Source depopulation potential and surface-wave tomography using a crosscorrelation method in a scattering medium

GEOPHYSICS, VOL. 76, NO. 2 (MARCH-APRIL 2011); P. SA51–SA61, 8 FIGS.
10.1190/1.3535443

Source depopulation potential and surface-wave tomography using a crosscorrelation method in a scattering medium

Pierre Gouédard[1], Philippe Roux[2], Michel Campillo[2], Arie Verdel[3], Huajian Yao[4], and Robert D. van der Hilst[4]

ABSTRACT

We use seismic prospecting data on a 40×40 regular grid of sources and receivers deployed on a 1 km $\times$ 1 km area to assess the feasibility and advantages of velocity analysis of the shallow subsurface by means of surface-wave tomography with Green's functions estimated from crosscorrelation. In a first application we measure Rayleigh-wave dispersion curves in a 1D equivalent medium. The assumption that the medium is laterally homogeneous allows using a simple projection scheme and averaging of crosscorrelation functions over the whole network. Because averaging suppresses noise, this method yields better signal-to-noise ratio than traditional active-source approaches, and the improvement can be estimated a priori from acquisition parameters. We find that high-quality dispersion curves can be obtained even when we reduce the number of active sources used as input for the correlations. Such source depopulation can achieve significant reduction in the cost of active source acquisition. In a second application we compare Rayleigh-wave group velocity tomography from raw and reconstructed data. We can demonstrate that the crosscorrelation approach yields group velocity maps that are similar to active source maps. Scattering has an importance here as it may enhance the crosscorrelation performance. We quantify the scattering properties of the medium using mean free path measurements from coherent and incoherent parts of the signal. We conclude that for first-order velocity analysis of the shallow subsurface, the use of crosscorrelation offers a cost-effective alternative to methods that rely exclusively on active sources.

INTRODUCTION

The Green's function (GF) of a medium between two points A and B represents the record of ground motion that we would observe at A if a point source was applied at B (or vice versa). Weaver and Lobkis (2001, 2004), show that the two-point crosscorrelation of wavefields produced by randomly isotropic noise sources converges toward the GF, that is, including all reflections, scattering, and propagation modes. Various experimental, numerical, and theoretical studies (e.g., Wapenaar, 2006; Godin, 2007; Gouédard et al., 2008c) have since shown the validity of this principle also for wavefields produced by a sufficiently random distribution of deterministic seismic sources, or if wave randomization is caused by multiple scattering in heterogeneous media (Campillo and Paul, 2003; Gouédard et al., 2008b). In most seismological applications, however, the crosscorrelation is dominated by energy associated with surface waves because the near-surface location of receivers and most noise sources renders, effectively, a 2D acquisition geometry. We will refer to this band-limited, surface wave-dominated signal as the reconstructed GF.

The ability to estimate GFs from ambient noise has led to the development of passive surface-wave tomography, with successful applications in studies at a regional (e.g., Sabra et al., 2005; Shapiro et al., 2005; Yao et al., 2006; Stehly et al., 2009) or local (Picozzi et al., 2009) scale. In seismic exploration, crosscorrelation has been used for what is known as "redatuming," but so far this

Manuscript received by the Editor 28 May 2010; revised manuscript received 17 September 2010; published online 23 March 2011.
[1]Massachusetts Institute of Technology, Department of Earth, Atmospheric and Planetary Sciences, Cambridge, Massachusetts, U.S.A. and Université Joseph Fourier, Institut des Sciences de la Terre, CNRS, Grenoble, France. E-mail: gouedard@mit.edu.
[2]Université Joseph Fourier, CNRS, Institut des Sciences de la Terre, Grenoble, France. E-mail: philippe.roux@obs.ujf-grenoble.fr; michel.campillo@obs.ujf-grenoble.fr; michel.campillo@ujf-grenoble.
[3]Formerly at Shell International Exploration & Production B. V., Rijswijk, The Netherlands. Presently at Delft University of Technology, Department of Geotechnology, Delft, The Netherlands. E-mail: arie.verdel@gmail.com.
[4]Massachusetts Institute of Technology, Department of Earth, Atmospheric and Planetary Sciences, Cambridge, Massachusetts, U.S.A. E-mail: hjyao@mit.edu; hilst@mit.edu.

mainly involved reflected seismic wavefields produced by active sources to ensure better propagation through the shallow subsurface and better imaging of deep reflectors (Bakulin and Calvert, 2006; Korneev and Bakulin, 2006; Schuster and Zhou, 2006).

There is growing interest in surface waves in seismic exploration, for instance for velocity analysis of the reservoir overburden (e.g., Campman and Riyanti, 2007; Socco and Boiero, 2008), and the GF reconstruction method can be used to enhance active source experiments. In this context, it is sometimes referred to as seismic interferometry. For instance, one could augment data coverage by turning receivers into virtual sources, or reduce costs by using fewer active sources and more virtual ones. However, the feasibility of this type of reconstructed surface-wave tomography and its performance relative to applications that rely exclusively on active sources are yet to be established.

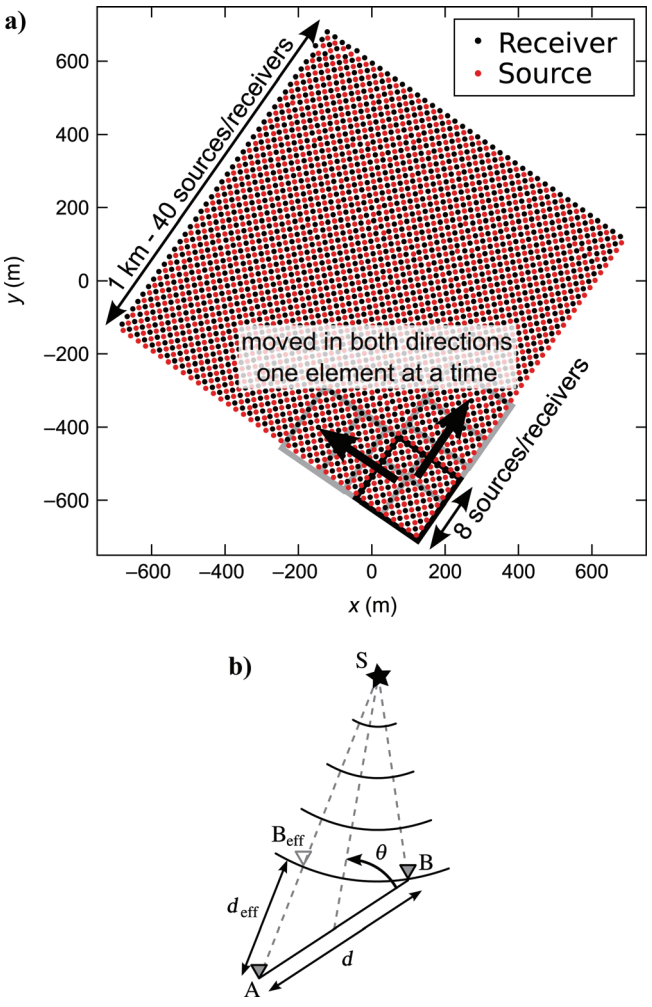

Figure 1. (a) Array configuration and bins used for the tomography. The black square indicates the first bin of 8 × 8 grid cells used for the 2D tomography. Every next bin (gray square) is obtained by moving this bin in both directions, by one element at a time, so the whole area is covered. (b) For a pair of receivers (A, B) (triangles) and a source S (star), we define the effective position B_{eff} of receiver B for which the source is aligned with the receiver pair. The crosscorrelation of the recordings at A and B for the source S yields the individual crosscorrelation for the equivalent distance $d_{eff} = d(A, B_{eff})$. The angle θ shows conventions used for the source azimuth definition.

The main objective of this paper is to introduce and illustrate concepts related to the use of reconstructed waveforms for the purpose of surface-wave tomography with seismic prospecting data; a detailed subsurface characterization is beyond the scope of this study. We will demonstrate the feasibility of measurements of wave speed from reconstructed GFs assuming laterally homogeneous (1D) and heterogeneous (2D) media. In both cases we show how active sources can be removed from the data set and replaced by virtual ones (that is, source depopulation) without losing resolution in the velocity analysis. Access to data from 1600 sources and seismic receivers deployed on a regular grid of 1×1 km allows us to compare velocity analysis from reconstructed GFs with that from active source data.

After introducing the correlation principle, in a first application we present how 1D-equivalent dispersion curves can be measured under a homogeneous medium approximation. This approximation allows the use of a projection scheme to stack crosscorrelation functions obtained from every source. The improved signal-to-noise ratio produced by this data redundancy allows us to replace active with virtual source data and still get high-quality dispersion curves. This source depopulation can lead to a substantial reduction in cost of active source experiments. Subsequently, we show that the gain in signal-to-noise ratio is a function of the targeted wave's amplitude and the geometry of the network.

In a second application, we drop the 1D approximation and determine lateral variations in medium velocity using surface-wave tomography. Comparison of the group velocity maps obtained from active and crosscorrelation methods demonstrates the promise of the latter for velocity analysis of near-surface structures (e.g., reservoir overburden). We show that source depopulation by a factor of four does not significantly change the final tomographic model.

The last part of this paper is devoted to the characterization of the scattering in the medium, which we will show is heterogeneous, from the analysis of the coherent and incoherent parts of the wavefields.

THE SEISMIC NETWORK

For our study we use data from a high-resolution survey of a 1×1 km carbonate (karst) area in northern Oman conducted by Petroleum Development Oman (PDO). The data were acquired with a 40×40 grid of geophones and sources (both with 25×25 m spacing), but with the source and receiver grids shifted with respect to one another by half a grid distance in both directions, that is, 12.5 m (see Figure 1a). Seismic vibrator trucks were used as the source on each node of the source grid, emitting sweeps (frequency modulated signal) in the 8–120 Hz frequency band. Each recording point is a cluster of 12 vertical-component SM-4 geophones, sampling the wavefield at 2 ms. The records are first crosscorrelated with the vibroseis source-sweep function (pulse compression). Because the autocorrelation of a sweep is close to a Ricker wavelet, this crosscorrelation is similar to — but easier to apply and more stable than — deconvolution of the source function. Signal records are then cut to a 4-s duration and downsampled to 8 ms to generate what is hereinafter referred to as the "raw" data. For further information, refer to Herman and Perkins (2006) and Gouédard et al. (2008b).

The complete data set, consisting of 1600×1600 source-receiver time-domain signals, constitutes an exhaustive measurement of the transfer function of the half-space medium over a 1-km^2 area.

Because of the 2D acquisition geometry, the data set includes mainly Rayleigh (surface) waves. We note that the results presented below could also have been obtained by using time windows centered on the arrival of the fundamental mode Rayleigh wave (e.g., through application of a group velocity cut-off), but because the results are controlled by this mode anyway and because the use of a shorter window would not lead to significant reduction in computational expense, we decided to use the full 4-s time windows.

The 4-s records are used as input for the crosscorrelation method, as any other diffuse wavefield would have been. We will process the reconstructed GFs and compare the results to those obtained by processing the data from an active-source configuration, that is, without performing the crosscorrelation.

1D EQUIVALENT DISPERSION CURVES FROM A HOMOGENEOUS MEDIUM APPROXIMATION

Crosscorrelation procedure

In a first application we consider the medium to be laterally homogeneous (1D), and we measure phase velocity dispersion through frequency-slowness analysis (which involves a time-space to frequency-wavenumber Fourier transform). The conventional crosscorrelation method considers a pair of receivers (A, B) with sources spread around it, such that the wavefield after summation over all sources is isotropic. In such a geometry, only sources located inside the endfire lobes of this pair (i.e., within a small aperture around a line through A–B) contribute to the averaged correlation function, whereas contributions from other directions vanish in the averaging (Roux et al., 2004). This can be understood from stationary phase (Snieder, 2004) or Fresnel zone arguments (Yao and van der Hilst, 2009). The width of the endfire lobe depends on the ratio between the distance d, between receivers A and B and the dominant wavelength of the data.

If considering a single source located inside the endfire lobe, the correlation function does not correspond to a GF anymore, as there is no averaging over sources spread over the whole Fresnel zone, but it nevertheless allows us to measure phase delays between the receivers. We will refer to such a correlation function as an individual crosscorrelation function (ICC).

The ICC and the reconstructed GF both allow measurements of wave speeds, but the ICC lacks near-field effects (such as the $\pi/4$ phase shift for the Rayleigh wave). The ICC yields correct phase delays for all direct, reflected (from a horizontal interface at depth, as we assumed a 1D medium), and refracted waves, as for all these waves the wavepaths to each receiver lie in the vertical plane going through the source and both receivers (i.e., the wavefront remains circular when projected onto the surface).

If the medium is laterally homogeneous we can benefit from axial symmetry. Because in homogeneous media the wavefronts are circular, the waveform recorded at receiver B, for a given source S, is identical at any point of the wavefront going through B (again, this is valid for all direct, reflected, and refracted waves). We can thus define an effective position of the receiver B_{eff} on this wavefront (Figure 1b), for which the source S is in the alignment of the receiver pair (A, B_{eff}).

Point B_{eff} is the image of B by a circular projection onto the line joining the source S and receiver A. By computing the crosscorrelation of the recordings in A and B for this particular source S, we obtain the ICC for an equivalent distance $d_{eff} = d(A, B_{eff})$, hereinafter referred to as the offset. In other words, the correlation function obtained from a misaligned source with respect to the receiver pair, which in conventional processing may vanish in the summation over all source contributions, may be considered as the ICC for an effective distance d_{eff} that is smaller than the distance d, between A and B. Notice that for a homogeneous medium, defining A_{eff} and $d_{eff} = d(A_{eff}, B)$ is equivalent.

The offset d_{eff} is defined with respect to the source azimuth and may be different for each source. Considering only two receivers, and a distribution of sources surrounding them (i.e., with a good sampling of the $0°-360°$ azimuth range), one can build an ICC section with offsets d_{eff} densely spread from 0 (for sources with azimuth $90°$ or $270°$) to the real distance $d = d(A,B)$ (for sources with azimuth 0 or $180°$) — see also (Gouédard et al., 2008a). One can then consider several sources and stack the ICC sections obtained from each of them.

Application of this projection concept to the Oman data set (that is, $1600 \times 1599/2$ receiver pairs and 1600 sources) yields $\sim 2.10^9$ possible crosscorrelation functions, along with the associated effective distances. Because in a homogeneous medium the ICC (and GF) depends only on the offset d_{eff} and not on the location, we stack correlation functions associated with the same small (1 m) offset bin. The resulting ICC section is presented in Figure 2a. As expected from the acquisition geometry, this record section is dominated by Rayleigh waves. A 2D Fourier transform is applied to this ICC section to transform it to the frequency-wavenumber space, using zero-padding to have 1024 sampling points in time and 512 in space, and tapering the data in space. The absolute value of this frequency-wavenumber spectrum is then interpolated to the frequency-slowness space and normalized by its maximum amplitude. The resulting spectrum is presented on Figure 2b.

For comparison we averaged the raw data in a similar way: For each receiver we stacked all records for sources at the same offset from that receiver (as above, we use 1 m offset bins), which resulted in a seismic section for each receiver. We then averaged these seismic sections over all receivers. The resulting averaged seismic section is presented in Figure 2c. This raw seismic section involves 800 times fewer traces than the section obtained from crosscorrelation (Figure 2a). The frequency-slowness spectrum from raw data (Figure 2d — obtained using the same processing parameters as for Figure 2b) reveals the same features as the spectrum from the ICC data (Figure 2b), but the dynamic range is much smaller (25 dB versus 55 dB). The superior dynamic range in the ICC section can be explained by a better signal-to-noise ratio thanks to the much larger data redundancy at each offset, especially at the smaller distances. The projection scheme also allows sampling at short offsets, where the raw section (Figure 2c) has gaps.

The stack and corresponding power spectrum of the raw data reveal the existence of waves with higher phase speeds than the surface waves (arrows in Figure 2c, d). These probably represent refracted waves. They are not visible in the time-domain section of the crosscorrelations (Figure 2a), but their appearance in the slowness-frequency spectrum (Figure 2b, arrow) suggests that they exist in the ICC data also, but with amplitudes that are far less than that of the surface waves that dominate the correlations. We remark that the correlation method is unable to recover the actual relative amplitude between waves, even though each wave is present in the reconstruction (see also the section about signal-to-noise gain below).

Source depopulation

Crosscorrelation can be used to create a virtual source at each receiver location. Because active source acquisition is expensive, here we explore if we can reduce the number of active sources (that is, replace them with a virtual source) while maintaining data quality, as measured either in the space-time or in the slowness-frequency domain. We refer to this concept as source depopulation.

Figure 3a shows again the reconstructed section produced by crosscorrelation of records produced by all 1600 sources, and Figure 3b represents the corresponding slowness-frequency spectrum. Figure 3c and d presents the reconstructed seismic section and spectrum, respectively, for crosscorrelation (for all 1600 receivers) of data produced by active sources at only 100 sites (that is, one out of every four source sites in each direction), and Figure 3e, f and g, h represents similar outputs for four active sources (one at each corner of the network) and for one active source (at a corner), respectively.

The resolution of the seismic sections and the frequency-slowness spectra does not decrease significantly when reducing the number of sources from 1600 (Figure 3a, b) to 100 (Figure 3c, d). When considering only four sources, the time-domain representation stays very good (Figure 3e), but the data quality starts to degrade slightly and the frequency-slowness spectrum becomes more noisy (Figure 3f). In particular, the high-velocity wave mentioned above is no longer visible. The dispersion of the fundamental Rayleigh mode can nevertheless still be measured. When using a single source, the fundamental Rayleigh wave is still well reconstructed in the seismic section (Figure 3g), but it becomes harder to measure dispersion from the frequency-slowness diagram (Figure 3h). These results illustrate the fact that when using crosscorrelation methods, the resolution is mainly controlled by the receiver network. Indeed, what allows us to have usable measurements even for a single source is the large number of receivers and, thus, the large number of averaged correlation functions ($1600 \times 1599/2 \approx 1.3 \cdot 10^6$ for Figure 3g).

We stress that the spatial sampling in Figure 3g is still very good, despite using a single source. To achieve comparable sampling in a classical source/receiver setup using a single source one would either have to use many more receivers or at least a more dedicated geometry, which would not be regular.

Source depopulation makes it possible to obtain high-quality seismic sections and slowness-frequency spectra with significantly fewer source sites than in conventional experiments. In addition to the obvious cost reduction, this is an attractive option in situations where sources are not easy to set up, for example in densely populated areas or when dealing with rugged topography or dense vegetation.

One can also consider when dealing with the reciprocal case, namely the receiver depopulation process, with practical advantages for cases where the receiver effort exceeds the source effort. At the seabottom, for instance, this could be used to reduce the number of (very expensive) ocean bottom seismometers needed to meet a given image resolution objective.

As presented here, the source depopulation process is optimized for surface waves, which dominate the reconstructed GFs for the acquisition geometry used. Source depopulation could theoretically be used for any kind of wave (we showed how a refracted wave was impacted, even while it was not the main target), but the geometry of the array has to be adapted to the targeted wave.

SIGNAL-TO-NOISE RATIO GAIN FOR CROSSCORRELATION METHODS

In this section we show how the gain in signal-to-noise ratio (when using crosscorrelation) can be predicted for a given array geometry. This can also be seen as the answer to the following question: What are the geometry requirements to reach a desired

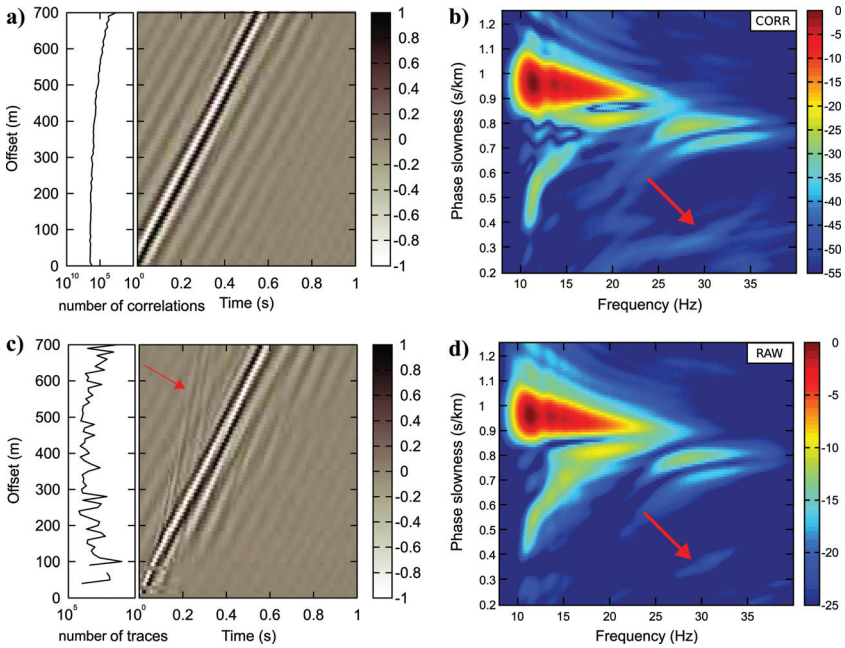

Figure 2. (a) Individual crosscorrelation (ICC) section obtained from the correlation in the 1D approximation. A Rayleigh wave is dominating. The inset on the left indicates the number of correlation functions effectively averaged at each offset. ICCs are normalized in amplitude at each offset. (b) Normalized frequency-slowness diagram (in decibel scale) obtained from the ICC section. (c and d) Same as (a and b) but obtained from raw data. The two diagrams in (b and d) are very similar, apart from a difference in the amplitude dynamic range. Red arrows point to a refracted wave in the time/offset and the frequency/slowness domains, wherever it is visible.

signal-to-noise ratio gain? We focus here on the signal-to-noise ratio gain for surface waves, but similar reasoning can be used for body waves.

We denote by $a_R(r) < 1$ the relative amplitude (as function of offset r) of the refracted wave with respect to the Rayleigh waves in the raw seismic section. Similarly, we denote by $a_C(r) < 1$ the relative amplitude of the same refracted wave in the correlated seismic section. Because we consider surface waves, so that the average field is not fully equipartitioned, and because correlation is mathematically a product in the Fourier domain, energy ratios between the different waves in the reconstructed GFs are squared:

$$a_C(r) = a_R^2(r). \quad (1)$$

Indeed, in Figure 2b the refracted wave (indicated by red arrow) has amplitudes that are twice as large (in negative dB) as in Figure 2d.

The signal-to-noise ratio in a raw seismic section at distance r can be written

$$\mathrm{S/N}_R(r) = \mathrm{S/N}_R^{\circ}(r)\sqrt{N_R(r)}, \quad (2)$$

where $\mathrm{S/N}_R^{\circ}(r)$ is the signal-to-noise ratio for one trace and $N_R(r)$ the number of traces averaged for the offset r, that is, the number of receivers located at offset $r \pm 1$ m for a source, summed over all sources, as plotted in the inset of Figure 2c.

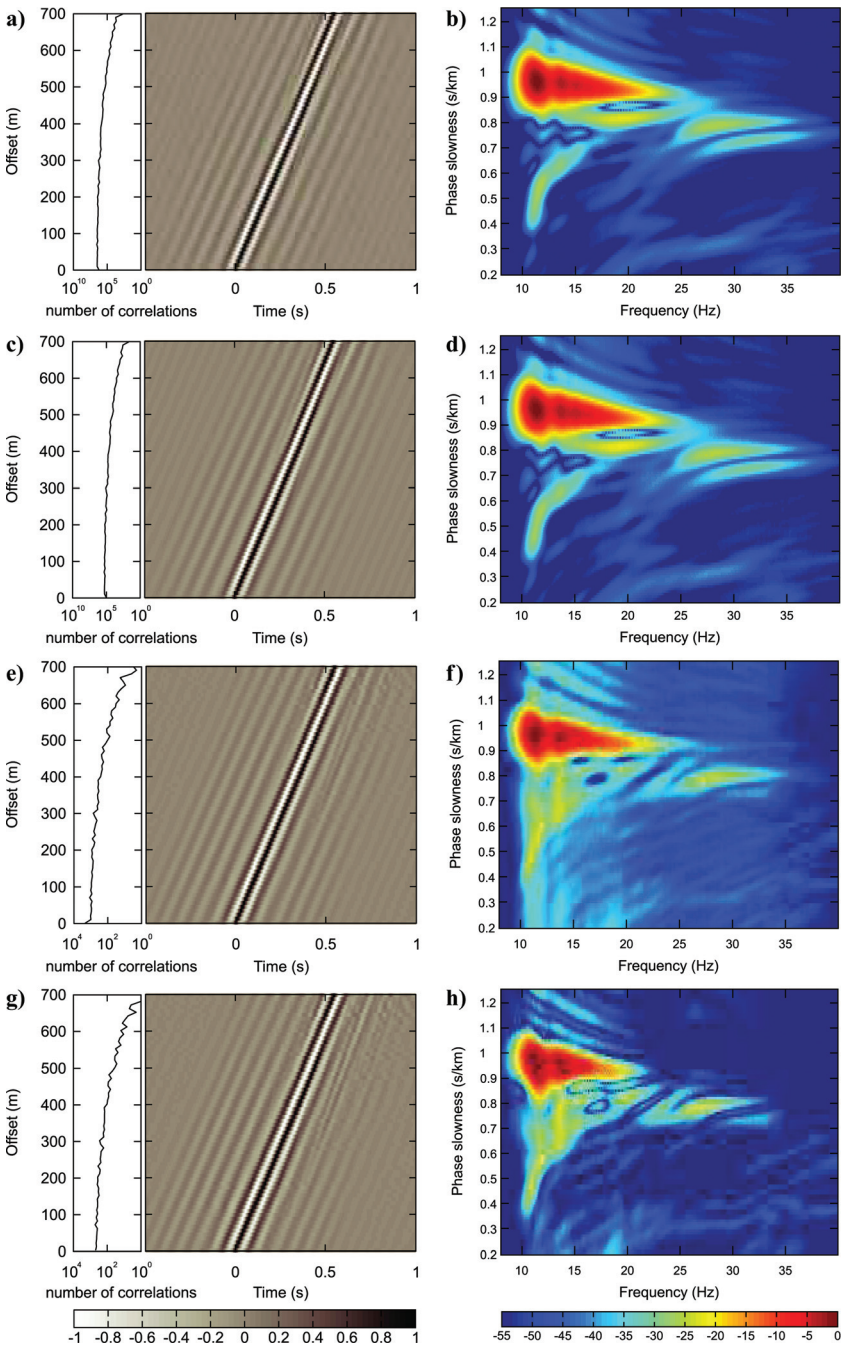

Figure 3. Reconstructed seismic sections, with the number of correlation functions effectively averaged at each offset indicated on the left (left row), and corresponding frequency-slowness diagram (right row) using fewer and fewer sources in the averaging process. First line (a and b) is obtained using all the 1600 available sources (we note that these figures are the same as Figure 2a and b); second line (c and d) is obtained using 100 sources (one over four in both directions); third line (e and f) is for four sources (one at each corner of the network); last line (g and h) is obtained using only one source, located at a corner of the network.

The relation between $N_R(r)$ and the total number of sources and receivers in the data set depends on the distribution of effective distances, which itself depends on the array geometry. Likewise, we introduce the signal-to-noise ratio in the correlated seismic section:

$$\mathrm{S/N}_C(r) = \mathrm{S/N}_C^{\circ}(r)\sqrt{N_C(r)}, \tag{3}$$

where $\mathrm{S/N}_C^{\circ}(r)$ is the signal-to-noise ratio for one correlation function, and $N_C(r)$ the number of correlation functions averaged for the offset r (within a 1-m bin, as plotted in the inset of Figure 2a). We define the signal-to-noise ratio for the refracted wave as $\mathrm{S/N}_R^a(r) = a_R\mathrm{S/N}_R(r)$ for the raw section and $\mathrm{S/N}_C^a(r) = a_C\mathrm{S/N}_C(r)$ for the correlated section. Using equation 1, the gain in signal-to-noise ratio can be written as

$$G^a(r) = \frac{\mathrm{S/N}_C^a}{\mathrm{S/N}_R^a} = \frac{\mathrm{S/N}_C^{\circ}}{\mathrm{S/N}_R^{\circ}}\frac{a_C}{a_R}\sqrt{N_C/N_R} = \frac{\mathrm{S/N}_C^{\circ}}{\mathrm{S/N}_R^{\circ}}a_R\sqrt{N_C/N_R}. \tag{4}$$

This gain depends on the offset r and can be smaller than 1 (for small a_R and/or small N_C/N_R ratio). For the geometry considered here, when using all the 1600 source sites available, the ratio N_C/N_R appears to decrease exponentially with offset (Figure 4).

The gain G^a depends also on the targeted wave amplitude a_R, and the smaller the amplitude, the more averaging is needed to achieve a given signal-to-noise gain. Because of the high quality of the data considered here, the signal-to-noise ratio for the ICC functions is comparable to that of the raw data; that is, $\mathrm{S/N}_C^{\circ}/\mathrm{S/N}_R^{\circ} \approx 1$. Given the values of a_R from Figure 2d and the N_C/N_R ratio from Figure 4, $G^a(r)$ is smaller than one for offsets greater than ~500 m but close to 40 at small offsets.

Equation 3 gives insight into how many averaged crosscorrelation functions are needed at each offset for the refracted wave to appear above the noise level. Using the value of a_C measured from Figure 2b, and assuming $\mathrm{S/N}_C^{\circ} = 10$ (which is lower than the true value), we get

$$N_C > \frac{1}{(a_C\mathrm{S/N}_C^{\circ})^2} \approx 100. \tag{5}$$

This is in agreement with what we see in Figure 3, where the refracted wave falls below the noise level when using four or fewer sources (Figure 3e and f). For this number of sources, the inset of Figure 3e shows that the number of averaged crosscorrelation functions is lower or close to 100 for more than half of the offsets. This allows us to predict, qualitatively, and for a given wave amplitude, how many sources and receivers one should use, and in which geometry, to have this wave appear above the noise level.

Notice that the example taken from the previous section and used here for numerical validation represents a 1D case. Equation 4 is still valid in 2D or 3D geometries. However, in the particular case of our data set, we used a nontraditional plane-wave-based approach for the 2D case, as explained in the next section. This approach changes the signal-to-noise gain, as traces are stacked in a different way, and equation 4 is no longer valid. A similar formula could nevertheless be derived in this case.

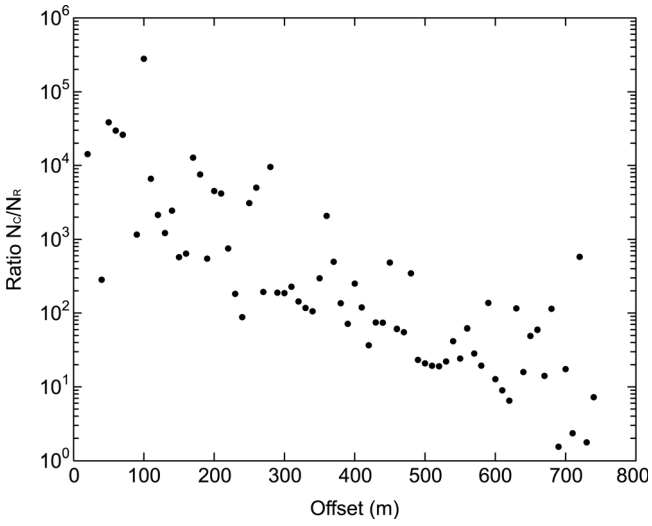

Figure 4. Ratio N_C/N_R as a function of offset r in our particular geometry, when using all the 1600 source sites available in the data set. This ratio decreases exponentially with r. Dispersion around the average curve is explained by the incompleteness of the data set as a result of the removal of some bad traces.

2D HETEROGENEITY FROM SURFACE-WAVE TOMOGRAPHY

In the first part of this paper we assumed lateral homogeneity of the medium and measured Rayleigh wave dispersion curves that represent an integration of the medium properties over the entire studied region. Here we conduct a tomographic inversion of surface-wave dispersion curves for lateral heterogeneity. For this purpose we use group velocity measurements of the fundamental mode of the Rayleigh waves, and we produce group velocity maps for different frequencies. Looking at the mean velocity values from Figure 2 and the processing frequency band (10–25 Hz), these maps reflect the heterogeneity of the medium at relatively shallow depth (~30 m).

Group velocity maps from raw data

A straightforward approach to surface-wave tomography would be to measure fundamental mode Rayleigh wave dispersion for all source-receiver combinations (that is, over a range of azimuths and offsets) and then use these data to produce 2D group velocity maps. Picking of the direct arrivals in individual records proved challenging, however, in particular for the larger offsets (perhaps as a result of the effects of scattering).

To overcome this difficulty, we measured surface-wave arrivals (using a plane-wave approach) on 8×8 subgrids, or bins, which were moved across the entire 40×40 grid one grid increment at a time (see Figure 1a). Each of these bins encompassed 64 sources and receivers (recall that the sources are located on a grid that is staggered with respect to the receiver grid). For each bin the sources on one of the sides represent a source array, and the receivers on the opposite site form a receiver array.

At each receiver we summed the traces from the eight sources of the source array to form a plane-wave beam, and measured the arrival of the maximum of the envelope of the Rayleigh

wave. This time measurement is attributed to a path through that receiver and perpendicular to the receiver array.

Each source array thus yields eight measurements, and for each bin we obtain 32 data points (for the four plane waves propagating from left, right, top, and bottom as shown in Figure 5). The bin is then shifted laterally over one grid unit (Figure 1a) and the measurement repeated for four new source and receiver arrays.

After measuring all traveltimes, we used a tomographic inversion to constrain the lateral variations in group velocity of the fundamental mode Rayleigh wave in the 1 km^2 area. For this inversion we use a quasi-Newton algorithm due to Tarantola and Valette (1982) and Tarantola (1987), using straight rays. The starting model is homogeneous, with a group velocity of 1200 m/s measured from seismic sections of Figure 2. Smoothing is added to the inversion via a model covariance matrix with a correlation length of 100 m. The starting velocity model was measured from the data, and we expect the velocity deviations from this model, as well as their effect on raypaths, to be small. The weak nonlinearity justifies inversion in a single iteration. Indeed, tests showed that a second iteration does not change the final model much. With this method we produced a group velocity map in the frequency band 10–25 Hz (Figure 6a), with a variance reduction compared to the homogeneous model of 61.3%.

Group velocity maps from reconstructed GFs

We applied the processing explained in the previous section to the reconstructed GFs obtained from crosscorrelation. For this application the bins included receivers only, and the traveltime measurements were made on the reconstructed GFs between selected receivers. Here again, we applied the above-mentioned

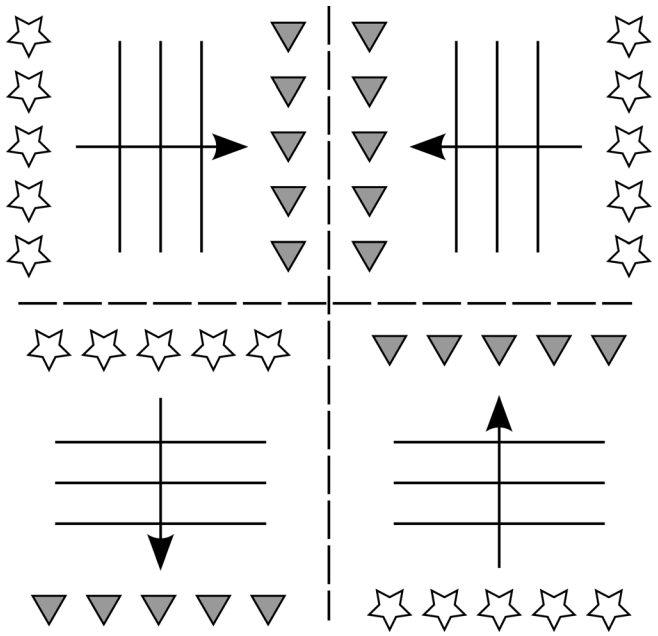

Figure 5. In each bin, four lines of sources are used to beamform four plane waves, propagating through the bin in four directions, and recorded at four lines of receivers.

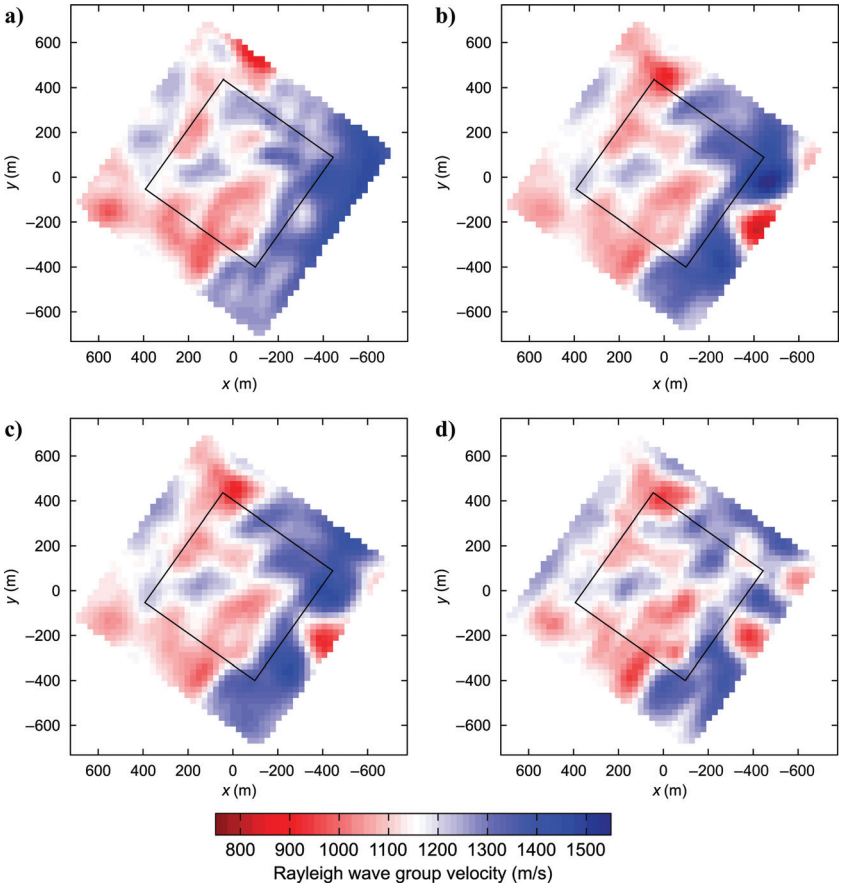

Figure 6. (a) Rayleigh-wave group velocity model obtained by inverting raw data, for the 1×1 km area, in the 10–25 Hz frequency band. (b) Similar map obtained by applying the same algorithm to the GF, reconstructed using the crosscorrelation technique using 1600 source sites. The two approaches compare well, apart from the boundaries of the area where the crosscorrelation method performs poorly. The black square shows the confidence zone for the crosscorrelation method. (c and d) Same as (b), but using only 400 and 100 source sites, respectively.

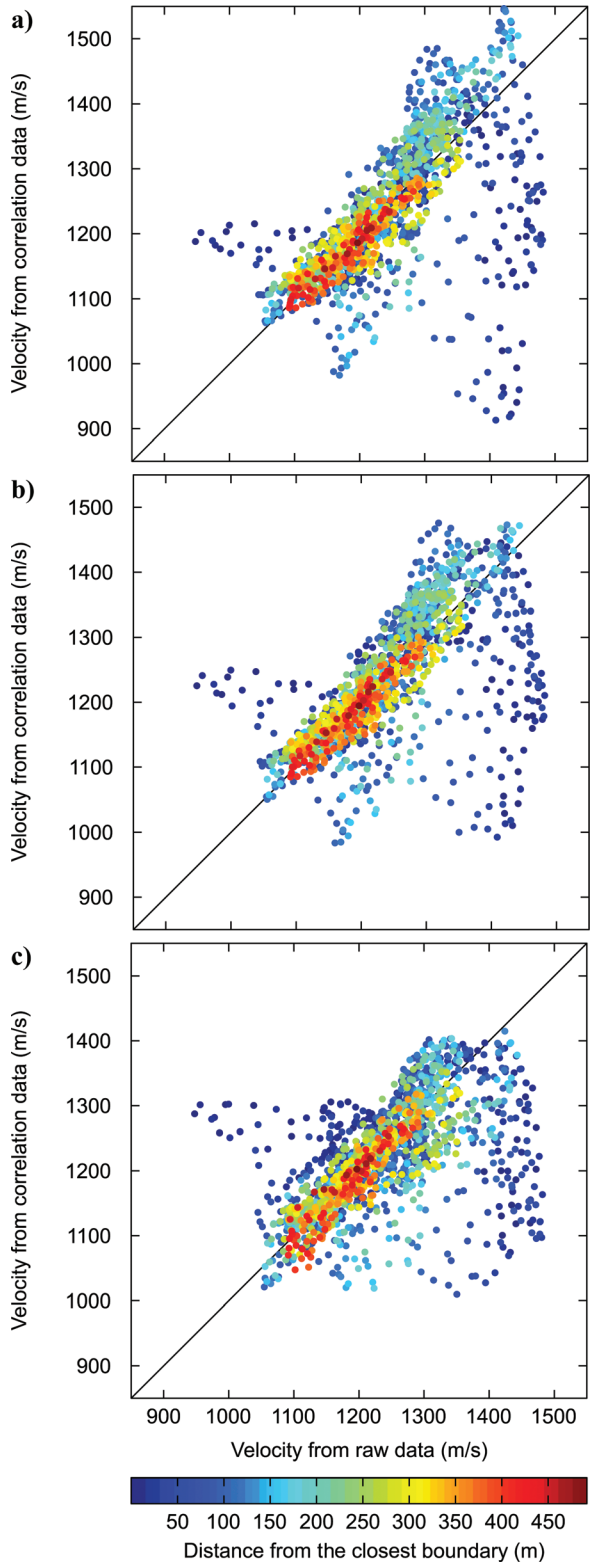

Figure 7. (a) Comparison of models inferred from raw and reconstructed data. For each grid cell, the velocity found from reconstructed GFs as shown in Figure 6b is plotted versus the velocity found in the raw data tomography of Figure 6a. Color indicates the distance from the cell to the closest boundary of the 1-km square network. The black line is the identity reference line. (b) and (c) Same as (a), but using only 400 and 100 source sites for the GF reconstruction, respectively.

plane-wave-based picking method, with plane waves generated using lines of virtual sources. Crosscorrelation functions were computed from the recordings of all available sources to obtain a high signal-to-noise ratio.

Using all sources generates a quasi-isotropic illumination for the selected pair. This isotropic illumination approximation is not valid at the boundaries of the network, as the source array is not larger than the receiver array, and the traveltime measurements in these areas are inaccurate. Around the center of the array, the wavefield generated by all sources is not strictly isotropic either because there is little chance to have, for a given pair, the exact same number of sources in each direction. A nonisotropic source distribution is nevertheless not an issue, as several authors have shown that the error in traveltime measurements made in such cases is small and can be predicted and corrected for (Fan and Snieder, 2009; Weaver et al., 2009; Yao and van der Hilst, 2009; Froment et al., 2010).

The resulting Rayleigh-wave group velocity model, which produces a variance reduction of 46.57%, is shown in Figure 6b. The velocity map is similar to Figure 6a. Figure 7 presents a more detailed comparison of these maps by plotting, for each cell of the grid, the crosscorrelation model against the model inferred from the raw data. Differences are located mainly within ~200 m from the boundaries of the area; here the crosscorrelation performs poorly because the source distribution considered for GF reconstruction is inappropriate (Gouédard et al., 2008b). The black square in Figure 6 shows the confidence zone, where these boundary problems are not occurring. Small discrepancies within the confidence zone between the two models can be explained by measurement errors and inversion uncertainty. For grid cells located around the center of the array, Figure 7 shows that the velocities of Figure 6a and b are close.

Source depopulation

As shown previously, an important advantage of the correlation method is the potential for source depopulation, that is, it could allow the use of fewer sources for the same image quality. Indeed, for a dense receiver grid, sources at the periphery of the area should be sufficient to reconstruct all the GFs needed for the tomography. Unfortunately, we cannot verify this expectation with the given data set because not all receivers recorded all sources; in particular, the wavefield from sources at one side of the network was not recorded by all receivers at the opposite side.

We will nevertheless apply source depopulation and see how the group velocity images change when using fewer sources as an input for GF reconstruction. We use the same picking procedure as above, but this time we use only one out of every two (or four) source sites in each direction, resulting in 400 (or 100) source sites in total. The group velocity maps produced from these data are presented in Figure 6c and d. A comparison with the maps from the raw data is presented in Figure 7b and c.

When using 400 source sites, instead of the 1600 used for the original tomography (Figure 6b), the group velocity model remains the same. When using only 100 source sites, the first-order features remain well retrieved, although some resolution is lost compared to the previous models. Figure 7 shows that no systematic bias is introduced in the velocity model when reducing the number of source sites. The confidence zone is slightly reduced, however, which was expected because removal of

source sites made it harder to have a suitable source distribution to reconstruct the GF between a pair of receivers located close to the boundaries. We recall that the existence of the confidence zone as well as its size and shape are direct consequences of the source- and receiver-array configuration, which suggests that they can be optimized through appropriate array design.

Source depopulation was very effective in the 1D case discussed in the previous section because one source site allowed measurement of Rayleigh-wave dispersion (Figure 3h). Averaging is less effective in the heterogeneous case considered here because we need to retain the ability of different Rayleigh wave travel paths to add different constraints on structure. Nevertheless, modest source depopulation is possible and economically attractive.

CHARACTERIZATION OF SCATTERING

Mean free path: Coherent versus incoherent signal

As stated before, the data set was obtained in a carbonate (karst) area, and the medium is scattering, as shown by the maps in Figure 6 where typical sizes for heterogeneities are comparable to the wavelength ($\lesssim$100 m). This is important as scattering randomizes the wavefield and can lead to equipartition of energy, which helps with the retrieval of the GF when the source distribution is less than optimal (Campillo and Paul, 2003; Gouédard et al., 2008b). Characterization of scattering is therefore an important step when one wants to use crosscorrelation methods. In the following we try to measure the elastic mean free path ℓ, which is the characteristic distance traveled by the wave between two scattering events, and which classically is used as the scattering length of a propagation medium.

The elastic mean free path traditionally is measured from the decrease of the amplitude of an incident plane wave in a scattering medium. Indeed, in a lossless medium, the plane-wave amplitude, averaged over different realizations of the medium, exponentially decreases with propagation distance with a characteristic length defined as the elastic scattering mean free path. Creating a plane wave with our data set requires combining several sources together into a source subarray, whose total field is averaged over several receivers as shown in Figure 5. The distance between the source and the receiver subarrays ranges from 150 to 250 m depending on the receiver locations. Different realizations of the medium can be evaluated by shifting the setup over the whole network, as in Figure 1a.

However, two experimental limitations make this traditional measurement of the mean free path unsuccessful with the data set we used. First, the plane-wave approximation assumes both the source and the receiver subarrays to be in the far-field of each other, which corresponds, at 15 Hz, to a minimum offset of ~300 m, where our maximum offset is 250 m. Second, the size of the network is too small to perform enough independent realizations, which results in a bias in the measure of the average incident field. Furthermore, this method does not account for reverberations in the medium, which are present in our data as suggested by the presence of a refracted wave as seen in the first part of this paper.

These difficulties can be overcome by using another approach to measuring the scattering mean free path. The coherent-to-incoherent intensity ratio was introduced by De Rosny and Roux (2001) as a way to normalize the point-to-point intensity of the coherently averaged field by the point-to-point incoherently averaged intensity. It allows measuring the scattering mean free path of small scatterers inside a reverberant cavity. Working point-to-point between every source and every receiver is the key here to increasing the number of independent realizations and suppressing the far-field approximation associated with the source and the receiver subarrays.

We consider the signal $s_n(t,\mathbf{r})$ issued from source n and recorded on every single receiver. We then perform a coherent average of all the signals recorded at the same distance from the source $s_n(t,r)=\langle s_n(t,\mathbf{r})\rangle_{|\mathbf{r}|=r}$. This first average over the receiver positions is an ensemble average over the medium because the source-receiver paths all are different. The second step consists of averaging $s_n(t,r)$ over all sources such that $s(t,r)=\langle s_n(t,r)\rangle_n$. We then obtain the coherent intensity I_c:

$$I_c(t,r)=s^2(t,r)=\langle\langle s_n(t,\mathbf{r})\rangle_{|\mathbf{r}|=r}\rangle_n^2. \quad (6)$$

On the other hand, the incoherent intensity I_i, also called total intensity by Ishimaru (1978, chap. 14), is the average of the field intensity over the same realizations:

$$I_i(t,r)=\langle\langle s_n^2(t,\mathbf{r})\rangle_{|\mathbf{r}|=r}\rangle_n, \quad (7)$$

with the same ensemble average on the sources and the receivers. It has been shown that the ratio of the coherently averaged intensity $I_c(t,r)$ to the incoherently averaged intensity $I_i(t,r)$ provides a direct measurement of the scattering mean free path as (De Rosny and Roux, 2001)

$$R(r)=R(t=r/c,r)=\frac{I_c(t=r/c,r)}{I_i(t=r/c,r)}=\exp\left(-\frac{r}{\ell}\right), \quad (8)$$

where c is the wave speed in the medium.

As in the plane-wave measurement of the mean free path, the coherent intensity takes into account the scattering properties of the propagation medium by taking the intensity of the averaged field. The incoherent intensity is then just a normalization factor that accounts for geometrical spreading associated to the point-to-point measurement as well as the medium reverberation and attenuation, if any.

Application to the data

The measurement of coherent-to-incoherent intensity ratio $R(r)$ in the 10–25 Hz frequency band, following Equation 8, is presented in Figure 8. The group velocity used is 1200 m/s, as measured from the seismic sections of Figure 2. The logarithmic dependence of the ratio $R(r)$ as a function of the propagation distance for the Rayleigh wave is linear for offsets from 0 to 200 m. A linear fit is used to invert for the scattering mean free path of the Rayleigh wave, as expected from equation 8, yielding $\ell=374$ m. This value tells us that the wavefield is equipartitioned in energy after ~1800 m distance ($\sim 5\times\ell$, see Paul et al. [2005]), or after ~1.5 s in the recording time, which corresponds to what was found by Gouédard et al. (2008b) using the same data set.

Using coda instead of direct waves for the GF reconstruction, as illustrated by Gouédard et al. (2008b), reduces the importance of the source distribution, and yields reliable GFs in situations where direct waves fail. However, as discussed by Gouédard et al. (2008b), convergence toward the GF is slower when using

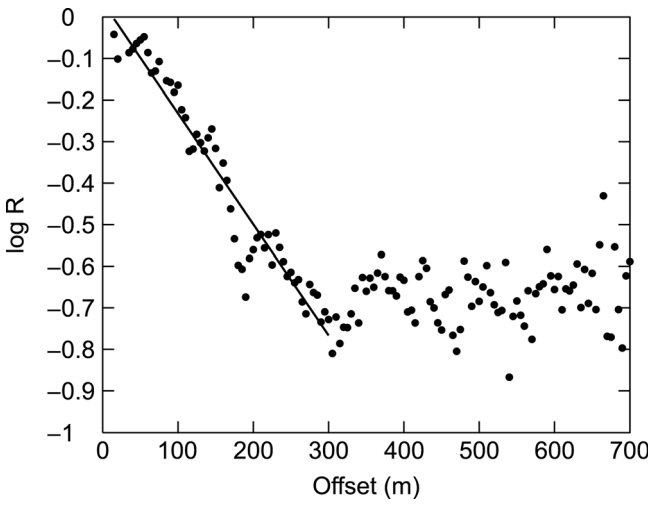

Figure 8. Ratio $R(r)$, in logarithmic scale. The linear fit (done in the 0–200 m offset range) yields a value of $\ell = 374$ m for the scattering mean free path for the Rayleigh wave.

coda waves, and for a given geometry and record length the signal-to-noise ratio is smaller with coda than with direct waves. This low signal-to-noise ratio prevents us from using coda wave tomography for this data set.

Figure 8 shows a plateau of the coherent-to-incoherent intensity ratio $R(r)$ after 300 m. This plateau may be linked to the number of independent realizations that can be obtained in our acquisition geometry, as described by De Rosny and Roux (2001). In our case, this number cannot be determined easily because for offsets greater than 300 m the Rayleigh wave interferes with some reflected waves, and the intensity ratio $R(r)$ is not representative of a single wave anymore.

Nevertheless, we can expect the number of independent realizations to be small, despite the large number of source/receiver pairs used, by comparing the value of the scattering mean free path $\ell = 374$ m for the Rayleigh wave to the size of the kilometer-squared network.

CONCLUSIONS

We showed the feasibility of velocity analysis with surface-wave tomography using Green's functions (and/or individual correlation functions) estimated from crosscorrelation, first using a homogeneous medium approximation, and then working with the actual laterally inhomogeneous medium. Velocity measurements from raw and correlated data are of comparable accuracy, with sometimes a superior signal-to-noise ratio from the correlation approach as shown in the homogeneous approach. The gain (or loss) in signal-to-noise ratio depends on array geometry and the targeted wave amplitude and can be calculated in advance.

A major advantage of crosscorrelation methods is that they require fewer sources, as the lack of data can be compensated using virtual sources, so that source depopulation can be applied. We have shown that (up to a certain level) this source depopulation process can be applied without compromising accuracy under either the homogeneous or the laterally inhomogeneous approximations. Source depopulation is more efficient when the homogeneous approximation is used because it allows more averaging. Scattering in the medium enhances the GF reconstruction from crosscorrelation by reducing effects of uneven source distribution.

We proposed a method to characterize the scattering, by evaluating the elastic mean free path of the medium from the ratio between the coherent and the incoherent intensity. The value found for the mean free path indicates that coda waves are present in the records, but limitations of the data set (in particular, the number of sources recorded by each receiver and record length) prevent us from using them efficiently.

ACKNOWLEDGMENTS

This work was part of a collaboration among Shell International Exploration & Production B. V., Rijswijk, The Netherlands, Université Joseph Fourier, Grenoble, France, and Massachusetts Institute of Technology, Cambridge, USA, initiated by the late Gérard Herman. His support and his inspiration are greatly acknowledged. The authors thank the Ministry of Oil and Gas of the Sultanate of Oman, Petroleum Development of Oman and Shell Research for permission to use the data and publish these results. The authors also are thankful to Xander Campman, from Shell, for his involvement in this project, as well as Fabian Ernst, from Shell, for helpful comments. Yuanzhong Fan, Matteo Picozzi, L. Valentina Socco and an anonymous reviewer are acknowledged for their constructive remarks, which helped improve the paper. This work was partially supported by a Shell Research grant.

REFERENCES

Bakulin, A., and R. Calvert, 2006, The virtual source method: Theory and case study: Geophysics, **71**, no. 4, SI139–SI150, doi:10.1190/1.2216190.

Campillo, M., and A. Paul, 2003, Long-range correlations in the diffuse seismic coda: Science, **299**, 547–549, doi:10.1126/science.1078551.

Campman, X., and C. D. Riyanti, 2007, Non-linear inversion of scattered seismic surface waves: Geophysical Journal International, **171**, 1118–1125, doi:10.1111/j.1365-246X.2007.03557.x.

De Rosny, J., and P. Roux, 2001, Multiple scattering in a reflecting cavity: Application to fish counting in a tank: Journal of the Acoustical Society of America, **109**, 2587–2597, doi:10.1121/1.1369101.

Fan, Y., and R. Snieder, 2009, Required source distribution for interferometry of waves and diffusive fields: Geophysical Journal International, **179**, 1232–1244, doi:10.1111/j.1365-246X.2009.04358.x.

Froment, B., M. Campillo, P. Roux, P. Gouédard, A. Verdel, and R. L. Weaver, 2010, Estimation of the effect of nonisotropically distributed energy on the apparent arrival time in correlations: Geophysics, **75**, no. 5, SA85–SA93, doi:10.1190/1.3483102.

Godin, O. A., 2007, Emergence of the acoustic Green's function from thermal noise: Journal of the Acoustical Society of America, **121**, EL96–EL102, doi:10.1121/1.2430764.

Gouédard, P., C. Cornou, and P. Roux, 2008a, Phase-velocity dispersion curves and small-scale geophysics using noise correlation slantstack technique: Geophysical Journal International, **172**, 971–981, doi:10.1111/j.1365-246X.2007.03654.x.

Gouédard, P., P. Roux, M. Campillo, and A. Verdel, 2008b, Convergence of the two-point correlation function toward the Green's function in the context of a seismic prospecting data set: Geophysics, **73**, no. 6, V47–V53, doi:10.1190/1.2985822.

Gouédard, P., L. Stehly, F. Brenguier, M. Campillo, Y. Colin de Verdière, E. Larose, L. Margerin, P. Roux, F. J. Sánchez-Sesma, N. M. Shapiro, and R. L. Weaver, 2008c, Cross-correlation of random fields: mathematical approach and applications: Geophysical Prospecting, **56**, 375–393, doi:10.1111/j.1365-2478.2007.00684.x.

Herman, G. C., and C. Perkins, 2006, Predictive removal of scattered noise: Geophysics, **71**, no. 2, V41–V49, doi:10.1190/1.2187808.

Ishimaru, A., 1978, Wave propagation and scattering in random media: Academic Press.

Korneev, V., and A. Bakulin, 2006, On the fundamentals of the virtual source method: Geophysics, **71**, no. 3, A13–A17, doi:10.1190/1.2196868.
Paul, A., M. Campillo, L. Margerin, E. Larose, and A. Derode, 2005, Empirical synthesis of time-asymmetrical Green functions from the correlation of coda waves: Journal of Geophysical Research, Solid Earth, **110**, B8, B08302, doi:10.1029/2004JB003521.
Picozzi, M., S. Parolai, D. Bindi, and A. Strollo, 2009, Characterization of shallow geology by high-frequency seismic noise tomography: Geophysical Journal International, **176**, 164–174, doi:10.1111/j.1365-246X.2008.03966.x.
Roux, P., W. A. Kuperman, and the NPAL Group, 2004, Extracting coherent wave fronts from acoustic ambient noise in the ocean: Journal of the Acoustical Society of America, **116**, 1995–2003, doi:10.1121/1.1797754.
Sabra, K. G., P. Gerstoft, P. Roux, W. A. Kuperman, and M. C. Fehler, 2005, Surface wave tomography from microseisms in Southern California: Geophysical Research Letters, **32**, L14311, doi:10.1029/2005GL023155.
Schuster, G. T., and M. Zhou, 2006, A theoretical overview of model-based and correlation-based redatuming methods: Geophysics, **71**, no. 4, SI103–SI110, doi:10.1190/1.2208967.
Shapiro, N. M., M. Campillo, L. Stehly, and M. H. Ritzwoller, 2005, High-resolution surface-wave tomography from ambient seismic noise: Science, **307**, no. 5715, 1615–1618, doi:10.1126/science.1108339.
Snieder, R., 2004, Extracting the Green's function from the correlation of coda waves: A derivation based on stationary phase: Physical Review E: Statistical, nonlinear, and soft matter physics, **69**, 046610, doi:10.1103/PhysRevE.69.046610.
Socco, L. V., and D. Boiero, 2008, Improved Monte Carlo inversion of surface wave data: Geophysical Prospecting, **56**, 357–371, doi:10.1111/j.1365-2478.2007.00678.x.
Stehly, L., B. Fry, M. Campillo, N. M. Shapiro, J. Guilbert, L. Boschi, and D. Giardini, 2009, Tomography of the Alpine region from observations of seismic ambient noise: Geophysical Journal International, **178**, 338–350, doi:10.1111/j.1365-246X.2009.04132.x.
Tarantola, A., 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier.
Tarantola, A., and B. Valette, 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Reviews of Geophysics, **20**, no. 2, 219–232, doi:10.1029/RG020i002p00219.
Wapenaar, K., 2006, Nonreciprocal Green's function retrieval by cross correlation: Journal of the Acoustical Society of America, **120**, EL7–EL13, doi:10.1121/1.2208153.
Weaver, R., B. Froment, and M. Campillo, 2009, On the correlation of non-isotropically distributed ballistic scalar diffuse waves: Journal of the Acoustical Society of America, **126**, 1817–1826, doi:10.1121/1.3203359.
Weaver, R. L., and O. I. Lobkis, 2001, Ultrasonics without a source: Thermal fluctuation correlations at MHz frequencies: Physical Review Letters, **87**, 134301, doi:10.1103/PhysRevLett.87.134301.
Weaver, R. L., and O. I. Lobkis, 2004, Diffuse fields in open systems and the emergence of the Green's function (L): Journal of the Acoustical Society of America, **116**, 2731–2734, doi:10.1121/1.1810232.
Yao, H., and R. D. van der Hilst, 2009, Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to SE Tibet: Geophysical Journal International, **179**, 1113–1132, doi:10.1111/j.1365-246X.2009.04329.x.
Yao, H., R. D. van der Hilst, and M. V. de Hoop, 2006, Surface-wave array tomography in SE Tibet from ambient seismic noise and two-station analysis—I. Phase velocity maps: Geophysical Journal International, **166**, 732–744, doi:10.1111/j.1365-246X.2006.03028.x.