periods only, after which the dispersion curve is 'tracked' from longer periods to shorter periods, extending it with one zero at a time. Extension to shorter periods stops once it is not possible to unambiguously determine the next  $c_n$  (e.g. Ekström 2014). In the supporting material, we present two additional examples: one where it is straightforward to track the phase-velocity dispersion up to 0.4 Hz (S7) and one where it is impossible to track unambiguously the phase velocity up to frequencies higher than approximately 0.25 Hz (S8).

#### 5.5 Temporal stability

Recently, time-lapse monitoring using time-lapse (henceforth 'lapse') crosscorrelations of ambient seismic surface wave noise has attracted a great deal of attention (e.g. Obermann *et al.* 2014). Such studies generally focus on the coda of the crosscorrelations, where the term coda refers to the later arriving, multiply scattered waves. The reason to use the coda is twofold. First, the multiple scattering implies that these waves have sampled the subsurface very densely and hence are more sensitive to tiny mechanical and structural changes in that subsurface. Second, the coda of the lapse crosscorrelations has proved less sensitive to temporal variations in the illumination pattern than the direct (first) arrival (e.g. Hadziioannou *et al.* 2009; Colombi *et al.* 2014; Lehujeur *et al.* 2015). Reducing the sensitivity of lapse crosscorrelations to the illumination pattern is therefore of immediate interest for such studies. Just as in the previous section, we concentrate on the virtual-source responses retrieved at station TE07.

The low number of time windows that qualify for the application of SI by MDD between 0.15 and 0.2 Hz (see Supporting Information Fig. S6), does not allow the retrieval of stable lapse responses at these frequencies, whereas at frequencies below 0.1 Hz our short-periods sensors cannot recover vertical ground velocity successfully (see Appendix B). Above 0.20 Hz the number of selected windows increases again, but their distribution in time is rather patchy (Fig. 10), rendering the retrieval of stable lapse responses infeasible. Consequently, we restrict our attention to the responses between 0.10 and 0.15 Hz. This is also the frequency band for which the application of SI by MDD to all selected windows appeared to be most successful. The rather narrow frequency band does not lend itself well for analysis in the time domain, that is, after inverse Fourier transformation. We therefore investigate the stability of the retrieved lapse responses in the frequency domain. Note that this implies that both the direct arrival and the coda are included in the evaluated lapse responses.

We separately evaluate the variation of the phase and amplitude of the lapse responses. The phases and amplitudes of the lapse responses retrieved at station TE07 from virtual sources TN06,...,TN17 are considered (only responses between 0.10 and 0.15 Hz though). The lapse CC3 responses are obtained by averaging over a total of 12000 time windows, excluding gaps. This implies that the difference between the start and the end of the total period of time, over which individual responses are averaged, varies between approximately 2.5 months and 5 months [recall that (i) an overlap of 50 per cent has been used, (ii) that data between 10:00 am and 11:00 pm were discarded, and (iii) that larger data gaps exist due to many offline stations]. Lapse responses are computed at increments of 100 time windows and they are placed in time by simply averaging the times associated with the 12 000 individual time



**Figure 17.** Distributions of amplitude (left) and phase (right) of the lapse responses with respect to their mean values. Distributions are computed from all virtual-source responses between 0.10 and 0.15 Hz which are retrieved at station TE07. Amplitudes of the lapse responses are normalized with respect to the mean amplitude for each frequency, virtual source-receiver pair, and method individually. The colours of the histograms correspond to the kind of responses and comply with results presented in the other figures: CC3 is blue, CC1 is red and MDD is green.

windows. The lapse CC1 responses are computed by taking into account only those time windows for which  $|\hat{p}_y^{(TN)}| < \hat{p}_{\text{thr}}^{(TN)}$  (TE07) and  $\hat{p}_x^{(TE)} > \hat{p}_{\text{thr}}^{(TE)}$  (TE07). The lapse MDD responses are in turn obtained by MDD of these lapse CC1 responses with a lapse PSF computed from the same set of selected time windows. In the Supporting Information (Fig. S9) we present the temporal variation of the phase and amplitude at 0.12 Hz of the lapse responses from virtual source TN07.

We have no reason to believe that the subsurface surface wave response at low frequencies might have changed over the course of 2012 (which could be the case, for example, due to a large nearby earthquake; Obermann *et al.* 2014). Therefore, in case the retrieved lapse virtual-source responses would coincide with the actual-medium responses, their phases should be constant over time. We observe that this is not the case for any of the retrieved responses (Fig. 17). In this figure we systematically compare the magnitude of the variations in phase and amplitude between the different methods. The phases and amplitudes of the lapse responses from virtual sources TN06,...,TN17 are used (only responses between 0.10 and 0.15 Hz though). We infer from Fig. 17 that especially the variation in amplitude of the lapse MDD responses is significantly less than that of the lapse CC1 and CC3 responses. Although we do observe a slight improvement in the stability of the MDD responses with respect to the CC1 responses for some virtual source-receiver pairs (e.g. Supporting Information Fig. S9), a systematic improvement is not recognizable in the histograms. We conclude from Fig. 17 that the application of SI by MDD improves the stability of the lapse responses obtained by SI by crosscorrelation.

## 6 CONCLUSIONS

We applied SI by MDD to recordings of vertical particle velocity of ambient seismic noise recorded by a seismic array in Malargüe, Argentina. Using the configuration of this array and assuming a (very simple) laterally homogeneous subsurface, we modelled syn-

thetic data. We show that application of SI by MDD to the synthetic data results in retrieved surface wave responses that approximate the actual-medium responses significantly better than surface wave responses obtained through SI by crosscorrelation. In application to field data, we mainly focused on the phase of the retrieved responses. In case of an anisotropic illumination pattern, we find that SI by MDD improves the phases of responses obtained through SI by crosscorrelation. This improvement is most pronounced for pairs of virtual sources and receivers whose separation is small with respect to the wavelength. We find that the sign of the phase correction resulting from the MDD process coincides with the sign of the phase correction one would expect based on the illumination pattern. Our results have implications for tomographic studies exploiting the ambient seismic field through the application of SI by crosscorrelation. Many of those studies often discard virtual-source responses computed between stations separated by less than a few wavelengths. Our results indicate that the application of SI by MDD allows those virtual-source responses to be corrected and hence the paths associated with those station couples to be included in the tomographic inversion. For our data set, the phase velocities extracted from the retrieved virtual-source responses reveal that, especially at frequencies higher than approximately 0.4 Hz, surface wave velocities increase towards the east. Most likely, this can be attributed to the progressive thinning of the sedimentary infill of the Malargüe basin in that direction.

We find that for station separations larger than a few wavelengths, the application of SI by MDD does not yield much improvement. There can be several reasons for this, most of them associated with the violation of one or more of the assumptions underlying SI by MDD. First, the PSF is assumed to sample the incoming wavefield sufficiently densely. For the shorter wavelengths in this study, however, the nominal station separation of 2 km along the TN-line may give rise to spatial aliasing. Second, too much body-wave energy would deteriorate the effectiveness of SI by MDD for surface waves (Figs C1 and C2 indicate that quite some body-wave energy arrives at 0.32 Hz during the selected time windows). Third, the medium is required to be laterally invariant along the line of receivers used to construct the PSF. The dispersion curves in Fig. 8 suggest that this assumption does not hold at higher frequencies ( $>0.2$  Hz), which adversely affects the quality of the virtual-source responses retrieved through SI by MDD. Finally, the theory assumes the surface wave wavefield to be dominated by a single surface wave mode. Especially between 0.2 and 0.4 Hz, however, multiple surface wave modes cannot be ruled out (e.g. Supporting Information Fig. S5).

Although we did not investigate the amplitude of the retrieved virtual-source responses computed from all (selected) time windows, we did evaluate the stability of the amplitudes of the lapse responses. Limiting ourselves to frequencies between 0.1 and 0.15 Hz, we found that the amplitudes of the responses retrieved through SI by MDD are more stable and hence more precise. We did not observe a significant increase in stability of the phases of the responses retrieved through SI by MDD. We finally note that we experienced that the application of SI by MDD to real seismic noise data is not trivial. Because the spatial deconvolution requires the PSF and CCF to be computed from the same data, interpolation is required in case virtual sources (or auxiliary receivers) are offline. Moreover, and apparent from this study, the requirements on the station geometry, spatial sampling, energy flux and medium may often not be fulfilled. For the application of SI by MDD to recordings of ambient seismic (surface wave) noise, one should therefore assess its potential on an individual basis. This study shows that when (most) conditions are

fulfilled, more accurate phase-velocity estimates and more stable time-lapse responses are obtained.

## ACKNOWLEDGEMENTS

CW is financially supported by the Netherlands Research Centre for Integrated Solid Earth Science (ISES). The research of DD is supported by the Division for Earth and Life Sciences (ALW) with financial aid from the Netherlands Organization for Scientific Research (NWO) with grant VIDI 864.11.009. The authors thank IRIS-PASSCAL for providing the seismic equipment and the Argentine Ministry of Science, Technology and Production Innovation for the financial support connected to the transportation of the equipment. The authors also thank Pierre Auger Observatory and the department of Civil Defense of Malargüe for the help during the data acquisition. Finally, we would like to thank two anonymous reviewers for their constructive comments.

## REFERENCES

- Aki, K., 1957. Space and time spectra of stationary stochastic waves, with special reference to microtremors, *Bull. Earthq. Res. Inst. Univ. Tokyo*, **35**, 415–457.
- Aki, K. & Richards, P.G., 2002. *Quantitative Seismology*, 2nd edn, University Science Books.
- Bakulin, A. & Calvert, R., 2006. The virtual source method: theory and case study, *Geophysics*, **71**(4), S1139–S1150.
- Boschi, L., Weemstra, C., Verbeke, J., Ekström, G., Zunino, A. & Giardini, D., 2013. On measuring surface wave phase velocity from station-station cross-correlation of ambient signal, *Geophys. J. Int.*, **192**, 346–358.
- Campillo, M. & Paul, A., 2003. Long-Range Correlations in the Diffuse Seismic Coda, *Science*, **299**(5606), 547–549.
- Colombi, A., Chaput, J., Brenguier, F., Hillers, G., Roux, P. & Campillo, M., 2014. On the temporal stability of the coda of ambient noise correlations, *C. R. Geosci.*, **346**(11–12), 307–316.
- Draganov, D., Campman, X., Thorbecke, J., Verdel, A. & Wapenaar, K., 2009. Reflection images from ambient seismic noise, *Geophysics*, **74**(5), A63–A67.
- Dziewonski, A.M. & Anderson, D.L., 1981. Preliminary reference earth model, *Phys. Earth planet. Inter.*, **25**(4), 297–356.
- Ekström, G., 2014. Love and Rayleigh phase-velocity maps, 5–40 s, of the western and central USA from USArray data, *Earth planet. Sci. Lett.*, **402**, 42–49.
- Ekström, G., Abers, G. & Webb, S., 2009. Determination of surface-wave phase velocities across USArray from noise and Aki's spectral formulation, *Geophys. Res. Lett.*, **36**(18), L18301, doi:10.1029/2009GL039131.
- Grobbe, N., van der Neut, J. & Vidal, C.A., 2013. Flux-normalized elastodynamic wavefield decomposition using only particle velocity recordings, *SEG Technical Program Expanded Abstracts*, **32**, 4693–4698.
- Gualtieri, L., Stutzmann, E., Capdeville, Y., Arduin, F., Schimmel, M., Mangeney, A. & Morelli, A., 2013. Modelling secondary microseismic noise by normal mode summation, *Geophys. J. Int.* **193**(3), 1732–1745.
- Hadziioannou, C., Larose, E., Coutant, O., Roux, P. & Campillo, M., 2009. Stability of monitoring weak changes in multiply scattering media with ambient noise correlation: laboratory experiments, *J. acoust. Soc. Am.*, **125**(6), 3688–3695.
- Halliday, D. & Curtis, A., 2008. Seismic interferometry, surface waves and source distribution, *Geophys. J. Int.*, **175**(3), 1067–1087.
- Kimman, W.P. & Trampert, J., 2010. Approximations in seismic interferometry and their effects on surface waves, *Geophys. J. Int.*, **182**(1), 461–476.
- Kimman, W.P., Campman, X. & Trampert, J., 2012. Characteristics of seismic Noise: fundamental and higher mode energy observed in the northeast of the Netherlands, *Bull. seism. Soc. Am.*, **102**(4), 1388–1399.
- Kraemer, P., Silvestro, J., Achilli, F. & Brinkworth, W., 2011. Kinematics of a Hybrid Thick-thin-skinned Fold and Thrust Belt Recorded in Neogene Syntectonic Wedge-top Basins, Southern Central Andes Between 35 and 36 S, Malargüe, Argentina, in *Thrust Fault-Related Folding*, chap. 11, p. 401, eds McClay, K., Shaw, J. & Suppe, J., American Association of Petroleum Geologists.
- 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**(B05), B05302, doi:10.1029/2009JB006918.
- Lehujeur, M., Vergne, J., Schmittbuhl, J. & Maggi, A., 2015. Characterization of ambient seismic noise near a deep geothermal reservoir and implications for interferometric methods: a case study in northern Alsace, France, *Geotherm. Energy*, **3**.
- Longuet-Higgins, M.S., 1950. A theory of the origin of microseisms, *Phil. Trans. R. Soc. A*, **243**(857), 1–35.
- Malcolm, A.E., Scales, J.A. & Van Tiggelen, B.A., 2004. Extracting the Green function from diffuse, equipartitioned waves, *Phys. Rev. E*, **70**, 015601, doi:10.1103/PhysRevE.70.015601.
- McNamara, D.E. & Buland, R.P., 2004. Ambient Noise Levels in the Continental United States, *Bull. seism. Soc. Am.*, **94**(4), 1517–1527.
- Mulargia, F., 2012. The seismic noise wavefield is not diffuse, *J. acoust. Soc. Am.*, **131**(4), 2853–2858.
- Nishitsuji, Y., Ruigrok, E., Gomez, M. & Draganov, D., 2014. Global-phase H/V spectral ratio for delineating the basin in the Malargüe Region, Argentina, *Seismol. Res. Lett.*, **85**(5), 1004–1011.
- Obermann, A., Froment, B., Campillo, M., Larose, E., Planès, T., Valette, B., Chen, J.H. & Liu, Q.Y., 2014. Seismic noise correlations to image structural and mechanical changes associated with the Mw 7.9 2008 Wenchuan earthquake, *J. geophys. Res.*, **119**, 3155–3168.
- Park, C.B., Miller, R.D. & Xia, J., 1998. Imaging dispersion curves of surface waves on multi-channel record, *SEG Expanded Abstracts*, **17**(1), 1377–1380.
- Prieto, G.A., Lawrence, J.F. & Beroza, G.C., 2009. Anelastic Earth structure from the coherency of the ambient seismic field, *J. geophys. Res.*, **114**(B07), B07303, doi:10.1029/2008JB006067.
- Ruigrok, E. et al., 2012. Malargüe seismic array: design and deployment of the temporary array, *Eur. Phys. J. Plus*, **127**(10), 126, doi:10.1140/epjp/i2012-12126-7.
- Salvermoser, J., Hadziioannou, C. & Stähler, S., 2015. Structural monitoring of a highway bridge using passive noise recordings from street traffic, *J. acoust. Soc. Am.*, **138**(6), 3864–3872.
- Schuster, G.T., Yu, J., Sheng, J. & Rickett, J., 2004. Interferometric/daylight seismic imaging, *Geophys. J. Int.*, **157**(2), 838–852.
- Sens-Schönfelder, C. & Wegler, U., 2006. Passive image interferometry and seasonal variations of seismic velocities at Merapi Volcano, Indonesia, *Geophys. Res. Lett.*, **33**(21), L21302, doi:10.1029/2006GL027797.
- Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, **31**(7), 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.
- Snieder, R., 2004. Extracting the Green's function from the correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E*, **69**(4), 046610, doi:10.1103/PhysRevE.69.046610.
- Snieder, R. & Fleury, C., 2010. Cancellation of spurious arrivals in Green's function retrieval of multiple scattered waves, *The J. acoust. Soc. Am.*, **128**(4), 1598–605.
- Snieder, R., van Wijk, K., Haney, M. & Calvert, R., 2008. Cancellation of spurious arrivals in Green's function extraction and the generalized optical theorem, *Phys. Rev. E*, **78**(3), 36606, doi:10.1103/PhysRevE.78.036606.
- Stehly, L., Campillo, M. & Shapiro, N.M., 2006. A study of the seismic noise from its long-range correlation properties, *J. geophys. Res.*, **111**(B10), B10306, doi:10.1029/2005JB004237.
- 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.
- Tsai, V.C., 2011. Understanding the amplitudes of noise correlation measurements, *J. geophys. Res.*, **116**(B09), B09311, doi:10.1029/2011JB008483.

- van Dalen, K.N., Wapenaar, K. & Halliday, D.F., 2014. Surface wave retrieval in layered media using seismic interferometry by multidimensional deconvolution, *Geophys. J. Int.*, **196**, 230–242.
- van Dalen, K.N., Mikesell, T.D., Ruigrok, E.N. & Wapenaar, K., 2015. Retrieving surface waves from ambient seismic noise using seismic interferometry by multidimensional deconvolution, *J. geophys. Res.*, **120**, 944–961.
- van Manen, D.-J., Curtis, A. & Robertsson, J.O.A., 2006. Interferometric modeling of wave propagation in inhomogeneous elastic media using time reversal and reciprocity, *Geophysics*, **71**(4), SI47–SI60.
- van der Neut, J., Thorbecke, J., Mehta, K., Slob, E. & Wapenaar, K., 2011. Controlled-source interferometric redatuming by crosscorrelation and multidimensional deconvolution in elastic media, *Geophysics*, **76**(4), SA63–SA76.
- Wapenaar, K. & Fokkema, J., 2006. Green's function representations for seismic interferometry, *Geophysics*, **71**(4), SI33–SI46.
- Wapenaar, K. & van der Neut, J., 2010. A representation for Green's function retrieval by multidimensional deconvolution, *The J. acoust. Soc. Am.*, **128**(6), EL366–EL371.
- Wapenaar, K., Ruigrok, E., van der Neut, J. & Draganov, D., 2011a. Improved surface-wave retrieval from ambient seismic noise by multi-dimensional deconvolution, *Geophys. Res. Lett.*, **38**(1), L01313, doi:10.1029/2010GL045523.
- Wapenaar, K., van der Neut, J., Ruigrok, E., Draganov, D., Hunziker, J., Slob, E., Thorbecke, J. & Snieder, R., 2011b. Seismic interferometry by crosscorrelation and by multidimensional deconvolution: a systematic comparison, *Geophys. J. Int.*, **185**(3), 1335–1364.
- Wapenaar, K., van der Neut, J. & Thorbecke, J., 2012. On the relation between seismic interferometry and the simultaneous-source method, *Geophys. Prospect.*, **60**(4), 802–823.
- 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.
- Weemstra, C., Boschi, L., Goertz, A. & Artman, B., 2013. Seismic attenuation from recordings of ambient noise, *Geophysics*, **78**(1), Q1–Q14.
- Weemstra, C., Snieder, R. & Boschi, L., 2015. On the estimation of attenuation from the ambient seismic field: inferences from distributions of isotropic point scatterers, *Geophys. J. Int.*, **203**, 1054–1071.
- Yao, H., van der Hilst, R.D. & de Hoop, M.V., 2006. Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis - I. Phase velocity maps, *Geophys. J. Int.*, **166**, 732–744.
- Yokoi, T. & Margaryan, S., 2008. Consistency of the spatial autocorrelation method with seismic interferometry and its consequence, *Geophys. Prospect.*, **56**(3), 435–451.
- Zhan, Z., Ni, S., HelMBERGER, D.V. & Clayton, R.W., 2010. Retrieval of Moho-reflected shear wave arrivals from ambient seismic noise, *Geophys. J. Int.*, **182**, 408–420.

## SUPPORTING INFORMATION

Supplementary data are available at [GJIRAS](http://gji.oxfordjournals.org/) online.

**Figure S1.** Comparison of the different methods and/or data sets (abbreviations explained in the main text) using synthetic data.

**Figure S2.** ( $f$ ;  $k$ )-transform computed from virtual-source responses at the receivers of the TE-array using SI by crosscorrelation.

**Figure S3.** ( $f$ ;  $k$ )-transforms computed from virtual-source responses of two different subsets of receivers of the TE-array.

**Figure S4.** ( $f$ ;  $k$ )-transform computed from virtual-source responses at the receivers of the TN-array using SI by crosscorrelation.

**Figure S5.** ( $f$ ;  $k$ )-transforms computed from virtual-source responses of two different subsets of receivers of the TN-array.

**Figure S6.** Number of time windows selected for the application of SI by MDD for receiver TE07.

**Figure S7.** Zeros of  $J_0(\omega_i r/c_n)$  and  $\mathbf{H}_0(\omega_i r/c_n)$  for a set of phase velocities  $c_n(\omega_i)$  at those frequencies  $\omega_i$  for which the real and

imaginary parts of the retrieved virtual-source response are zero, respectively.

**Figure S8.** Zeros of  $J_0(\omega_i r/c_n)$  and  $\mathbf{H}_0(\omega_i r/c_n)$  for a set of phase velocities  $c_n(\omega_i)$  at those frequencies  $\omega_i$  for which the real and imaginary parts of the retrieved virtual-source response are zero, respectively.

**Figure S9.** Temporal variations of the phase (bottom) and amplitude (top) of the lapse crosscorrelations at 0.12 Hz for virtual source TN07 and receiver TE07.

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

## APPENDIX A: SIMPLIFICATION OF THE REPRESENTATION THEOREMS

In this appendix we introduce the correlation- and convolution-type representation theorems and show how they can be simplified for our purpose. In Appendix A1 we simplify the correlation-type representation theorem for the application of SI by crosscorrelation. In Appendix A2 we simplify the convolution-type representation theorem for the application of SI by MDD.

### A1 Simplification of the representation theorem of the correlation type

Consider a volume  $\mathbb{V}$  enclosed by an arbitrarily shaped boundary  $\mathbb{S}$  with outward pointing normal vector  $\mathbf{n} = (n_1, n_2, n_3)$ . In  $\mathbb{V}$ , the elastodynamic medium can be arbitrarily inhomogeneous and anisotropic, with the physical quantities describing its response being a function of the Cartesian coordinate vector  $\mathbf{x} = (x_1, x_2, x_3)$  and angular frequency  $\omega$ ;  $x_3$  is positive downward. For brevity, we indicate the frequency dependency of a variable by a hat, e.g.  $\hat{G} \equiv \hat{G}(\omega)$ . Assuming the medium in  $\mathbb{V}$  is lossless, a correlation-type representation theorem for the Green's function between two points  $\mathbf{x}$  and  $\mathbf{x}_R$ , both in  $\mathbb{V}$ , can be obtained (van Manen *et al.* 2006; Wapenaar & Fokkema 2006):

$$\begin{aligned} \hat{G}_{im}^u(\mathbf{x}_R, \mathbf{x}) - \hat{G}_{im}^{u*}(\mathbf{x}_R, \mathbf{x}) \\ = \oint_{\mathbb{S}} \left[ \hat{G}_{in}^u(\mathbf{x}_R, \mathbf{x}_S) n_j c_{njkl}(\mathbf{x}_S) \partial_k^S \hat{G}_{ml}^{u*}(\mathbf{x}, \mathbf{x}_S) \right. \\ \left. - n_j c_{njkl}(\mathbf{x}_S) \partial_k^S \hat{G}_{il}^u(\mathbf{x}_R, \mathbf{x}_S) \hat{G}_{mn}^{u*}(\mathbf{x}, \mathbf{x}_S) \right] dx_S. \end{aligned} \quad (\text{A1})$$

A Green's function  $\hat{G}_{im}^u(\mathbf{x}_R, \mathbf{x})$  represents particle displacement (hence the superscript  $u$ ) in the  $i$  direction at  $\mathbf{x}_R$  due to an impulsive force in the  $m$  direction at  $\mathbf{x}$ . Summation is invoked over repeated indices. The asterisk denotes complex conjugation and hence the products on the right-hand side correspond to crosscorrelations in the time domain. The integral is over the coordinate  $\mathbf{x}_S$  on the boundary  $\mathbb{S}$  and  $c_{njkl}(\mathbf{x}_S)$  denotes the stiffness of the medium along that boundary. The spatial derivatives act at  $\mathbf{x}_S$  and hence a term  $n_j c_{njkl} \partial_k^S \hat{G}_{il}^u(\mathbf{x}_R, \mathbf{x}_S)$  represents the displacement at  $\mathbf{x}_R$  due to a deformation-rate (dipole) source at  $\mathbf{x}_S$ . Eq. (A1) states that the Green's function between two arbitrary points inside  $\mathbb{V}$  can be retrieved if both the stiffness along  $\mathbb{S}$  and the displacement (and its spatial derivative) along  $\mathbb{S}$ , due to an (impulsive) source at each of these two points, are known. The left-hand side is the Fourier transform of  $G_{im}^u(\mathbf{x}_R, \mathbf{x}, t) - G_{im}^u(\mathbf{x}_R, \mathbf{x}, -t)$ .

The correlation-type Green's function representation has been scrutinized for surface wave Green's functions travelling along the surface of laterally homogeneous, isotropic media by Halliday & Curtis (2008). Among other things, these authors investigate the effect of the source types. Importantly, they find that cross-modal terms may arise in case only point sources of force exist on  $\mathbb{S}$ . The authors therefore propose to separate different surface wave modes prior to crosscorrelation. In this work, we assume that a single surface wave mode dominates the ambient surface wave wavefield, which renders the separation of different surface wave modes unnecessary.

Assuming only point sources of force exist on  $\mathbb{S}$ , the correlation-type Green's function representation associated with a single surface wave mode can be approximated by (Halliday & Curtis 2008; Kimman & Trampert 2010),

$$\begin{aligned} \hat{G}_{im}^u(\mathbf{x}_R, \mathbf{x}) - \hat{G}_{im}^{u*}(\mathbf{x}_R, \mathbf{x}) \\ \approx -i\omega\hat{A}\oint_{\mathbb{S}} \hat{G}_{in}^u(\mathbf{x}_R, \mathbf{x}_S) \hat{G}_{mn}^{u*}(\mathbf{x}, \mathbf{x}_S) d\mathbf{x}_S, \end{aligned} \quad (\text{A2})$$

where  $\hat{A}$  is a mode-and frequency-dependent scale factor and where the (non-index)  $i$  denotes the imaginary unit. The modal scale factor does not affect the phase of the retrieved Green's function.

Eq. (A2) still requires sources at depth and also allows  $\mathbf{x}_R$  and  $\mathbf{x}$  to be at depth. In this work, we restrict ourselves to vertical-component recordings of the particle velocity at the Earth's surface. Using the Fourier convention  $\hat{f}(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t} dt$ , we therefore consider the vertical component of the Green's tensor for particle velocity, that is,  $\hat{G}_{3m}^v \equiv i\omega\hat{G}_{3m}^u$ . Provided the subsurface is laterally homogeneous and only surface wave responses are present,  $\hat{G}_{3m}^v$  will only contain fluctuations due to Rayleigh waves. In practice, this is of course not the case, but for the frequencies considered (microseism frequency band) it is a good approximation. Furthermore, Rayleigh waves induced by the coupling of the oceans with the solid Earth are predominantly excited by vertical forcing (Gualtieri *et al.* 2013). We therefore consider  $\hat{G} \equiv \hat{G}_{33}^v (\equiv i\omega\hat{G}_{33}^u)$ , where the indices and superscript  $v$  are dropped to reduce notational burden. Eq. (A2) can hence be simplified to

$$\begin{aligned} \hat{G}(\mathbf{x}_R, \mathbf{x}) + \hat{G}^*(\mathbf{x}_R, \mathbf{x}) \\ \approx \hat{B}\oint_{\mathbb{S}} \hat{G}(\mathbf{x}_R, \mathbf{x}_S) \hat{G}^*(\mathbf{x}, \mathbf{x}_S) d\mathbf{x}_S, \end{aligned} \quad (\text{A3})$$

where  $\hat{B}$  is a scale factor different from  $\hat{A}$ , due to the fact that we have neglected sources at depth. Note that  $\mathbf{x}_R, \mathbf{x}$ , as well as the contour  $\mathbb{S}$  are now all confined to the surface. The left-hand side is the Fourier transform of  $G_{33}^v(\mathbf{x}_R, \mathbf{x}, t) + G_{33}^v(\mathbf{x}_R, \mathbf{x}, -t)$ . This approximation complies with practical applications of SI and is quite accurate in case sources are situated in the far-field (Wapenaar & Fokkema 2006). A gap in the surface wave illumination pattern of the receivers, however, implies that the boundary  $\mathbb{S}$  is not closed. Halliday & Curtis (2008) show that this may severely hinder the retrieval of  $\hat{G}(\mathbf{x}_R, \mathbf{x}) + \hat{G}^*(\mathbf{x}_R, \mathbf{x})$ .

## A2 Simplification of the representation theorem of the convolution type

Let us consider a configuration different from the one associated with eq. (A1). Assume  $\mathbf{x}$  to be outside  $\mathbb{S}$  and rename it  $\mathbf{x}_S$ ;  $\mathbf{x}_R$  remains in  $\mathbb{V}$ . For this configuration a convolution-type representa-

tion theorem for the Green's function between  $\mathbf{x}_R$  and  $\mathbf{x}_S$  can be obtained (Aki & Richards 2002; van Dalen *et al.* 2014),

$$\begin{aligned} \hat{G}_{im}^u(\mathbf{x}_R, \mathbf{x}_S) \\ = \oint_{\mathbb{S}} \left[ \hat{G}_{in}^u(\mathbf{x}_R, \mathbf{x}) n_j c_{njkl}(\mathbf{x}) \partial_k \hat{G}_{lm}^u(\mathbf{x}, \mathbf{x}_S) \right. \\ \left. - n_j c_{njkl}(\mathbf{x}) \partial_k \hat{G}_{il}^u(\mathbf{x}_R, \mathbf{x}) \hat{G}_{nm}^u(\mathbf{x}, \mathbf{x}_S) \right] d\mathbf{x}. \end{aligned} \quad (\text{A4})$$

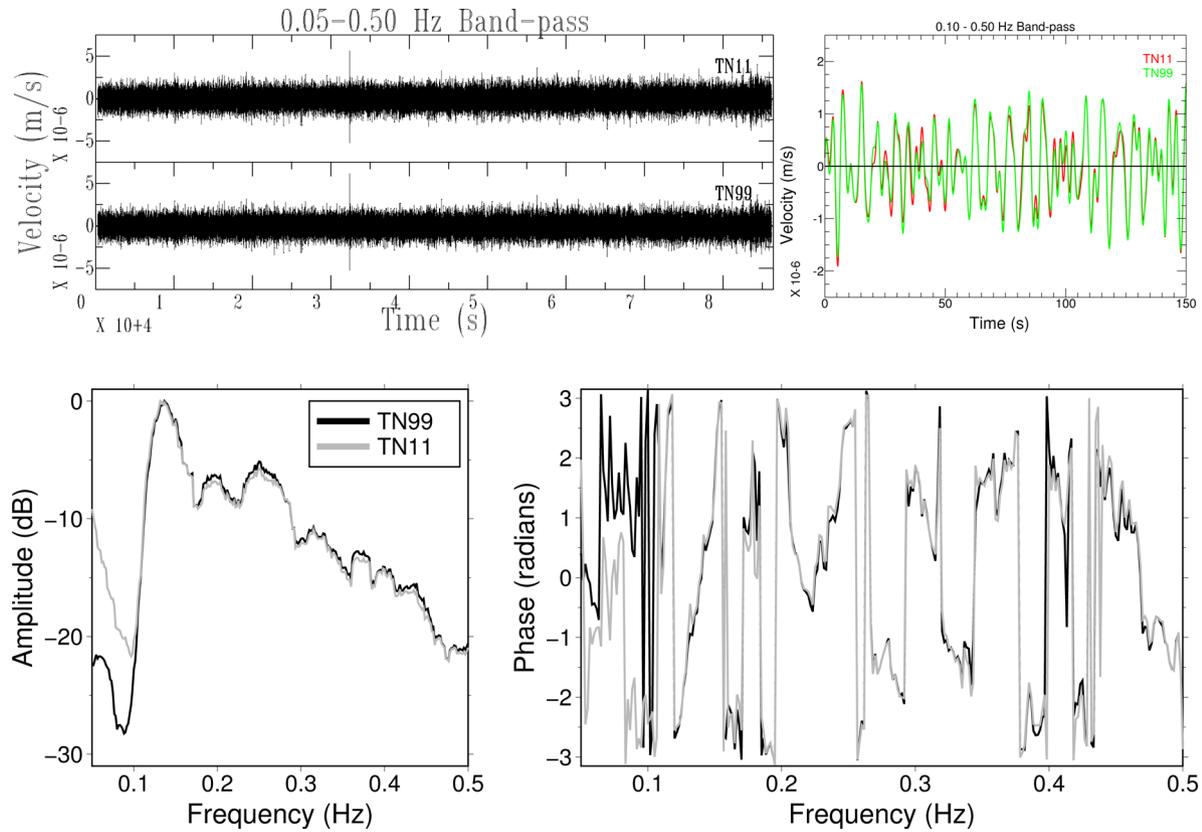
The Green's functions are again force-source Green's functions for particle displacement. The coordinates on the integration surface  $\mathbb{S}$  are denoted by  $\mathbf{x}$  in this case. The spatial derivatives, therefore, act at  $\mathbf{x}$  and hence a term  $n_j c_{njkl} \partial_k \hat{G}_{il}^u(\mathbf{x}_R, \mathbf{x})$  represents the displacement at  $\mathbf{x}_R$  due to a deformation-rate (dipole) source at  $\mathbf{x}$ . A Green's function  $\hat{G}_{in}^u(\mathbf{x}_R, \mathbf{x})$  is defined in a 'reference medium' that may be different from the actual medium on and outside  $\mathbb{S}$  (hence the overbar). These Green's functions propagate the wavefield from the boundary to  $\mathbf{x}_R$ . Since complex conjugate signs are absent on the right-hand side, the products correspond to convolutions in the time domain. Eq. (A4) implies that the Green's function between two arbitrary points  $\mathbf{x}_R$  and  $\mathbf{x}_S$ , the former inside  $\mathbb{S}$ , the latter outside  $\mathbb{S}$ , can be retrieved if both the stiffness and the displacement (and its spatial derivative) along  $\mathbb{S}$ , due to an (impulsive) source at each of these two points, are known. After simplification of the integrand, however, it is the spatial derivative of the Green's function associated with the reference medium that will be retrieved through SI by MDD; the Green's functions between  $\mathbf{x}_S$  and  $\mathbf{x}_R$ , and  $\mathbf{x}_S$  and  $\mathbf{x}$  are related to the observations. Unlike the correlation-type representation theorem, eq. (A4) does not rely on the principle of time-reversal. Importantly, therefore, the convolution-type representation theorem is also valid for media with losses.

The convolution-type representation theorem is examined by van Dalen *et al.* (2014) considering exclusively surface wave Green's functions in a laterally homogeneous and isotropic medium. The reference Green's function and the Green's function associated with the actual medium coincide for this specific medium. The analysis is similar to the one by Halliday & Curtis (2008) for the correlation-type representation theorem in the sense that it approximates the integral in eq. (A4) by considering only those points for which the integrand's phase is stationary with respect to  $\mathbf{x}$ . And just as for the correlation-type representation theorem, these authors find that replacing dipole sources by monopole sources along  $\mathbb{S}$  may give rise to cross-modal terms. Considering a single surface wave mode, however, they show that the convolution-type Green's function representation can be approximated by

$$\begin{aligned} \hat{G}_{im}^u(\mathbf{x}_R, \mathbf{x}_S) \\ = i\omega\hat{A}\oint_{\mathbb{S}} \hat{G}_{in}^u(\mathbf{x}_R, \mathbf{x}) \hat{G}_{nm}^{u(\text{in})}(\mathbf{x}, \mathbf{x}_S) d\mathbf{x}. \end{aligned} \quad (\text{A5})$$

The superscript (in) indicates that only inward-propagating energy associated with  $\hat{G}_{nm}^u(\mathbf{x}, \mathbf{x}_S)$  is considered (i.e. waves propagating into  $\mathbb{V}$  through  $\mathbb{S}$ ). This ensures that non-physical contributions associated with stationary points  $\mathbf{x}$  on the side of  $\mathbf{x}_R$  opposite to  $\mathbf{x}_S$  vanish. These contributions are a direct consequence of the replacement of the dipole sources by monopole sources.

Like eq. (A2), eq. (A5) generally does not comply well with reality. For example, the fact that responses are required at depth along  $\mathbb{S}$  does not reconcile with the data analysed in this study. Under the same assumptions and for the same definitions as the



**Figure B1.** Comparison between the vertical-component recordings by TN99 (broad-band) and TN11 (short-period). Instrument responses are removed prior to comparison. Time-series (top left) represent the vertical particle velocity on 2012 May 1. At the bottom, power and phase spectra are compared (left and right, respectively). Power spectra are averages of individual power spectra computed from individual time windows with a length of 10 min (144 in total). For display purposes, these power spectra are subsequently smoothed over a bandwidth of 0.02 Hz. Similarly, the phases at the bottom right are computed from the complex spectra after these have been averaged over 144 individual 10 min time windows and 0.02 Hz. At the top right, the vertical particle velocity is compared for a period of 2.5 min, starting at 01:00:00 UTC on the same day.

ones required to arrive at eq. (A3), eq. (A5) can be simplified to

$$\hat{G}(\mathbf{x}_R, \mathbf{x}_S) = \hat{D} \oint_{\mathbb{S}} \hat{G}(\mathbf{x}_R, \mathbf{x}) \hat{G}^{(in)}(\mathbf{x}, \mathbf{x}_S) d\mathbf{x}, \quad (\text{A6})$$

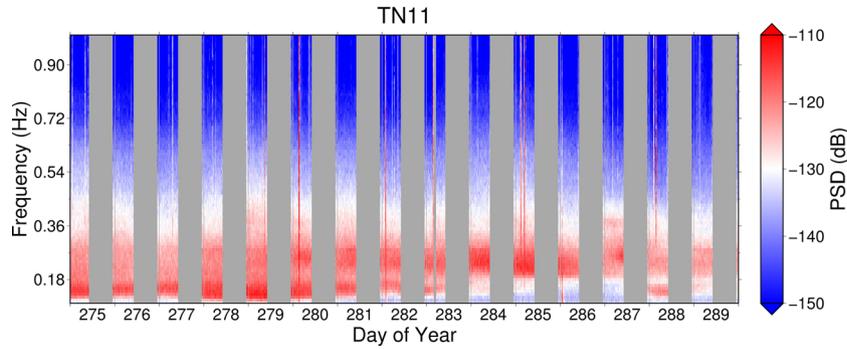
where  $\hat{D}$  is another scale factor, different from both  $\hat{A}$  and  $\hat{B}$ . The locations  $\mathbf{x}_R$ ,  $\mathbf{x}_S$ , as well as the contour  $\mathbb{S}$  are now all confined to the surface. Eqs (A3) and (A6) are strictly valid only for single surface wave modes.

## APPENDIX B: DATA CHARACTERISTICS

We use a full-day vertical-component recording by TN99 to assess the quality of the long-period ambient-noise signal recorded by the vertical component of TN11. We aim to exploit ambient seismic surface waves at frequencies dominated by microseism energy and therefore focus on frequencies between 0.05 and 0.5 Hz. Fig. B1 juxtaposes the vertical component ambient seismic noise recorded by TN99 and TN11. The noise was recorded on the first of May (day randomly chosen) and the recordings have been corrected for the instrument response. The band-pass-filtered time series (Fig. B1, top left) appear, qualitatively at least, rather similar. This observation is corroborated by the power and amplitude spectra (Fig. B1, bottom left and right, respectively): Down to about 0.1 Hz these spectra

agree very well. Despite the fact that the amplitude response of the short-period sensor falls off at about 6 dB per octave below 2 Hz, the ambient seismic field is successfully recovered down to 0.1 Hz. Below this frequency, the amplitude of the ambient field does not fall within the dynamic range of the short-period sensors (the apparent higher power of TN11's noise below 0.1 Hz is simply due to amplified self-noise). This becomes evident after comparing the phase responses, which exhibit a clear mismatch below this frequency, and is confirmed upon comparison of a short period of recording (2.5 min), which is filtered between 0.1 and 0.5 Hz (Fig. B1, top right). All in all, this comparison demonstrates that the amplitude of the ambient seismic field and the dynamic range of the short-period sensors permits the application of seismic interferometry down to about 0.1 Hz.

Upon scrutiny of the data, it turned out that the recordings of a significant number of stations contained diurnal spikes of an unknown nature in the frequency band of interest. Data recorded before 10:00 am local time and after 11:00 pm local time, however, appeared to be (almost) unaffected. We therefore discard the data recorded between those two times and proceed with the remainder of the recordings. Fig. B2 presents the power spectral densities of the recordings of station TN11 for a period of 15 d. The spectrogram is exemplary for the entire data set: the microseisms' energy is clearly distinguishable and its peak varies between 0.10 and 0.30 Hz. The vertical stripes of high energy are due to intermittent earthquakes.



**Figure B2.** Power spectra between 0.1 and 1.0 Hz computed from the recordings of station TN11 for the first half of October 2012. Dark grey vertical bands correspond to periods of time during which the recordings by a number of stations contained spikes of unidentified nature.

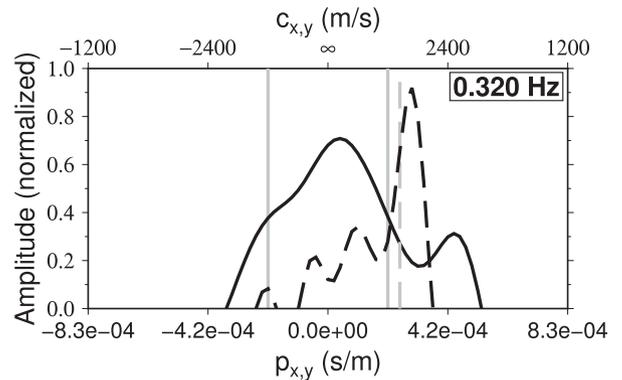
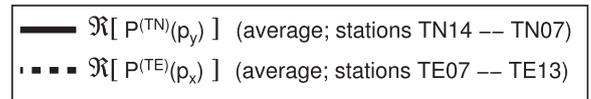
**APPENDIX C: APPLICATION TO AMBIENT VIBRATIONS ABOVE 0.20 Hz**

In this appendix, we present the results of both the slowness analysis and the application of SI by MDD for ambient signal above 0.2 Hz. In the main text the results for ambient signal below 0.2 Hz are presented. Of course, just as in Section 5, the length of the time window employed is 10 min and an overlap of 50 per cent between consecutive windows is used.

**C1 Slowness analysis**

Fig. C1 present the result of the slowness analysis at 0.32 Hz. Just as in Fig. 9 in the main text,  $\Re[\hat{P}^{(TE)}]$  (bottom) is only shown for time windows during which the energy flux is favourable for the retrieval of a virtual-source response at station TE07. Contrary to the seismic noise at 0.12 Hz, most energy is propagating in a northerly direction along the TN-line. This agrees with the higher amplitude causal part of the virtual-source responses along the TN-line in the frequency band 0.20–0.40 Hz (Fig. 7 in the main text). We observe that just as at 0.12 Hz, significant body-wave energy appears to be present. We furthermore observe that the selected time windows are not distributed uniformly with time but have a rather irregular temporal distribution.

Fig. C2 presents the average  $\Re[\hat{P}^{(TN)}]$  (solid line) and  $\Re[\hat{P}^{(TE)}]$  (dashed line) at 0.32 Hz. It reveals that the selected time windows are, on average, associated with recordings of ambient surface wave noise with a positive y component. The average of  $\Re[\hat{P}^{(TE)}]$  peaks of course just above  $\hat{p}_x^{(TE)} = \hat{p}_{\text{trh}}^{(TE)}$ , but does also show some energy

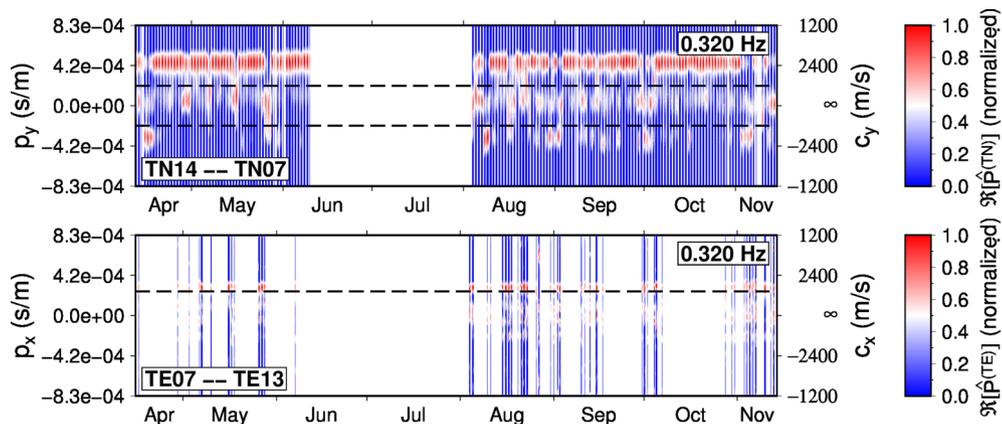


**Figure C2.** Idem as Fig. 11 in the main text, but at 0.32 Hz.

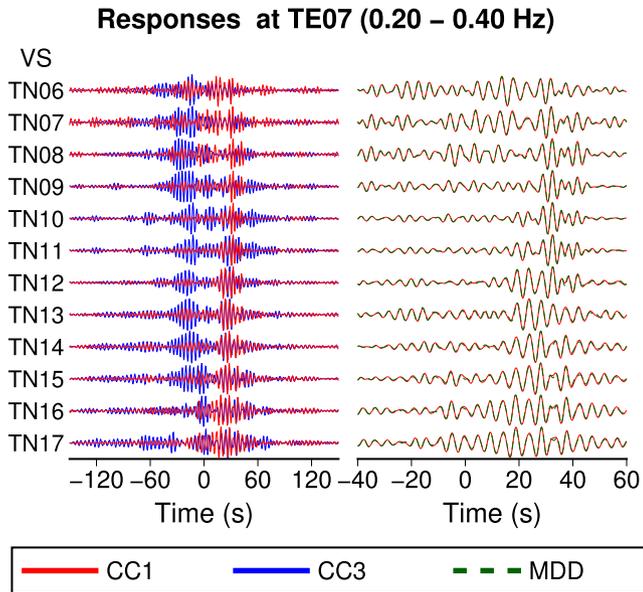
at lower  $\hat{p}_x^{(TE)}$ , which most likely can be attributed to body waves hitting the T-array from below.

**C2 Results**

Fig. C3 shows the virtual-source responses retrieved at station TE07 filtered between 0.2 and 0.4 Hz. The processing involved for the retrieval of these responses is described in Section 5.2. Just as for the virtual-source responses between 0.1 and 0.2 Hz, we have restricted ourselves to the virtual sources TN06–TN17. The absence



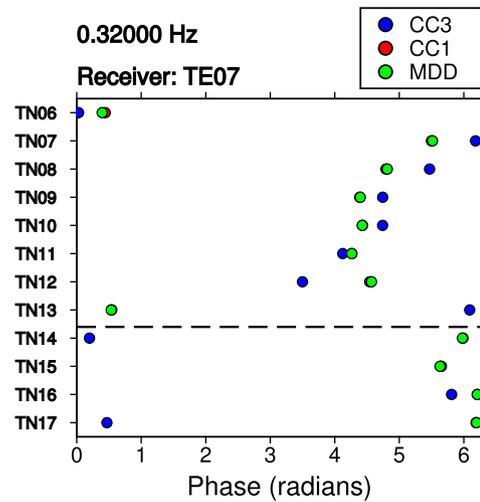
**Figure C1.** Idem as Fig. 9 in the main text, but at 0.32 Hz.



**Figure C3.** Comparison between the different virtual-source responses. Responses are retrieved at station TE07 and filtered between 0.2 and 0.4 Hz. Note the difference in scale along the time axis between left and right.

of time windows associated with a predominant flux of energy towards the west has reduced and in many cases simply eliminated the amplitudes of the acausal parts (CC1 versus CC3). Also, the CC1 responses are significantly more condensed in time than the CC3 responses. Multidimensional deconvolution of the CC1 responses does not appear to have an effect. Contrary to the 0.1–0.2 Hz frequency band, we do not observe a systematic time shift after application of SI by MDD.

Fig. C4 shows the phases of the virtual-source responses at 0.32 Hz. Contrary to Fig. 15 in the main text, we do not show a histogram of the number of selected windows per slowness vector. This is because, at this frequency, slownesses associated with the direction of the highest energy flux have not been determined using all stations, but only using subsets of stations. Such a histogram therefore does not give a correct estimate of the average directionality of the wavefield during the selected time windows. The phase difference between CC1 and MDD responses is almost negligible at this frequency. In the time domain, differences between these virtual-source responses between 0.20 and 0.40 Hz cannot even be observed (Fig. C3). One reason for the minor difference could be violation of one or more of the assumptions underlying SI by MDD. First, the PSF is assumed to sample the incoming wavefield sufficiently densely, such that the integral in eq. (11) is evaluated accurately. For phase velocities lower than  $1600 \text{ m s}^{-1}$ , however, the nominal station separation of 2 km along the TN-line may give rise to spatial aliasing at a frequency of 0.4 Hz. Spatial aliasing will surely not be an issue between 0.1 and 0.2 Hz. Second, too much



**Figure C4.** Idem as Fig. 15 in the main text, but at 0.32 Hz, but in this case the dashed black line indicates the location of a virtual source along the TN-line of which the azimuth of the line connecting it with station TE07 coincides with the azimuth of the slowness vector whose components are given by the maxima in Fig. C2.

body-wave energy would deteriorate the effectiveness of SI by MDD because the condition that the receivers at  $\mathbf{x}_R$  (i.e. the TE stations in our case) solely record energy that has traversed the contour  $\mathcal{S}_{\text{rev}}$  (i.e. the TN-line in our case) would be violated; Figs C1 and C2 indicate that quite some body-wave energy arrives at 0.32 Hz during the selected time windows. Third, the medium associated with eq. (11) is assumed laterally invariant (or at least smooth) along  $\mathcal{S}_{\text{rev}}$ . The dispersion curves in Fig. 8 suggest that this assumption may not hold at higher frequencies, which in turn would adversely affect the results obtained by SI by MDD at these higher frequencies. Finally, the derivations in Section 2 assume propagation of a single surface wave mode. In case multiple surface wave modes are present, the procedure followed in this study does not allow accurate retrieval of surface wave Green's functions. Instead, separation of modes would be required prior to crosscorrelation (Halliday & Curtis 2008; van Dalen *et al.* 2014). Especially between 0.2 and 0.4 Hz, multiple surface wave modes cannot be ruled out (e.g. Figure S5; supporting material). Another reason why we do not observe a significant phase correction after application of SI by MDD in this higher-frequency band could be the fact there is simply little to correct for. Note that the phase error predicted by Weaver *et al.* (2009) (eq. 28) is proportional to  $1/t\omega$ . This inverse proportionality to frequency and travelt ime (or station separation in a homogeneous medium) implies that the same illumination pattern at 0.32 Hz would result in a significantly smaller phase shift than it would at 0.12 Hz. In other words, the illumination pattern needs to be more rugged at higher frequencies and/or larger station separation to produce phase errors of similar magnitude.