9-2013

Using SVD for Improved Interferometric Green's Function Retrieval

Gabriela Melo
Massachusetts Institute of Technology

Alison Malcolm
Massachusetts Institute of Technology

Dylan Mikesell
Boise State University

Kasper van Wijk
Boise State University

Geophysical Journal International

Geophys. J. Int. (2013) **194,** 1596–1612
Advance Access publication 2013 June 27

doi: 10.1093/gji/ggt172

Using SVD for improved interferometric Green's function retrieval

Gabriela Melo,[1] Alison Malcolm,[1] Dylan Mikesell[2,3] and Kasper van Wijk[3]

[1] *Earth Resources Laboratory, Earth, Atmospheric, and Planetary Sciences Department, Massachusetts Institute of Technology, Cambridge, MA* 02139, *USA. E-mail: gmelo@mit.edu*
[2] *Géoazur, Université de Nice Sophia-Antipolis, CNRS, IRD, Observatoire de la Côte d'Azur,* 06560, *Sophia Antipolis, France*
[3] *Physical Acoustic Lab, Department of Geosciences, Boise State University, Boise, ID* 83725, *USA*

Accepted 2013 April 24. Received 2013 April 22; in original form 2012 June 21

SUMMARY

Seismic interferometry (SI) is a technique used to estimate the Green's function (GF) between two receiver locations, as if there were a source at one of the receiver locations. However, in many applications, the requirements to recover the exact GF are not satisfied and SI yields a poor estimate of the GF. For these non-ideal cases, we improve the interferometric GFs, by applying singular value decomposition (SVD) to the cross-correlations before stacking. The SVD approach preserves energy that is stationary in the cross-correlations, which is the energy that contributes most to the GF recovery, and attenuates non-stationary energy, which leads to artefacts in the interferometric GF. We apply this method to construct virtual shot gathers (for both synthetic and field data) and demonstrate how using SVD enhances physical arrivals in these gathers. We also find that SVD is robust with respect to weakly correlated random noise, allowing a better recovery of events from noisy data, in some cases recovering energy that would otherwise be completely lost in the noise and that the standard SI technique fails to recover.

Key words: Time-series analysis; Image processing; Numerical approximations and analysis; Interferometry.

INTRODUCTION

Seismic interferometry (SI), first suggested by Claerbout (1968), can be used to estimate the Green's function (GF) between two receivers, as if there were a source at one of the receiver locations, by cross-correlating the recorded seismic signal at the two stations and stacking the cross-correlations over many sources. The sources can be artificial sources (Schuster *et al.* 2004; Bakulin & Calvert 2006; van Wijk 2006; Mehta *et al.* 2007), earthquakes (Campillo & Paul 2003) or uncorrelated noise (Sabra *et al.* 2005a; Shapiro *et al.* 2005; Roux *et al.* 2005; Weaver 2005; Curtis *et al.* 2006; Godin 2006; Stehly *et al.* 2006). Independent of the source type, a requirement for accurate GF recovery is the receivers record energy from all directions. Unfortunately, this assumption is often not met in practice. As a result, we generally recover a partial estimate of the true GF. This raises the questions: How good an approximation to the GF can SI give in a particular scenario? Can we improve this estimated GF? This work addresses the second question; we present an approach to improve the accuracy of the estimated GF when there is an incomplete source distribution.

There are two general scenarios where SI has been shown to be useful. First, SI can be helpful in places where receivers can be planted but active sources cannot, due to physical or economic reasons. Secondly, even at places suitable for active sources, SI can be applied to re-organize the data in such a way that portions of the data that are normally not considered in traditional imaging techniques can be used. An example of such this is imaging with multiples. The majority of traditional imaging techniques use only singly scattered data. SI can be used to redatum the data in such a way that multiply scattered energy in the original data appears as single-scattered data, allowing for interferometric data to be processed with traditional tools. Generally, this increases the portion of the medium that can be imaged.

To recover the exact GF between two receivers requires that these receivers be surrounded by a closed surface of sources, with both monopole and dipole sources required for accurate amplitude estimates. A number of studies (see e.g. Schuster *et al.* 2004; Snieder 2004; Wapenaar *et al.* 2004a; Roux *et al.* 2005; Sabra *et al.* 2005b; Snieder *et al.* 2006) show that the sources that provide the main contribution to the GFs are the ones located along rays that pass through both receivers, and those in the Fresnel zone around these sources (Snieder 2004). This was shown by approximating the integral over sources using the stationary-phase method and showing that these are the sources for which the phase of the integrand (cross-correlations) is stationary. We refer to these sources as stationary sources. Assuming full source coverage, the rapidly varying energy emanated by sources outside the Fresnel zone destructively interfere; we refer to these sources as non-stationary sources. Incomplete

source coverage and lack of dipole sources results in a degradation of the quality of the recovered GF, which then needs to be carefully interpreted. Furthermore, since dipole sources are rarely available in practice, and source coverage is generally incomplete, here we focus on enhancing arrivals instead of recovering correct amplitudes.

There are various approaches to address the incomplete coverage problem and improve the accuracy of interferometric GFs (see e.g. Bakulin & Calvert 2006; Snieder *et al.* 2006; Wapenaar 2006; Mehta *et al.* 2007; Poliannikov & Willis 2011; van Wijk *et al.* 2011; King & Curtis 2012). A brief description of most of these can be found in the introduction section of King & Curtis (2012). Here we present an approach to alleviate the incomplete coverage problem in certain cases, using the singular value decomposition (SVD, see e.g. Golub & van Loan 1996). This decomposition, as explained below, identifies the stationary signal while suppressing the non-stationary energy in the GF. In addition to these properties, we find that SVD allows the recovery of phases obscured by noise in the regular interferometric GF.

We refer to the collection of cross-correlated traces for a pair of receivers, as the cross-correlogram (one trace for each source). In 2-D (in this work we only examine 2-D data), the cross-correlogram can be viewed as a matrix whose dimensions are time lags from cross-correlations and sources. Thus, by stacking the cross-correlogram along the source dimension, we obtain an interferometric GF. Poliannikov & Willis (2011) suggest viewing the cross-correlogram as the building block for performing interferometry and propose that it should be analysed and pre-processed when necessary before stacking. Here we follow this idea. As mentioned above, in general, there are two distinct types of energy in a cross-correlogram: energy that contributes to forming the interferometric GF (stationary energy) and energy that does not contribute to the GF (non-stationary energy). Stationary energy in the cross-correlogram is characterized by coherency, small wavenumber, and nearly in-phase events along the source dimension. Non-stationary energy, by contrast, is characterized by incoherency, larger wavenumber, and out-of-phase events along the source dimension. It is by separating these two parts of the energy in the cross-correlogram that we obtain more accurate GF estimates for non-ideal source distributions. We use SVD to perform this separation and enhance physical arrivals that are not properly recovered using standard stacking in SI. In this way, we can recover arrivals that would otherwise be obscured by noise. This is similar to the approach used in Freire & Ulrych (1988) and Ulrych *et al.* (1999) to increase the signal-to-noise ratio (SNR) and filter linear events.

Hansen *et al.* (2006) explained the relationship between singular values and frequency; large singular values correspond to low frequencies and small singular values correspond to high frequencies (here frequency refers to source wavenumber in the cross-correlogram.). The large singular values are associated with events that are in phase along the source dimension in the cross-correlogram. The nearly in-phase energy in the cross-correlogram corresponds to energy emanated from the stationary sources. We decompose the cross-correlogram using SVD, construct lower-rank approximations of the cross-correlograms (using the singular values that correspond to the arrivals we are interested in) and stack the lower-rank cross-correlogram to estimate the GF. This is based on the idea of approximating a matrix by another of a lower-rank presented in Eckart & Young (1936).

An intuitive justification for estimating interferometric GFs through a low-rank cross-correlogram comes from the relationship between singular values, frequency, and the stationary-phase method. As mentioned above, interferometric GFs can be obtained by approximating the integral over sources using the stationary-phase method. A solution of an integral obtained through the stationary-phase method is formed by keeping the slowly varying part of the integrand, which is where the integrand's phase is stationary. The stationary part of the integrand thus corresponds to low frequencies in the cross-correlogram space. As demonstrated in Hansen *et al.* (2006), low frequencies correspond to large singular values. Therefore, in drawing a connection between the continuous (GF obtained through an integral) and discrete (GF obtained through summation) cases, solving the SI integral using the stationary-phase method is related to stacking a low-rank approximation of the cross-correlogram obtained through SVD. We illustrate this with both synthetic and field data examples.

The examples we present here are based on the acoustic synthetic, elastic synthetic, and field data from Mikesell *et al.* (2009a,b) and Nichols *et al.* (2010), respectively. In the first example, we apply the SVD technique described above to a version of the acoustic synthetic data set used by Mikesell *et al.* (2009a) contaminated with weakly correlated Gaussian noise. We find that the virtual shot gather obtained through low-rank cross-correlograms created by retaining only the largest singular value, as in a low pass filter, has a larger SNR, thus enhancing the reflected wave that is obscured by the noise in the standard virtual shot gather. In the second example, we apply SVD to the elastic synthetic data set used by Mikesell *et al.* (2009b). They show an improvement in the SNR in cross-correlograms (and by consequence in the virtual shot gathers) in the presence of random noise, by stacking groups of cross-correlograms under the assumption of lateral homogeneity. We demonstrate further improvement in the SNR in the virtual shot gathers by incorporating the SVD technique. Finally, we present results obtained by applying the SVD technique to the field data collected at the Boise Hydrogeophysical Research Site (BHRS, Nichols *et al.* 2010). The source–receiver geometry is similar to the synthetic example. Contrary to the synthetic cases, in this example the reflected and refracted waves can be better recovered by ignoring the largest singular value, as in a high pass filter.

METHOD

In this section, we briefly review the SI method and the underlying assumptions. Then we discuss the proposed SVD method to improve the recovered GF.

Seismic interferometry

SI can be used with sources of different nature such as active sources, earthquakes, noise sources, etc. Here we focus on the active sources scenario. In this case, according to the theory, cross-correlation of the wavefield recorded by two receivers due to a set of monopole and dipole sources completely surrounding both receivers, followed by stacking over all the sources, gives the true impulse response between the receivers. This scenario holds as long as the medium is lossless. Later, we make assumptions and approximations to simplify the SI integral equation, making it suitable for the case when we have only monopole sources.

There are several derivations of the SI equations. For example, they can be derived using time-reversal arguments (Derode *et al.* 2003), representation theorems based on reciprocity theorems (Wapenaar *et al.* 2004b, 2006; Snieder 2007; Snieder *et al.* 2007), superposition of incoming plane waves (Weaver & Lobkis 2004), and the principle of stationary phase (Snieder 2004;

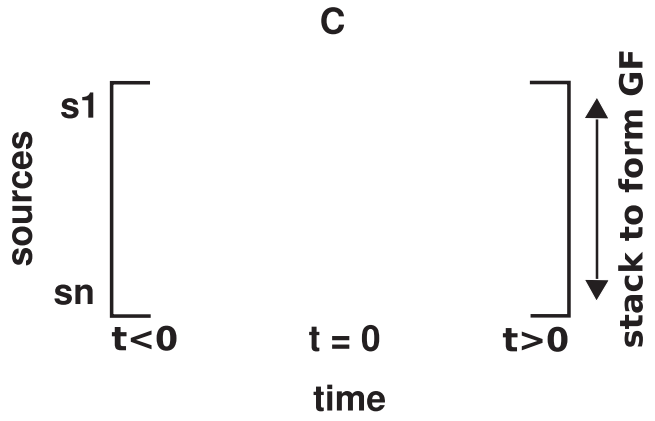

Figure 1. Cross-correlogram matrix $\boldsymbol{C}$. Stacking over sources gives the interferometric GF.

Roux *et al.* 2005). Here we present a brief summary based on the work of Wapenaar & Fokkema (2006) who derive SI equations from representation theorems based on reciprocity (for derivations of reciprocity, see e.g. Rayleigh 1896; Aki & Richards 1980; de Hoop 1988; Fokkema & van den Berg 1993).

Consider two independent acoustic states, at locations A and B, inside a volume V with boundary S, within a time-reversal invariant, lossless and arbitrarily inhomogeneous medium. Let $\hat{G}(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega)$ be the frequency domain GF for a receiver at $\boldsymbol{x}_A$ and a source at $\boldsymbol{x}_B$, where ω is the angular frequency. Based on the reciprocity theorem of the cross-correlation type and source–receiver reciprocity. Wapenaar & Fokkema (2006) show that the sum of the causal and anticausal GF for a receiver at $\boldsymbol{x}_A$ and a source at $\boldsymbol{x}_B$ is given by

$$\hat{G}(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega) + \hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega)$$
$$= \oint_S \frac{-1}{\mathrm{i}\omega\rho(\boldsymbol{x})} \{\hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}, \omega)\partial_i[\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}, \omega)]$$
$$-\partial_i[\hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}, \omega)]\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}, \omega)\}\, n_i \mathrm{d}S, \quad (1)$$

where $\hat{G}(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega)$ corresponds to the causal GF in the time domain, $\hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega)$ is the complex conjugate corresponding to the anticausal GF in the time domain, $\rho(\boldsymbol{x})$ is the density and n_i are the outward-pointing normal vector to the surface S. The physical interpretation of $\partial_i[\hat{G}(\boldsymbol{x}_A, \boldsymbol{x}, \omega)]$ is the GF from a dipole source at $\boldsymbol{x}$ recorded at $\boldsymbol{x}_A$ and $\hat{G}(\boldsymbol{x}_A, \boldsymbol{x}, \omega)$ is the GF from a monopole source at $\boldsymbol{x}$ recorded at $\boldsymbol{x}_A$ [similar interpretation holds for $\partial_i[\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}, \omega)]$ and $\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}, \omega)$]. Cross-correlation in the time domain is equivalent to the product of these GFs in the frequency domain, as seen in the integrand terms, $\hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}, \omega)\partial_i[\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}, \omega)]$ and $\partial_i[\hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}, \omega)]\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}, \omega)$, of eq. (1).

The exact representation of the acoustic GF in eq. (1) requires the computation of two cross-correlation products involving both monopole and dipole sources. In order to make this representation

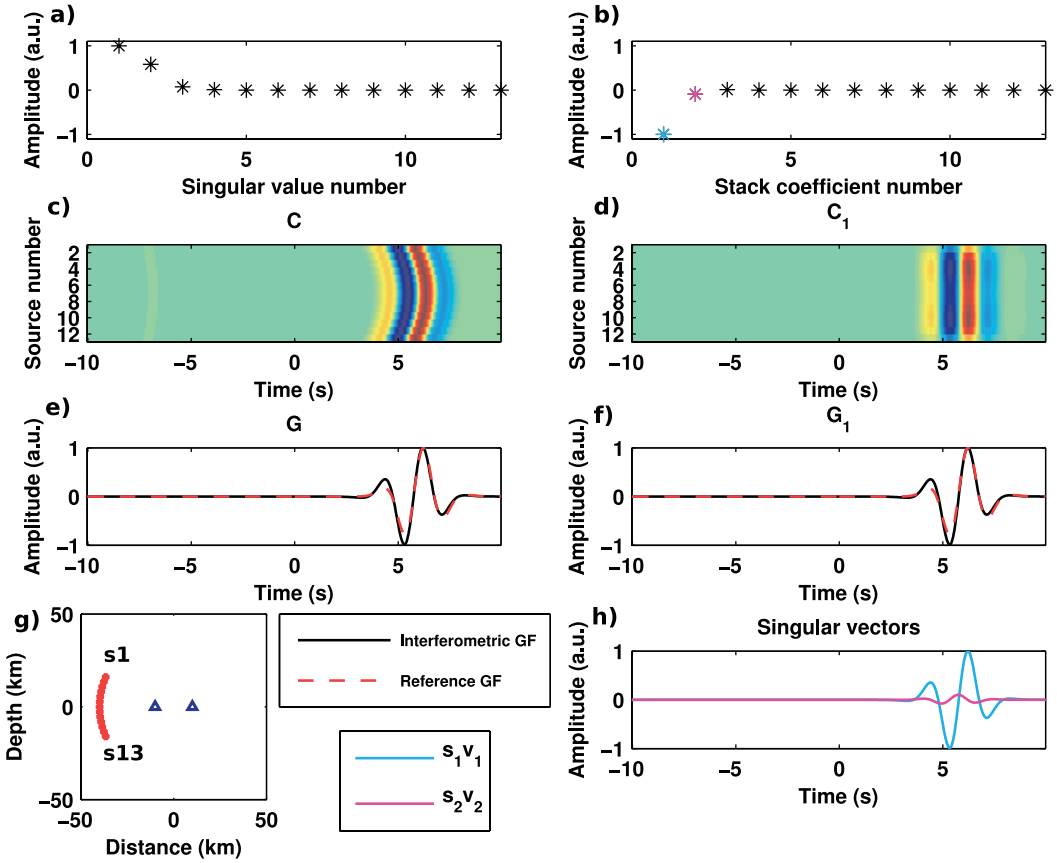

Figure 2. (a) Singular values, σ_k; (b) stack coefficients, s_k; (c) original cross-correlogram, $\boldsymbol{C}$; (d) rank-1 cross-correlogram, $\boldsymbol{C}_1$; (e) standard interferometric GF, G; (f) interferometric GF, G_1, obtained from $\boldsymbol{C}_1$; (g) source–receiver geometry with 13 evenly distributed sources (red stars) around the stationary zone to the left of the receivers (blue triangles); (h) first two singular vectors weighted by the respective stack coefficients. The GFs in (e) and (f) are similar. Even though (a) shows that there are two significant singular values to represent $\boldsymbol{C}$, (b) shows that G can be well represented with only one stack coefficient.

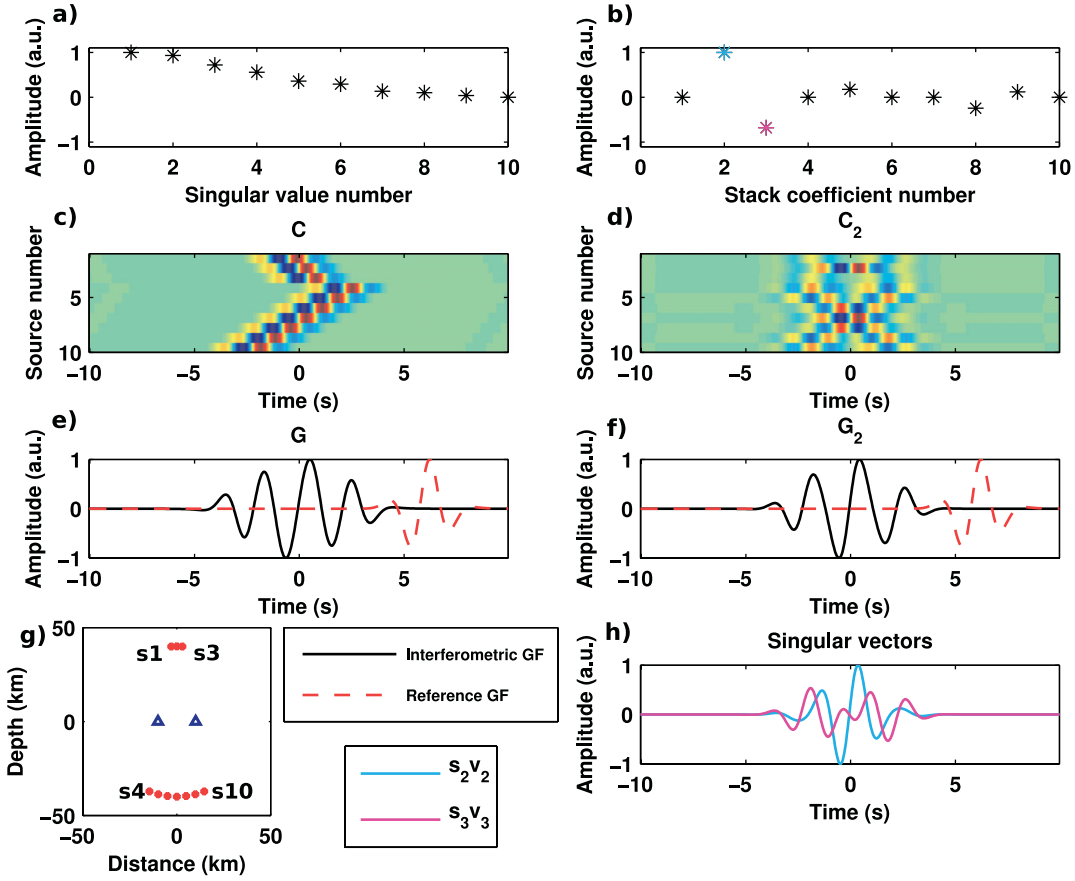

Figure 3. (a) Singular values, σ_k; (b) stack coefficients, s_k; (c) original cross-correlogram, $\boldsymbol{C}$; (d) rank-2 cross-correlogram, $\boldsymbol{C}_2$; (e) standard interferometric GF, G; (f) interferometric GF, G_2, obtained from $\boldsymbol{C}_2$; (g) source–receiver geometry with 3 and 7 sources (red stars) placed in two non-stationary zones with respect to the receivers (blue triangles); (h) singular vectors 2 and 3, weighted by the respective stack coefficients. The singular value spectrum in (a) shows a smooth decay and no significant singular value. The stack coefficient spectrum indicates two significant singular vectors, 2 and 3, contributing to the GF, however, as seen in (h), neither of them resembles the GF, as expected.

more tractable, one can make simplifications in order to, first, reduce the representation to one cross-correlation and, second, to require only monopole sources. First, assuming the medium is homogeneous at and outside of S (with constant velocity c and density ρ) and that no energy comes from outside into S, eq. (1) simplifies to

$$\hat{G}(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega) + \hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega) \approx \frac{2}{\mathrm{i}\omega\rho} \oint_S \partial_i[\hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}, \omega)]\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}, \omega) n_i \, \mathrm{d}S. \quad (2)$$

Secondly, assuming a high frequency regime and that the medium is smooth in a small vicinity around S, the normal derivative $\partial_i[\hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}, \omega)]$ in eq. (2) can be approximated as

$$\partial_i \hat{G}(\boldsymbol{x}_A, \boldsymbol{x}, \omega) n_i \approx \frac{-\mathrm{i}\omega}{c} |\cos[\alpha(\boldsymbol{x})]| \hat{G}(\boldsymbol{x}_A, \boldsymbol{x}, \omega), \quad (3)$$

where $\alpha(\boldsymbol{x})$ is the angle between the ray emanated from $\boldsymbol{x}$ and the normal to S. We assume that S is large enough so rays take off approximately normal to the integration surface S making $\alpha(\boldsymbol{x}) \approx 0$ and $\cos(\alpha(\boldsymbol{x})) \approx 1$. With these assumptions, eq. (2) simplifies to

$$\hat{G}(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega) + \hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega) \approx \frac{2}{\rho c} \oint_S \hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}, \omega)\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}, \omega) \, \mathrm{d}S. \quad (4)$$

In summary, the assumptions in eq. (4) are:

(i) The medium outside the integration surface S is homogeneous, such that no energy going outward from the surface is scattered back into the system.
(ii) The medium around the source is locally smooth.
(iii) All sources lie in the far-field (i.e. the distance from the source to the receivers and scatterers is large compared to the dominant wavelength).
(iv) Rays take off approximately normal to the integration surface S.

Due to these simplifications, the absolute amplitudes of the GF are lost in eq. (4) and errors in amplitude can be large in general. However, since the phase is unaffected, eq. (4) is considered suitable for most applications of SI.

Because our goal here is to enhance arrivals, in the transition from the continuous to the discrete case, we ignore the amplitude factor $\frac{2}{\rho c}$ in eq. (4). Thus, the interferometric GFs are obtained by summation of cross-correlations for all sources

$$\hat{G}(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega) + \hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}_B, \omega) \approx \sum_{i=1}^{N} \hat{G}^*(\boldsymbol{x}_A, \boldsymbol{x}_i, \omega)\hat{G}(\boldsymbol{x}_B, \boldsymbol{x}_i, \omega), \quad (5)$$

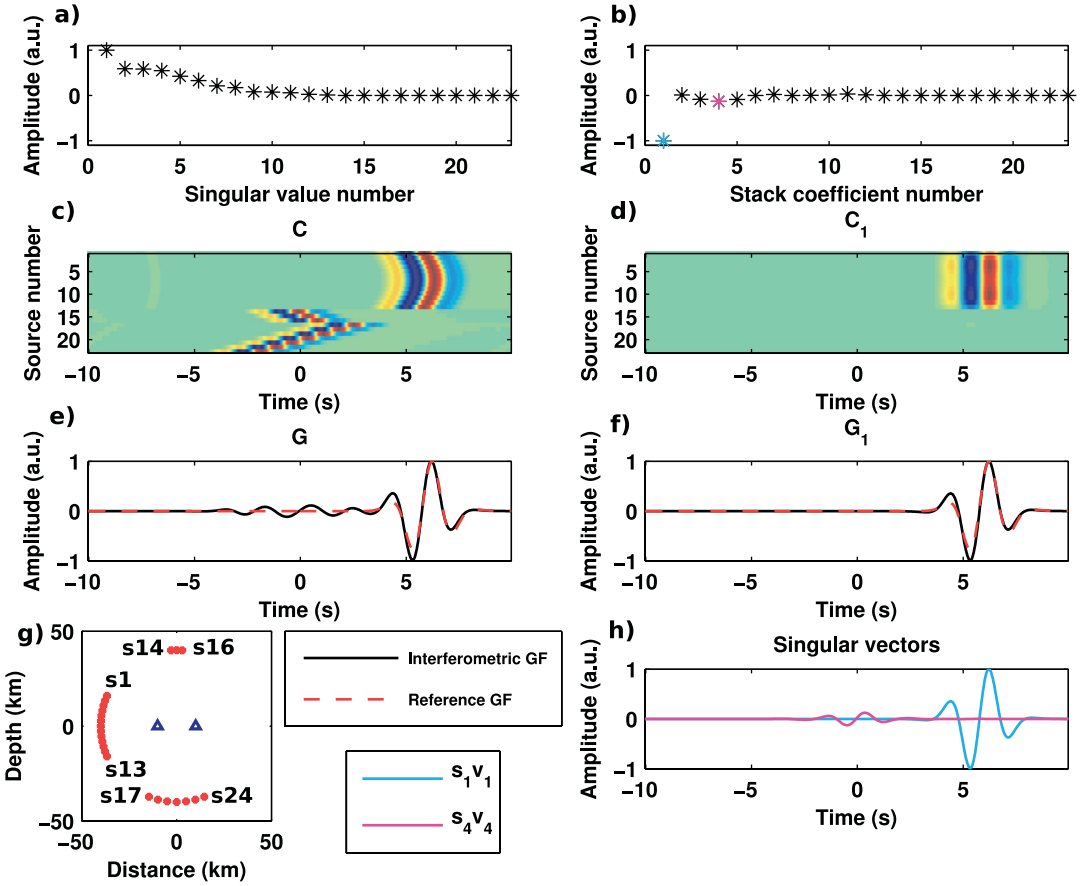

Figure 4. (a) Singular values, σ_k; (b) stack coefficients, s_k; (c) original cross-correlogram, $\boldsymbol{C}$; (d) rank-1 cross-correlogram, $\boldsymbol{C}_1$; (e) standard interferometric GF, G; (f) interferometric GF, G_1, obtained from $\boldsymbol{C}_1$; (g) source–receiver geometry with 13 sources (red stars) in the left-hand side stationary zone, and 3 and 7 sources in two non-stationary zones of the receivers (blue triangles). (h) Singular vectors 1 and 4 weighted by the respective stack coefficients. In (a), we observe a decay that is less smooth than in Fig. 3(a), with a break after singular value 1. In (b) it is clear that the interferometric GF is well represented by the first singular vector. In (f) the fluctuations are reduced and the GF is clearer than in (e). In (h) we see that the second most significant singular vector for the GF, according to (b), is singular vector 4. However, since it corresponds to non-stationary energy in the cross-correlogram, it should be ignored.

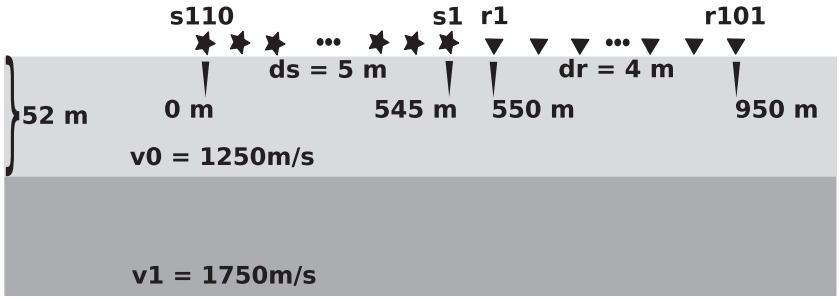

Figure 5. Source–receiver geometry for acoustic synthetic example.

where N is the number of sources. Next we discuss the decomposition of the collection of cross-correlations using SVD in order to isolate energy from stationary sources.

SVD of the cross-correlogram

Let $\boldsymbol{x}_i$ for $1 \leq i \leq N$ be the location of sources and τ be the time lags from cross-correlations in the time domain. We consider the cross-correlogram as the matrix $\boldsymbol{C} = \boldsymbol{C}(\boldsymbol{x}_i, \tau)$ (Fig. 1) where each row is the cross-correlation of the signals recorded at the two receivers from each source. Even though the derivation of the SI equations above are in the frequency domain, our implementation is in the time domain. Thus, the cross-correlogram $\boldsymbol{C}$ can be written as

$$\boldsymbol{C}(\boldsymbol{x}_i, \tau) = \int G(\boldsymbol{x}_A, \boldsymbol{x}_i, t+\tau) G(\boldsymbol{x}_B, \boldsymbol{x}_i, t)\, \mathrm{d}t. \tag{6}$$

Assuming M time samples, $\boldsymbol{C}$ is an $N \times 2M-1$ matrix. The interferometric GF is then obtained by stacking $\boldsymbol{C}$ over the source dimension,

$$G = G(\boldsymbol{x}_B, \boldsymbol{x}_A, t) + G(\boldsymbol{x}_B, \boldsymbol{x}_A, -t) = \sum_i \boldsymbol{C}(\boldsymbol{x}_i, \tau). \tag{7}$$

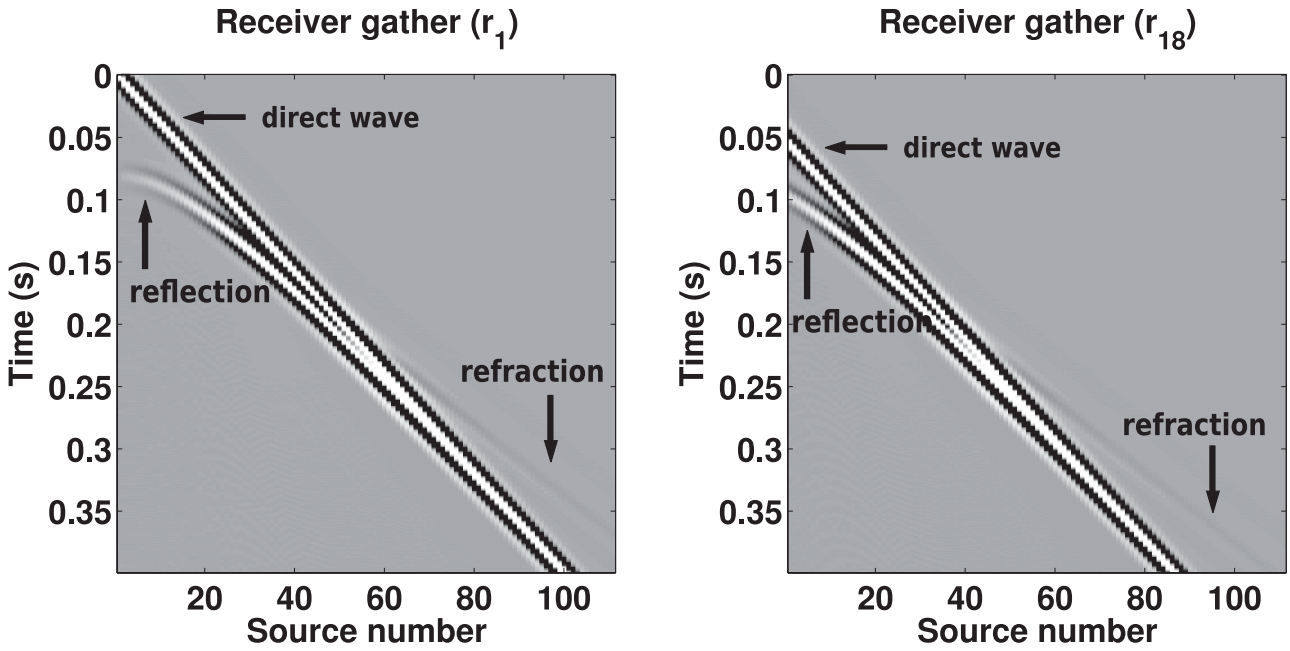

Figure 6. Receiver gathers for receivers r_1 (a) and r_{18} (b).

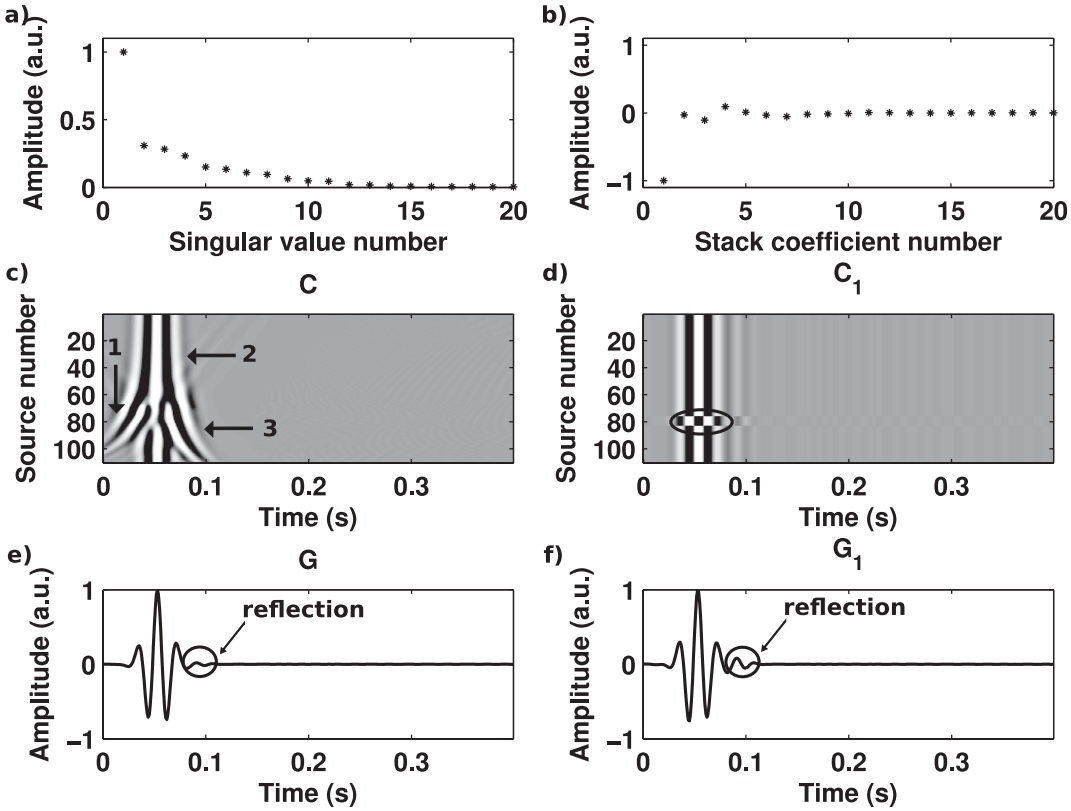

Figure 7. Cross-correlograms and GFs for receivers r_1 and r_{18}: (a) singular values, σ_k; (b) stack coefficients, s_k; (c) original cross-correlogram, $\boldsymbol{C}$; (d) rank-1 cross-correlogram, $\boldsymbol{C}_1$; (e) standard interferometric GF, G; (f) interferometric GF, G_1, obtained from $\boldsymbol{C}_1$. The spectra in (a) and (b) show that there is one significant singular value and stack coefficient. The corresponding singular vector has the same waveform as G_1, so we do not display it here. Note the enhanced reflection in (f) compared to (e). In (d), the cross-correlations for sources 68–75 have their phase reversed. This is because there are three events merging in this zone in the original cross-correlogram as seen in (c). Arrivals 1 and 3 dominate the cross-correlations for sources 68–75 and their phase is approximately the opposite of the linear event 2 in the merging zone. As SVD preserves linearity, it simply reverses the phases for these sources in $\boldsymbol{C}_1$.

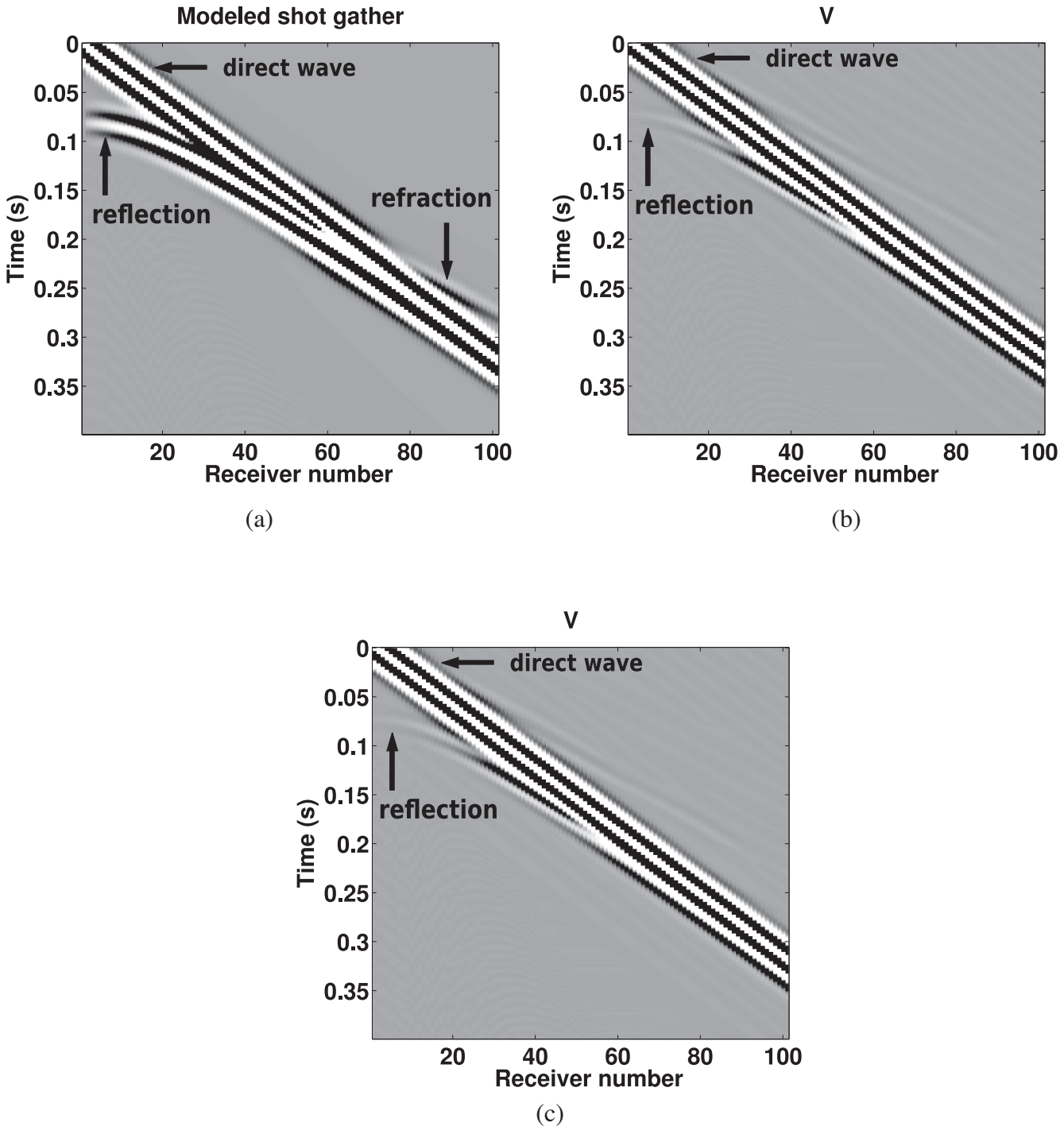

Figure 8. (a) Modelled shot gather with a source placed at the location of receiver r_1. We have convolved the modelled shot gather with a 40 Hz Ricker wavelet used as a source to simulate the change in source signature from SI. (b) Interferometric virtual shot record, V, for virtual source at r_1. Note the recovered direct and reflected waves. (c) Virtual shot gather obtained through rank-1 cross-correlograms, V_1. The reflection is enhanced when compared to the reflection in (b).

Next, we decompose $\boldsymbol{C}$ using SVD (see e.g. Golub & van Loan 1996, for a description of SVD). The SVD decomposition of the cross-correlogram is $\boldsymbol{C} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^t$, where $\boldsymbol{U}$ and $\boldsymbol{V}$ are the left and right singular vectors, respectively, and Σ is the diagonal matrix whose elements are the singular values of $\boldsymbol{C}$. Now we construct $\boldsymbol{\Sigma}_j$ by keeping j singular values of $\boldsymbol{\Sigma}$ and obtain a lower-rank approximation $\boldsymbol{C}_j = \boldsymbol{C}_j(\boldsymbol{x}_i, \tau) = \boldsymbol{U}\boldsymbol{\Sigma}_j\boldsymbol{V}^t$. As mentioned above, this is based on the idea of approximating a matrix by another of a lower-rank, as discussed in Eckart & Young (1936). Stacking the rows of $\boldsymbol{C}$ (eq. 7) gives the standard interferometric GF, G, and stacking the rows of the approximation $\boldsymbol{C}_j$ gives the modified interferometric GF, G_j.

$$G_j = G_j(\boldsymbol{x}_B, \boldsymbol{x}_A, t) + G_j(\boldsymbol{x}_B, \boldsymbol{x}_A, -t) = \sum_i \boldsymbol{C}_j(\boldsymbol{x}_i, \tau), \qquad (8)$$

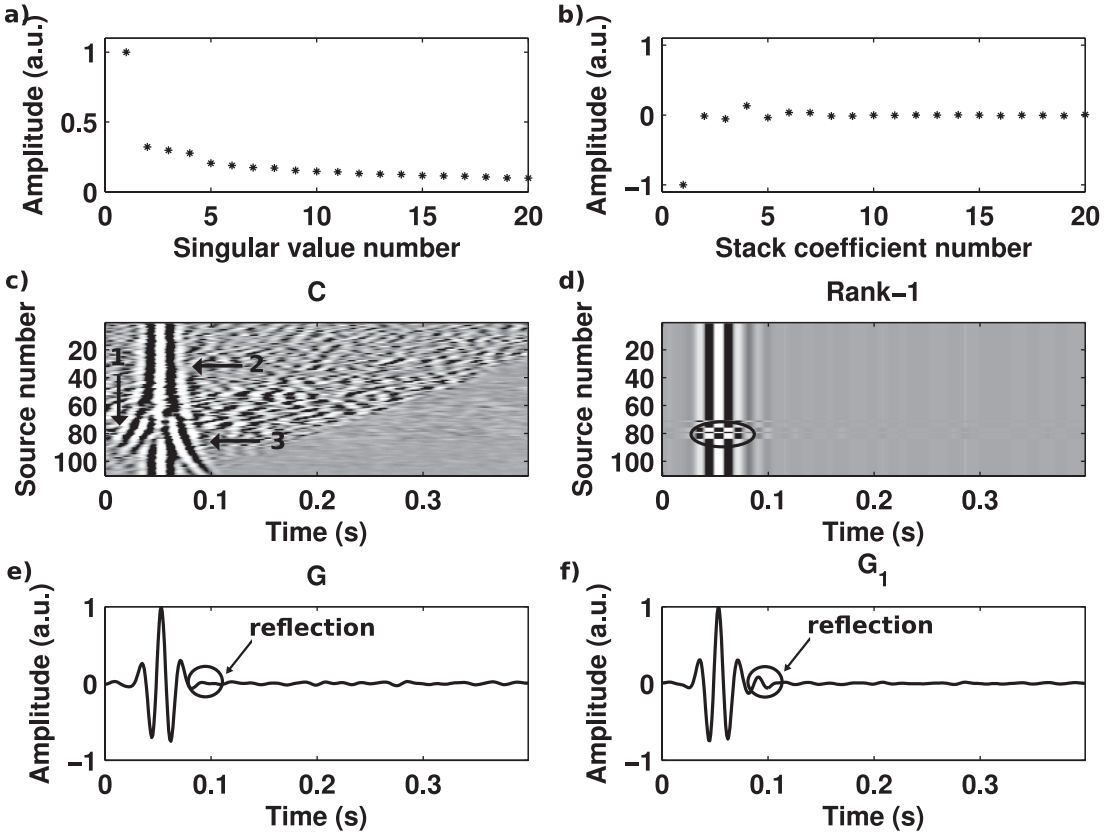

Figure 9. Cross-correlograms and GF for receivers r_1 and r_{18}: (a) singular values, σ_k; (b) stack coefficients, s_k; (c) original cross-correlogram, $\boldsymbol{C}$; (d) rank-1 cross-correlogram, $\boldsymbol{C}_1$; (e) standard interferometric GF, G; (f) interferometric GF, G_1, obtained from $\boldsymbol{C}_1$. Noise has been added to the synthetic data so that the reflection is difficult to see in the standard virtual shot gather. Similar to the clean data set in Fig. 7, the spectra in (a) and (b) show that there is one significant singular value and stack coefficient. The corresponding singular vector has the same waveform as G_1, so we do not display it here. Again, note the enhanced reflection in (f) compared to (e), as in Fig. 7. In (d), the cross-correlations for sources 68–75 have their phase reversed. This is because there are three events merging in this zone in the original cross-correlogram as seen in (c). Arrivals 1 and 3 dominate the cross-correlations for sources 68–75 and their phase is approximately the opposite of the linear event 2 in the merging zone. As SVD preserves linearity, it simply reverses the phases for these sources in $\boldsymbol{C}_1$.

According to the SVD based cross-correlogram decomposition, we note that G_j can be viewed as a weighted sum of the left singular vectors (rows of matrix $\boldsymbol{V}$). Let $\boldsymbol{e}$ be a vector of dimensions $1 \times N$, whose elements are all equal to 1. Here $\boldsymbol{e}$ is just an auxiliary vector we use to to write the stack of the rows of $\boldsymbol{U\Sigma}$ in matrix notation. Then, the interferometric GF can be written in matrix notation as

$$G = \boldsymbol{eC} = \boldsymbol{eU\Sigma V}^t = \boldsymbol{s}V^t \ , \tag{9}$$

where $\boldsymbol{s} = \boldsymbol{eU\Sigma}$ are the coefficients of the weighted sum of the singular vectors in $\boldsymbol{V}$. From here on we refer to these coefficients as stack coefficients. Let u_{ik} correspond to the elements of matrix $\boldsymbol{U}$, $\boldsymbol{v}_k$ correspond to the kth row of matrix $\boldsymbol{V}$ and σ_k be the singular values. Thus, eq. (7) can be rewritten as

$$G = \sum_k \sigma_k \left(\sum_i u_{ik} \right) \boldsymbol{v}_k = \sum_k s_k \boldsymbol{v}_k, \quad 1 \le k \le N. \tag{10}$$

In the examples that follow, the singular vectors shown are the rows of $\boldsymbol{V}$, and when selecting s_k's we consider their absolute value.

We now illustrate this procedure with a synthetic acoustic homogeneous model. The model for this example is a constant velocity and density model with no reflectors. Therefore, the GF consists of the direct wave only. We examine how we can approximate the true GF in three cases: (i) the case where there are stationary sources only, (ii) non-stationary sources only and (iii) both stationary and non-stationary sources. In all three cases there are gaps in the source distribution and, for comparison, all the GFs are normalized to have a peak amplitude of one.

First, we consider the case where the sources are only in the stationary-phase zone, as in Fig. 2(g). The energy from these sources contributes constructively to the GF. The spectra in Figs 2(a) and (b) show that while there are two significant singular values to represent $\boldsymbol{C}$ (Fig. 2c) only one stack coefficient, the first one, should be required to well-approximate the GF. Fig. 2(d) shows $\boldsymbol{C}_1$ constructed using only the first singular value/vector. Figs 2(e) and (f) show that the GF obtained from $\boldsymbol{C}$ and $\boldsymbol{C}_1$ are quite similar. This is a case where standard interferometry works well and the SVD technique is not necessary, although it is not detrimental. The singular vectors (Fig. 2h) are weighted by the corresponding stack coefficients ($s_k \boldsymbol{v}_k$) and, again, is clear that the GF is well approximated by the first singular vector.

In case (ii) we take only non-stationary sources (Fig. 3g). Ideally (i.e. assuming full source coverage), all of the non-stationary energy should cancel during the stack over sources. However, if there are gaps in the source distribution, residual energy will remain because of the imperfect cancellation of the non-stationary energy. The singular value spectrum in Fig. 3(a) shows a smooth decay, that is, there is no obvious truncation point of significant singular values other than when the values approach zero. On the other hand, the stack coefficient spectra shows that there are primarily two singular

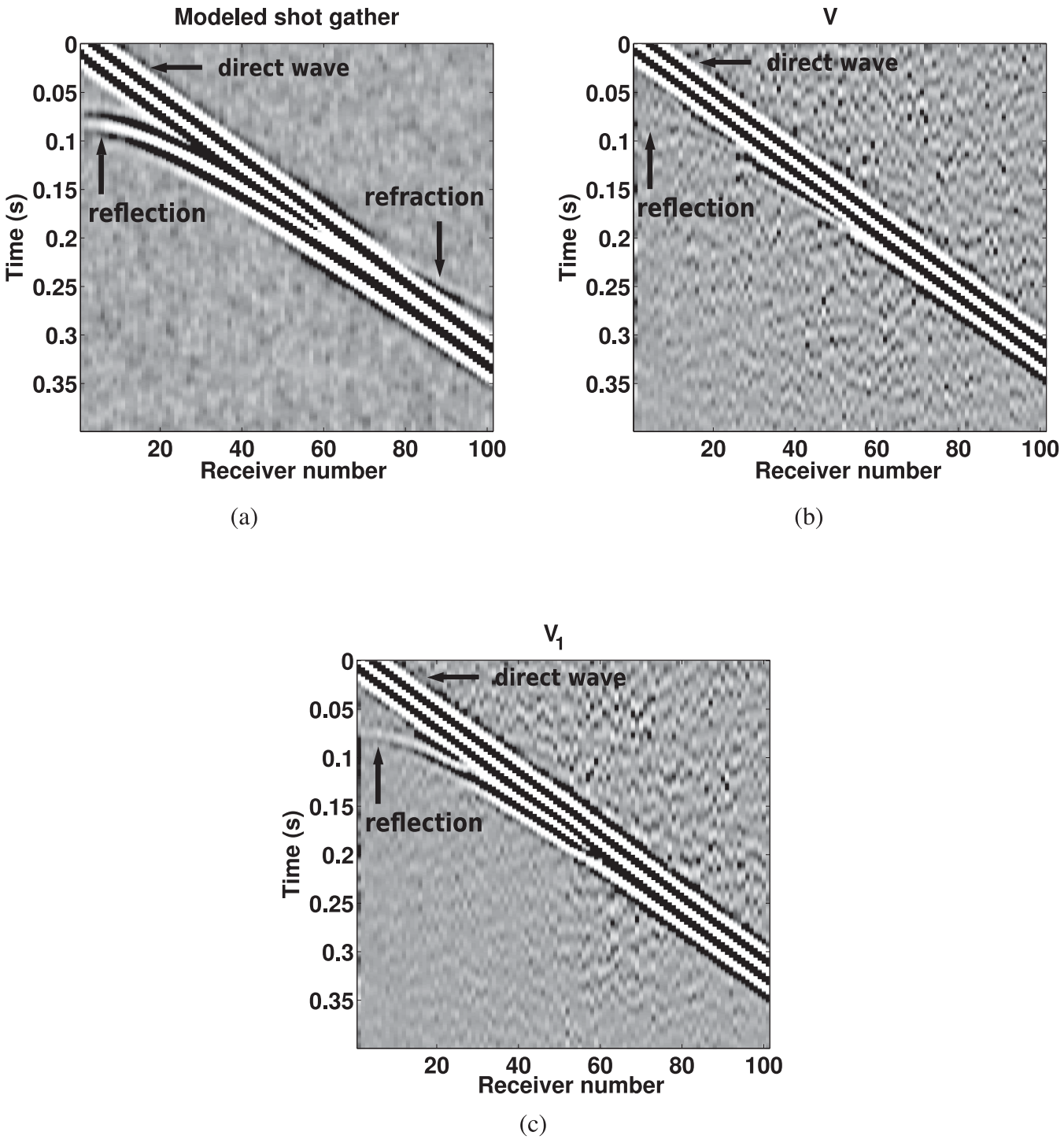

Figure 10. (a) Modelled shot gather, similar to Fig. 8(a), contaminated with noise. (b) Standard virtual shot gather, V, similar to Fig. 8(b). (c) Virtual shot gather obtained through rank-1 cross-correlogram, V_1. Note the enhanced reflected wave compared to the reflected wave in the standard virtual shot gather in Fig. 10(b).

vectors, 2 and 3, that contribute to the GF. Thus we construct a rank-2 cross-correlogram approximation, $\boldsymbol{C}_2$, in Fig. 3(d). In this case, $\boldsymbol{C}_2$ does not enhance any linearity and does not even resemble $\boldsymbol{C}$ (Fig. 3c). Singular vectors 2 and 3, weighted by the stack coefficients, are shown in Fig. 3(h). They correspond to the non-stationary energy and, contrary to the previous case, none of them resembles the GF.

Case (iii) mixes the two previous cases. Fig. 4(g) shows sources in stationary and non-stationary zones, but with gaps in between. The cross-correlogram (Fig. 4c) thus has energy contributing to the GF and energy that should cancel out completely. However, because of the gaps, it does not. In the singular value spectrum (Fig. 4a) we observe a mixture of the two previous cases, a break after the first singular value followed by smooth decay. Fig. 4(b) again indicates

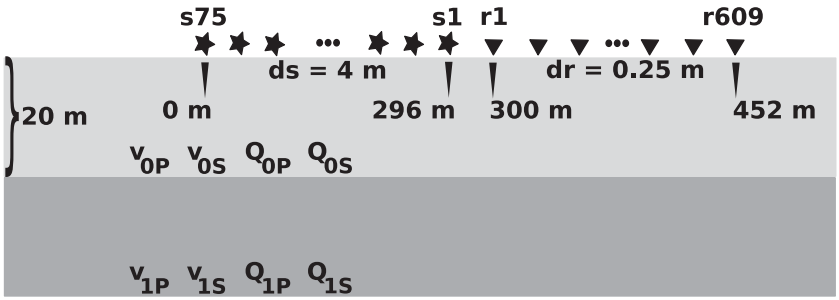

Figure 11. Source–receiver geometry for group 1 of elastic synthetic example.

that the interferometric GF can be well represented with only the first singular vector. Fig. 4(d) shows $\boldsymbol{C}_1$, constructed using only the first singular vector, which corresponds to the stationary energy. This rank-1 approximation thus suppresses the residual energy caused by the imperfect cancellation of non-stationary energy, and G_1 is more accurate than G as seen in Figs 4(e) and (f). Fig. 4(h) shows singular vectors 1 and 4 (the strongest stacking coefficients are number 1 and 4) weighted by the respective stack coefficients. An intuitive reason why SVD is able to capture stationary energy (here and in the following examples) in the cross-correlogram is because a rank-1 matrix obtained through SVD will consist of one row that best represents the original matrix, thus, qualitatively speaking, it will capture what is most in common among all rows, which is stationary energy. The same argument can be made for columns. We see again that, for instance, while singular vector 4 is significant for representing $\boldsymbol{C}$ it has little contribution to G, and because it corresponds to non-stationary energy in the cross-correlograms, it should be ignored.

DATA APPLICATIONS

In this section, we present three examples—acoustic synthetic, elastic synthetic and field data. We first apply the SVD technique to a version of the acoustic synthetic data set used by Mikesell *et al.* (2009a) contaminated with random noise. The virtual shot gather obtained through decomposing the cross-correlograms using SVD and retaining only the singular value corresponding to the largest stacking coefficient in absolute value, as in a low pass filter, has a larger SNR thus enhancing the reflected wave that is obscured by the noise in the standard virtual shot gather. In the second example we use the elastic synthetic data set used by Mikesell *et al.* (2009b), where we obtain improvements in the virtual shot gather's SNR by incorporating the SVD technique. Both synthetic wavefields were modelled using a spectral element method (Komatitsch & Vilotte 1998; Komatitsch & Tromp 2002). Finally, we present results obtained by applying the SVD technique to the field data collected at the BHRS (Nichols *et al.* 2010). In this case, the reflected and refracted waves can be better recovered by ignoring the largest singular value/stacking coefficient, as in a high pass filter.

Acoustic synthetic example

Here we apply the SVD technique discussed above to the same synthetic data set used by Mikesell *et al.* (2009a). Consider the 2-layer acoustic model shown in Fig. 5. The top layer has velocity $v_0 = 1250$ m s^{-1}, the bottom layer has velocity $v_1 = 1750$ m s^{-1} and density is constant throughout the model. A 2-D array of 110 sources is placed to the left of the receiver line (Fig. 5); the wavefield generated by each source is recorded at each receiver. The source is a 40 Hz Ricker wavelet.

Now, we create a virtual shot gather as if there were a source at receiver r_1 using SI. The receiver gather for receiver r_1 is shown in Fig. 6(a). The GF between r_1 and each of the other receivers is obtained with SI. For example, consider the cross-correlogram between receivers r_1 and r_{18}. The receiver gather for receiver r_{18} is shown in Fig. 6(b). The spectra in Figs 7(a) and (b) show that there is one significant singular value and stack coefficient. Figs 7(c) and (d) show the standard and the rank-1 causal cross-correlograms for receivers r_1 and r_{18}. Figs 7(e) and (f) show the standard cross-correlogram stack, G, and the rank-1 cross-correlogram stack, G_1. The amplitude of the reflected arrival is enhanced in G_1 in comparison with G.

Repeating this procedure for all receivers, we create a standard virtual shot gather, V (Fig. 8b) and a modified virtual shot gather, V_1, from the rank-1 cross-correlograms (Fig. 8c) for a virtual source at r_1. For comparison, Fig. 8(a) shows the modelled shot gather for a source at the location of receiver r_1. As is expected from the results in Fig. 7, the reflection is clearer in V_1 than in V. The refraction cannot be seen in either of the virtual shot gathers due to its low amplitude. Each trace in the shot records (modelled and virtual) are normalized individually such that all direct arrivals have a peak amplitude of 1, and all gathers are displayed on the same grey-scale.

We now add weakly-correlated Gaussian noise to the data, before cross-correlation, to test the SVD method's robustness with respect to noise. The noise level, about 1 per cent of the direct wave peak amplitude, was just enough so the refraction is lost and the reflection is very weak in the standard virtual shot gather. Similar to the clean data set, the spectra in Figs 9(a) and (b) show that there is one significant singular value and stack coefficient. Figs 9(c) and (d) show again the standard and rank-1 cross-correlograms for r_1 and r_{18}. Figs 9(e) and (f) show the respective interferometric GFs. We again see that the amplitude of the reflected wave is enhanced in G_1 in comparison with G. Fig. 10(a) shows the modelled shot gather plus noise. Further, as seen in Figs 10(b) and (c), the reflection is visible in V_1 whereas it is obscured by noise in V.

Elastic synthetic example

Next, we move to a more realistic, noisy, elastic synthetic data. Mikesell *et al.* (2009b) used this data set to show how to improve the SNR in cross-correlograms and, consequently, in the virtual shot gathers, by stacking groups of cross-correlograms under the assumption of lateral homogeneity. When lateral homogeneity does not hold, the same technique can be used if multiple-fold data is available. Here multiple-fold data means that for each source we produce multiple shot gathers that differ from each other only by the addition of a different realization of random noise. Assuming

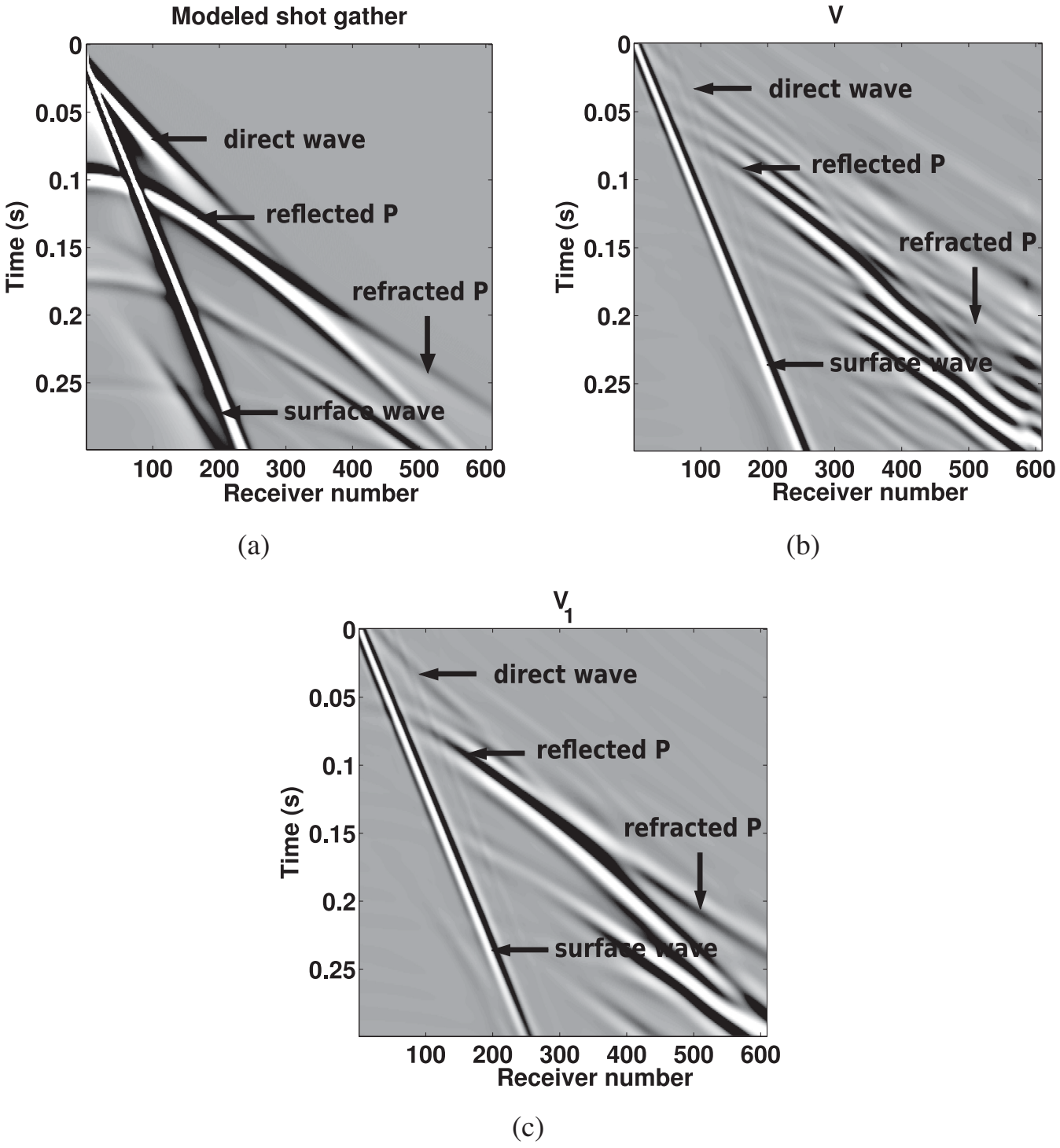

Figure 12. (a) Example of a shot gather, for synthetic elastic data, corresponding to a source located at the position of receiver r_1. (b) Standard virtual shot gather, V, for noise-free elastic data set. The arrivals are poorly recovered other than the Rayleigh wave. (c) SVD-enhanced virtual shot gather, V_1, for noise-free elastic data set. Note how the arrivals are better recovered in comparison with (b).

that multiple-fold data is available, we study four different ways to possibly improve the SNR in virtual shot gathers, including the SVD technique presented above, then apply and compare the four approaches to two different data folds.

The model consists of an array of sources and an array of receivers on the surface of a low velocity layer underlain by a faster velocity layer as depicted in Fig. 11. The source and receiver arrays are 296 and 152 m long, respectively, and the spacings are 4 and 0.25 m, respectively. The top layer is 20 m thick with velocities of $v_{0,P} = 1000$ m s^{-1} and $v_{0,S} = 400$ m s^{-1}. The lower half-space has velocities $v_{1,P} = 1550$ m s^{-1} and $v_{1,S} = 600$ m s^{-1}. In order to help attenuate free-surface multiples and better represent a near-surface of unconsolidated sediment, attenuation was included in the modelling. The Q values for the top layer are $Q_{1,P} = 60$ and

$Q_{1,S} = 10$, and for the bottom layer are $Q_{2,P} = 100$ and $Q_{2,S} = 30$. The wavefield was modelled using an impulsive vertical impact source with a dominant frequency of 120 Hz. More details about this data set can be found in Mikesell *et al.* (2009b).

We start by constructing virtual shot gathers from the noise-free data, and then we proceed to study a noisy version of this data set. Fig. 12(a) shows the shot gather for the source located at the position of receiver r_1. The SVD decomposition for all cross-correlograms in this case generally follow the pattern in the acoustic case of having the first stack coefficient as the most significant. Figs 12(b) and (c) show the standard and the SVD-enhanced virtual shot gathers, V and V_1, respectively. Most arrivals are better recovered and have correct arrival times in the SVD-enhanced virtual shot gather, while many of the arrivals are distorted in the standard virtual shot gather due to the presence of non-stationary energy that does not cancel completely during the cross-correlogram stack. This is clearly seen in the few annotated phases (direct, reflected *P*, and refracted *P*) in Figs 12(b) and (c). Note that there is a systematic phase difference in the source wavelet between the modelled and the virtual shot gathers. This happens because the source wavelet is squared in the cross-correlations (Snieder 2004; Wapenaar & Fokkema 2006).

Assume now we have a *n*-fold data set. Each single-fold data set consists of the clean data plus weakly correlated Gaussian noise, with a noise level of about 40 per cent of the direct wave peak amplitude, added to the modelled data before the cross-correlations. Enough noise was added such that, for each single-fold data, the events lose coherency in the cross-correlogram domain, and the arrivals are completely obscured by noise in the corresponding virtual shot gathers. To improve the SNR in virtual shot gathers, we study four different approaches.

In approach one, we use a stacking procedure where virtual shot gathers are constructed from stacked cross-correlograms, similar to what is done in Mikesell *et al.* (2009b). For each single-fold data we generate one cross-correlogram. Cross-correlograms are then stacked to form an *n*-fold cross-correlogram, which are then used to construct a standard virtual shot gather. The stacking greatly enhances coherency in the cross-correlogram space allowing the recovery of some of the events that were previously obscured by noise. Examples of such virtual shot gathers for two different folds (6 and 38) are shown in Figs 13(a) and 14(a). Comparing these virtual shot gathers with the modelled shot gather for the noise-free data in Fig. 12(a), we see that some events are still missing.

The second approach consists of reducing the noise by filtering the multiple-fold data for every source–receiver pair using SVD. Assume we have *n*-fold data. For every source–receiver pair, we collect the *n* corresponding recordings. Data in these traces are the same other than the noise, thus forming a rank-1 data set. Now we simply apply SVD to this set of similar traces to suppress the non-stationary energy, and use the first singular vector in the cross-correlogram. Applying SVD in this manner is an alternative to the common offset stack (as in approach one) and has the same goal of increasing the SNR. The virtual shot gathers are then constructed as normal. Figs 13(b) and 14(b) show the resultant virtual gathers for 6- and 38-fold data. These gathers have a better SNR than the respective gathers obtained through approach one, as seen in Figs 13(a) and 14(a).

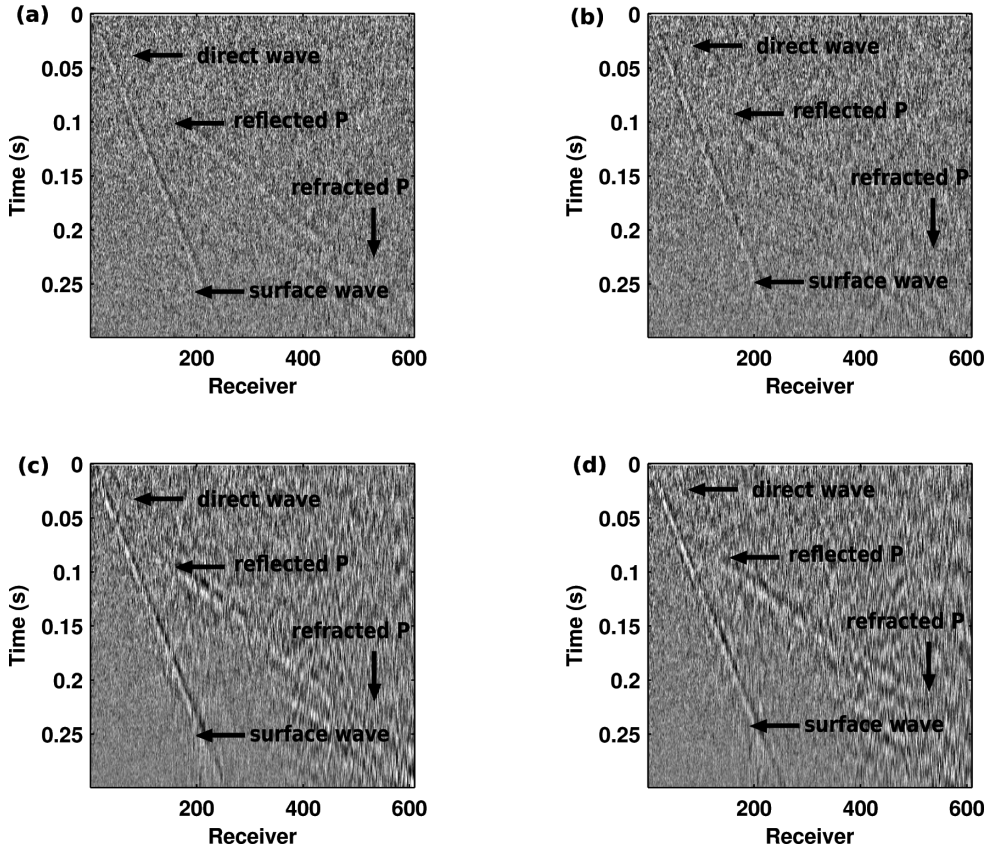

Figure 13. (a) 6-fold standard virtual shot gather; (b) virtual shot gather with 6-fold SVD-filtered data; (c) 6-fold SVD-enhanced virtual shot gather; (d) 6-fold double-SVD enhanced virtual shot gather.

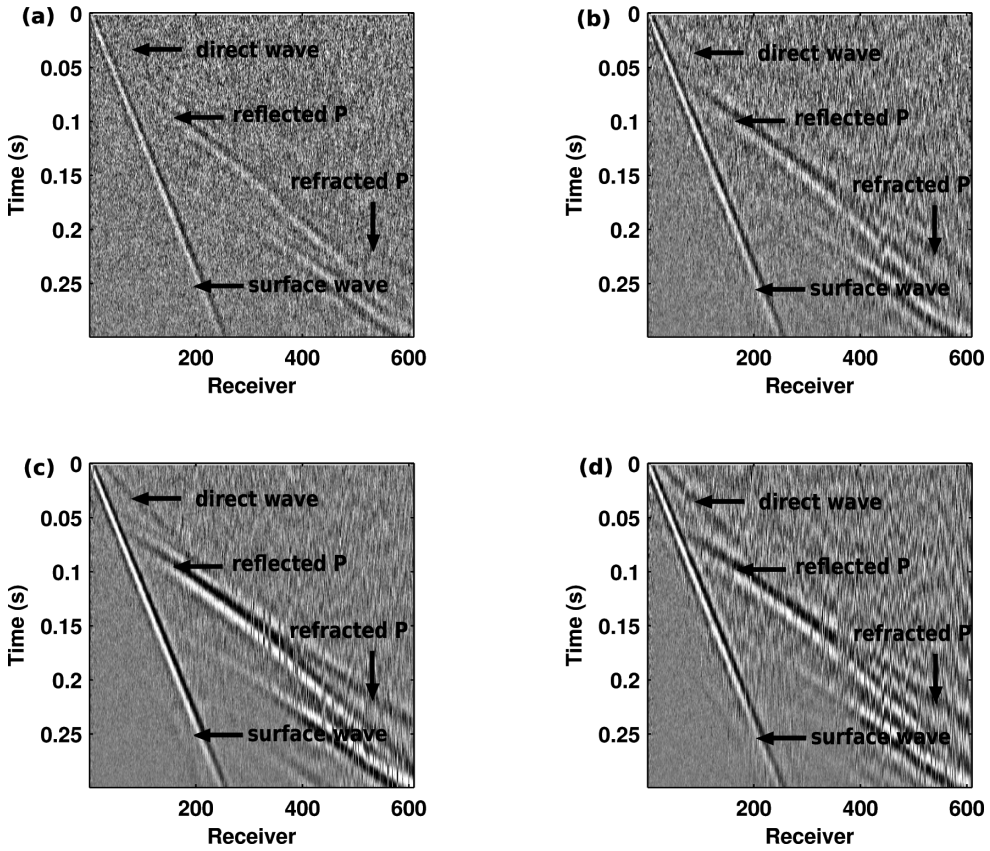

Figure 14. (a) 38-fold standard virtual shot gather; (b) virtual shot gather with 38-fold SVD-filtered data; (c) 38-fold SVD-enhanced virtual shot gather; (d) 38-fold double-SVD enhanced virtual shot gather.

The third approach is a combination of approach one and the SVD technique. First, the *n*-fold cross-correlograms are constructed through stacking, as in approach one. Then, we apply the SVD technique to these *n*-fold cross-correlograms to form *n*-fold SVD-enhanced virtual shot gathers (from rank-1 cross-correlograms). Figs 13(c) and 14(c) show the resultant gathers for 6- and 38-fold data. Comparing them with the respective gathers from the two previous approaches (Figs 13a, b and 14a, b), we see that the *n*-fold SVD-enhanced gathers converge faster with increased fold and have a much better SNR, revealing events that are not recovered in the previous two approaches.

The fourth approach is a combination of approach two and the SVD technique. We first construct the cross-correlograms as in the second approach and then apply the SVD technique. Here we also use rank-1 cross-correlograms. We call these double-SVD enhanced virtual shot gathers and they are shown in Figs 13(d) and 14(d). The SNR is better than in the virtual shot gathers from approach one. In addition, it is also generally better than the gathers from the approach two. However, gathers from the third approach generally have a better SNR and fewer phase reversals/oscillations, particularly at larger offsets. We therefore conclude that the key place to apply SVD is to the cross-correlograms and not to the common offset sections.

In all the cases for this elastic data set we used rank-1 approximations for all cross-correlograms. We analysed different ranks for the cross-correlograms in two different ways. First, we looked at SVD-enhanced virtual shot gathers (approaches three and four) from cross-correlograms of increasing ranks, but keeping the same rank number for all cross-correlograms for a given virtual shot gather. Secondly, we created virtual shot gathers from cross-correlograms of different ranks: we defined thresholds and constructed each cross-correlogram by keeping the stacking coefficients above a given threshold, that is, each trace in the virtual shot gather comes from a cross-correlogram of different rank. We found that for this particular data set there was visually no improvement over the rank-1 results. Therefore, we used the rank-1 cross-correlograms in all cases. For other data sets, individual analysis of the stacking coefficient spectra may be advantageous.

For this data set, in general, additionally applying SVD to the cross-correlograms leads to a better noise suppression than only stacking or SVD-filtering common offset data. Also, combining the SVD-filtered common offset data with SVD decomposed cross-correlograms gives, in general, little or no improvements.

Field data example

We now present the results from applying the SVD filtering technique to the field data set presented in Nichols *et al.* (2010). This data set is from a 2-D seismic survey conducted at the BHRS. Here we use data from an array of 108 receivers: 74 receivers with 1 m spacing in the centre of the line and 17 receivers with 0.25 m spacing at each end of the receiver line. The source is a 4 lb sledge hammer and the source array extends for 39.9 m starting at 0.1 m to the left of the first receiver; four shots were stacked every 0.1 m for the first 1.9 m and then shot spacing was increased to 1 m for another 38 m. Fig. 15 shows the source–receiver geometry. A time sampling interval of 0.25 ms was used. In general, energy from ground roll,

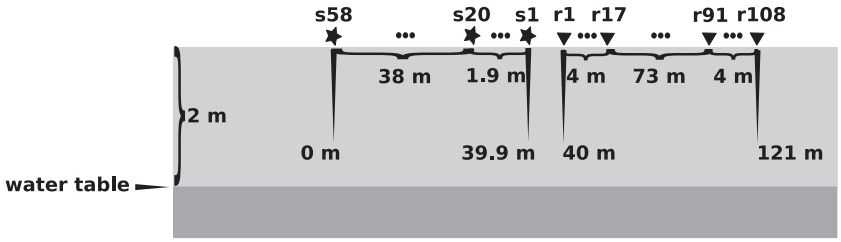

Figure 15. Source–receiver geometry for field data example. Stars and triangles represent sources and receivers, respectively. The spacing for the receiver array is 0.25 m for r_1–r_{17}, 1 m for r_{17}–r_{91} and 0.25 for r_{91}–r_{108}. The spacing for the source array is 0.1 m for s_1–s_{20} and 1 m for s_{20}–s_{77}. The average depth to the water-table is 2 m.

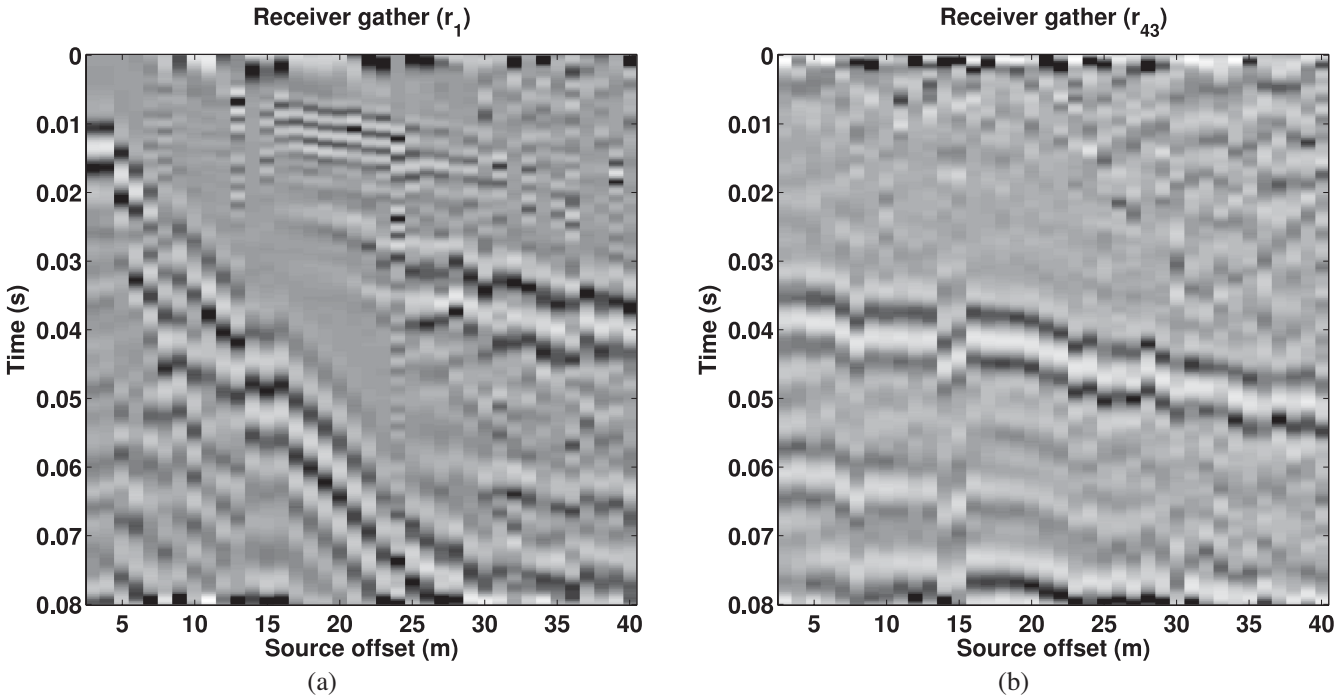

Figure 16. Examples of receiver gathers for receivers (a) r_1 and (b) r_{43}.

a refraction from the water-table (at approximately 2 m in depth), and deeper reflections can be seen in the shot and receiver gathers. The data were filtered and gained to suppress the ground roll and emphasize events associated with a water-table interface. For more details about these data the and geology of the field, see Nichols *et al.* (2010) or Mikesell & van Wijk (2011).

Similar to the synthetic examples above, the goal here is to produce virtual shot gathers for a virtual source located at the first receiver, r_1. In virtual shot gathers of two-layered models with the velocity in the deeper layer higher than in the shallower one, an artefact named the virtual refraction arises from the cross-correlation of head waves recorded at the two receivers. The virtual refraction can be used to estimate the depth to the interface and the velocity in the deeper layer (Mikesell *et al.* 2009a,b; Nichols *et al.* 2010) as well as input for a delay-time statics method (Mikesell *et al.* 2012). The virtual shot gather for the field data presented here has a very strong virtual refraction. This artefact is not so prominent in the synthetic examples above.

Similar to the synthetic examples above, GFs in the virtual shot gathers here correspond to one strong stack coefficient from the SVD decomposition. As an example, we study the SVD decomposition for the receiver pair r_1 and r_{43}. Figs 16(a) and (b) shows the receiver gathers for these two receivers. Fig. 17(c) shows the cross-correlogram between receivers r_1 and r_{43}. The singular value and stack coefficient spectra in Figs 17(a) and (b), respectively, show one strong singular value and stack coefficient. The first five (normalized) singular vectors are shown in Fig. 17(d). Note that singular vectors 1–3 mostly correspond to energy up to 0.05 s, whereas singular vectors 4–5, while also containing energy before 0.05 s, contain relatively more energy after 0.05 s compared to singular vectors 1–3. This indicates that higher order singular values correspond to later events.

Fig. 18(a) shows the virtual shot record, V, for the virtual source at r_1. The amplitudes of the virtual refraction are considerably stronger than the real refraction and reflection because the virtual refraction has more stationary energy in the cross-correlogram. Fig. 18(b) shows the virtual shot gather obtained through rank-1 cross-correlogram approximations, V_1. The virtual shot gather V_1 is dominated by the virtual refraction and the amplitudes of the real refraction and reflection are weak. This indicates that the largest singular value of the cross-correlogram corresponds primarily to the virtual refraction. Thus, we must look at lower singular values/stacking coefficients to enhance the real refraction and reflection, as previously observed from the singular vectors in Fig. 17(d). Fig. 18(c) shows a virtual shot gather, V_{n-1} (here there are 58 sources, thus $n = 58$), constructed by removing the largest singular

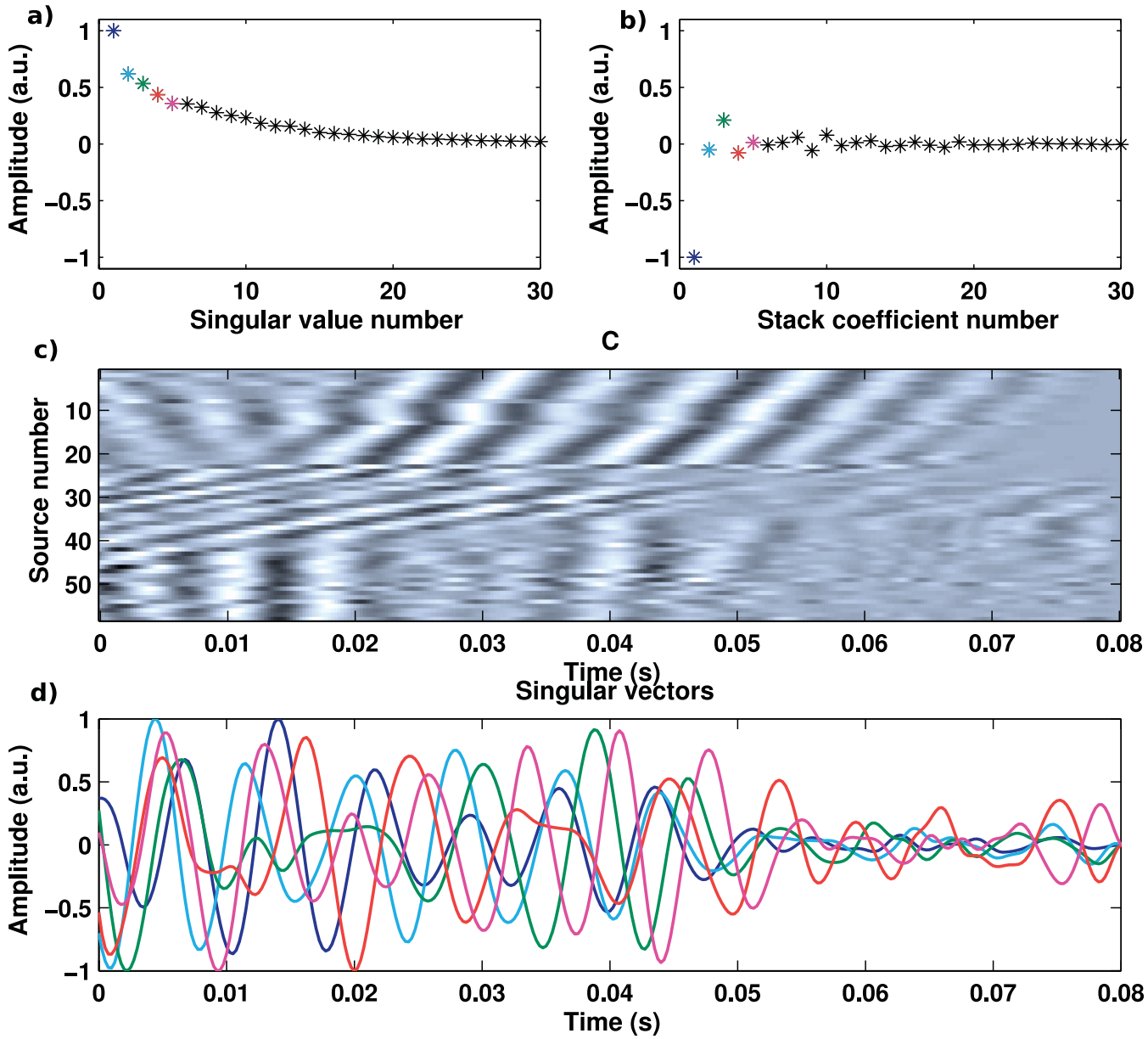

Figure 17. SVD decomposition components of cross-correlogram between receivers r_1 and r_{43}: (a) singular values, σ_k; (b) stack coefficients, s_k; (c) original cross-correlogram, $\boldsymbol{C}$; (d) first five singular vectors. For this particular receiver pair, there is one strong singular value and stack coefficient, which correspond to the first singular vector. Singular vectors 1–3 mostly correspond to energy up to 0.05 m, while singular vectors 4 and 5 contain relatively more energy after 0.05 s.

value of each of the SVD decomposition of the cross-correlograms. Looking at these three virtual shot gathers we see that the physical arrivals are enhanced in V_{n-1} compared to V and V_1.

In this particular field data set, the first singular value corresponds primarily to energy from the virtual refraction. Thus, contrary to the synthetic examples, the reflection phases were enhanced by ignoring the largest singular value. Even though the virtual refraction is not a physical arrival, it is an event with stationary energy in the cross-correlogram, so it is not considered to be noise as far as the SI-SVD method is concerned. In addition, as mentioned above, the virtual refraction can be used to invert for physical parameters. As previously mentioned, a rank-1 matrix obtained through SVD will consist of one row that best represents the original matrix, thus capturing what is most in common among all rows. As explained in Nichols *et al.* (2010), the critical offset for this data set is near the beginning of the source array, so most sources are at post-critical offsets. Therefore, the energy in the cross-correlogram corresponding to the virtual refraction is present in many sources (remember that each row in the cross-correlogram corresponds to one source), more precisely, all the sources from the critical offset until the end of the source array. Since the virtual refraction energy is the strongest signal present across most of the sources, it is captured by the first singular value. Other than having energy at most of the sources, the amplitude of the events in the cross-correlogram also plays a role. The field data was pre-processed in order to remove ground-roll and emphasize events related to the water table (refraction and reflection). Without this pre-processing, ground roll dominates the shot gathers and the virtual refraction, as well as other events corresponding to correlations between refracted and reflected waves, are much weaker. In this case the first singular value may not

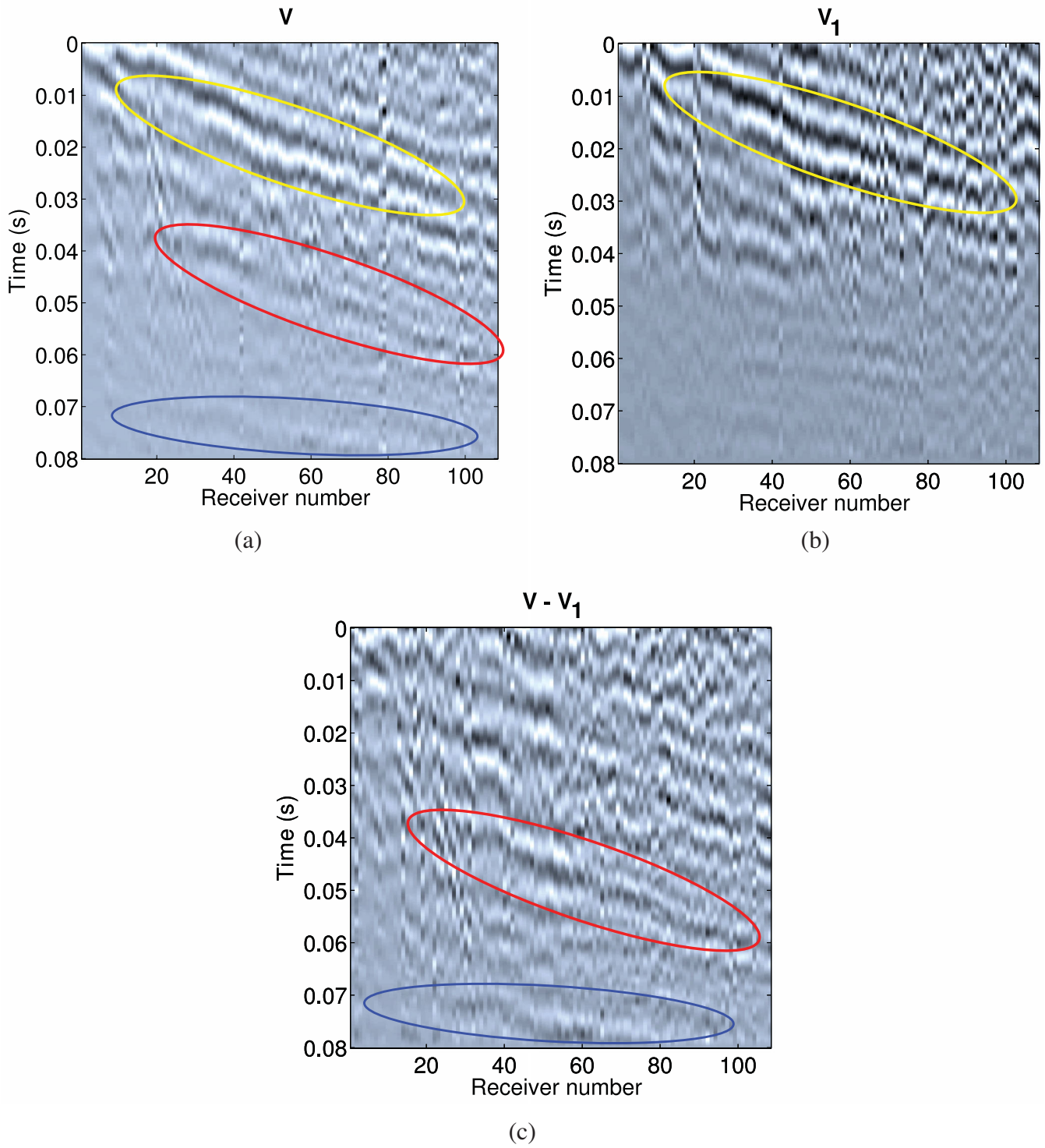

Figure 18. (a) Interferometric virtual shot record, V, for virtual source at r_1. Note the virtual refraction (yellow ellipse), real refraction (red ellipse), and reflection (blue ellipse). (b) Virtual shot gather obtained through rank-1 cross-correlogram, V_1. This virtual shot gather is dominated by energy from the virtual refraction. (c) Virtual shot gather, $V - V_1 = V_{n-1}$, obtained by ignoring largest singular value from the SVD decomposition of all cross-correlograms. The deeper events are now enhanced.

correspond primarily to the virtual refraction. A further quantitative study is necessary to determine in what situations which events are associated with which singular values. Meanwhile one can study this relationship by constructing virtual gathers with different singular values to see which events correspond to which singular values.

CONCLUSIONS

The accurate estimation of the GF with non-ideal source coverage remains a significant problem in SI. We have shown how using lower-rank cross-correlogram approximations, obtained through SVD, is a promising approach to alleviate this problem. The SVD

approach preserves stationary energy in the cross-correlogram, which is the energy that contributes most to GF recovery, and helps to attenuate the non-stationary energy that contributes primarily to artefacts in the interferometric GF. From the examples presented here, we see that different arrivals may correspond to different sets of singular values. This demonstrates that SVD is a powerful tool to filter events in the cross-correlogram, not only removing artefacts but also enhancing weaker arrivals. As a consequence, the lower-rank cross-correlograms obtained through SVD can lead to virtual shot gathers with clearer phases than standard virtual shot gathers, and may recover phases that are obscured by noise in the standard virtual gather.

ACKNOWLEDGEMENTS

We would like to thank Oleg Poliannikov, Pierre Gouedard, Bongani Mashele, Huajian Yao and Michael Fehler for the great inputs during our discussions. This work is supported by grants from the Department of Energy (DOE) and the founding members consortium at Earth Resources Laboratory (ERL).

REFERENCES

Aki, K. & Richards, P.G., 1980. *Quantitative Seismology: Theory and Methods,* W. H. Freeman, San Francisco.

Bakulin, A. & Calvert, R., 2006. The virtual source method: theory and case study, *Geophysics,* **71,** SI139–SI150.

Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science,* **299**(5606), 547–549.

Claerbout, J.F., 1968. Synthesis of a layered medium from its acoustic transmission response, *Geophysics,* **33,** 264–269.

Curtis, A., Gerstoft, P., Sato, H. & Snieder, R., 2006. Seismic interferometry—turning noise into signal, *Leading Edge,* **25,** 1082–1092.

de Hoop, A.T., 1988. Time-domain reciprocity theorems for acoustic wave fields in fluids with relaxation, *J. acoust. Soc. Am.,* **84**(5), 1877–1882.

Derode, A., Larose, E., Campillo, M. & Fink, M., 2003. How to estimate the Green's function of a heterogeneous medium between two passive sensors?, *Appl. Phys. Lett.,* **38,** 3054–3056.

Eckart, C. & Young, G., 1936. The approximation of one matrix by another of lower rank, *Psychometrika,* **1,** 211–218.

Fokkema, J. & van den Berg, P., 1993. *Seismic Applications of Acoustic Reciprocity,* Elsevier, Amsterdam, NY.

Freire, S. L.M. & Ulrych, T.J., 1988. Application of singular value decomposition to vertical seismic profiling, *Geophysics,* **53**(6), 778–785.

Godin, O.A., 2006. Recovering the acoustic Green's function from ambient noise cross crorrelation in an inhomogeneous moving medium, *Phys. Rev. Lett.,* **97,** 054301–054304.

Golub, G. & van Loan, C., 1996. *Matrix Computations,* The Johns Hopkins Univ. Press, Baltimore, MD.

Hansen, P.C., Kilmer, M.E. & Kjeldsen, R., 2006. Exploiting residual information in the parameter choice for discrete ill-posed problems, *BIT,* **46,** 41–59.

King, S. & Curtis, A., 2012. Suppressing nonphysical reflections in Green's function estimates using source-receiver interferometry, *Geophysics,* **77**(1), Q15–Q25.

Komatitsch, D. & Tromp, J., 2002. Spectral-element simulations of global seismic wave propagation–I. Validation, *Geophys. J. Int.,* **149,** 390–412.

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

Mehta, K., Bakulin, A., Sheiman, J., Calvert, R. & Snieder, R., 2007. Improving the virtual source method by wave-field separation, *Geophysics,* **72,** V79–V86.

Mikesell, D. & van Wijk, K., 2011. Seismic refraction interferometry with a semblance analysis on the crosscorrelation gather, *Geophysics,* **76**(5), SA77–SA82.

Mikesell, D., van Wijk, K., Calvert, A. & Haney, M., 2009a. The virtual refraction: useful spurious energy in seismic interferometry, *Geophysics,* **74**(3), A13–A17.

Mikesell, D., van Wijk, K., Colvert, A. & Haney, M., 2009b. Refraction interferometry for numerical surface seismic experiments, *SEG Tech. Prog. Expand. Abstr.,* **28**(1), 1350–1354.

Mikesell, T., van Wijk, K., Ruigrok, E., Lamb, A. & Blum, T., 2012. A modified delay-time method for statics estimation with the virtual refraction, *Geophysics,* **77**(6), A29–A33.

Nichols, J., Mikesell, T.D. & van Wijk, K., 2010. Application of the virtual refraction to near-surface characterization at the Boise Hydrogeophysical Research Site, *Geophys. Prospect.,* **58,** 1011–1022.

Poliannikov, O.V. & Willis, M.E., 2011. Interferometric correlogram-space analysis, *Geophysics,* **76**(1), SA9–SA17.

Rayleigh, J.W.S., 1896. *The Theory of Sound,* Vol. 2, Dover Publications Inc., New York, NY.

Roux, P., Sabra, K.G., Kuperman, W.A. & Roux, A., 2005. Ambient noise cross correlation in free space: theoretical approach, *J. acoust. Soc. Am.,* **117,** 79–84.

Sabra, K.G., Gerstoft, P., Roux, P., Kuperman, W.A. & Fehler, M., 2005a. Surface-wave tomography from microseisms in Southern California, *Geophys. Res. Lett.,* **32,** L14311, doi:10.1029/2005GL023155.

Sabra, K.G., Roux, P. & Kuperman, W.A., 2005b. Arrival-time structure of the time-averaged ambient noise cross-correlation function in an oceanic waveguide, *J acoust. Soc. Am.,* **117**(1), 164–174.

Schuster, T.G., Yu, J., Sheng, J. & Rickett, J., 2004. Interferometric/daylight seismic imaging, *Geophys. J. Int.,* **157,** 838–852.

Shapiro, N.M., Campillo, M., Stehly, L. & Ritzwoller, M.H., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science,* **307,** 1615–1618.

Snieder, R., 2004. Extracting the Green's function from correlation of coda waves: a derivation based on stationary phase, *Phys. Rev. E,* **69,** 046610.

Snieder, R., 2007. Extracting the Green's function of attenuating heterogeneous acoustic media from uncorrelated waves, *J. acoust. Soc.,* **121,** 2637–2643.

Snieder, R., Wapenaar, K. & Larner, K., 2006. Spurious multiples in seismic interferometry of primaries, *Geophysics,* **71**(4), SI111–SI124.

Snieder, R., Wapenaar, K. & Wegler, U., 2007. Unified Green's function retrieval by cross-correlation; connection with energy principles, *Phys. Rev. E,* **75,** 036103.

Stehly, L., Campilo, M. & Shapiro, N.M., 2006. A study of seismic noise from its long-range correlation properties, *J. geophys. Res.,* **111,** B10306.

Ulrych, T.J., Sacchi, M.D. & Graul, J.M., 1999. Signal and noise separation: art and science, *Geophysics,* **64**(5), 1648–1656.

van Wijk, K., 2006. On estimating the impulse response between receivers in a controlled ultrasonic experiment, *Geophysics,* **71,** SI79–SI84.

van Wijk, K., Mikesell, T.D., Schulte-Pelkum, V. & Stachnik, J., 2011. Estimating the Rayleigh-wave impulse response between seismic stations with the cross terms of the Green's tensor, *Geophys. Res. Lett.,* **38,** L16301, doi:10.1029/2011GL047442.

Wapenaar, K., 2006. Green's function retrieval by cross-correlation in case of one-sided illumination, *Geophys. Res. Lett.,* **33,** L19304, doi:10.1029/2006GL027747.

Wapenaar, K. & Fokkema, J., 2006. Green's function representations for seismic interferometry, *Geophysics,* **71**(4), S133–S146.

Wapenaar, K., Draganov, D., van der Neut, J. & Thorbecke, J., 2004a. Seismic interferometry: a comparison of approaches, *SEG Tech. Prog. Expand. Abstr.,* **23,** 1981–1984.

Wapenaar, K., Thorbecke, J. & Dragonov, D., 2004b. Relations between reflection and transmission responses of three-dimensional inhomogeneous media, *Geophys. J. Int.,* **156,** 179–194.

Wapenaar, K., Slob, E. & Snieder, R., 2006. Unified Green's function retrieval by cross correlation, *Phys. Rev. E.,* **97,** 234301.

Weaver, R.L., 2005. Information from seismic noise, *Science,* **307,** 1568–1569.

Weaver, R.L. & Lobkis, O.I., 2004. Diffuse fields in open systems and the emergence of the Green's function, *J. acoust. Soc. Am.,* **116,** 2731–2734.