GEOPHYSICS, VOL. 80, NO. 6 (NOVEMBER-DECEMBER 2015); P. V189–V199, 8 FIGS.
10.1190/GEO2015-0287.1

Closed-loop surface-related multiple elimination and its application to simultaneous data reconstruction

Gabriel A. Lopez[1] and D. J. Verschuur[1]

ABSTRACT

Surface-related multiple elimination (SRME) is one of the most commonly used methods for suppressing surface multiples. However, to obtain an accurate surface multiple estimation, dense source and receiver sampling is required. The traditional approach to this problem is performing data interpolation prior to multiple estimation. Although appropriate in many cases, this methodology fails when big data gaps are present or when relevant information is not recovered, e.g., near-offset data in shallow-water environments. We have developed a solution in which multiple estimation was performed simultaneously with data reconstruction, such that data reconstruction helped obtain better multiple estimates and in which the physical primary-multiple relationship helped constrain the data interpolation. To accomplish this, we proposed to extend the recently introduced closed-loop SRME (CL-SRME) algorithm to account for primary estimation in the case of coarsely sampled data. We achieved this by introducing a focal-domain parameterization of the primaries in a sparsity-promoting CL-SRME method. Results proved that the method was capable of reliably estimating primaries data in case of shallow water and with large undersampling factors.

INTRODUCTION

With the introduction of surface-related multiple elimination (SRME) (Berkhout, 1982; Verschuur et al., 1992; Berkhout and Verschuur, 1997; Weglein et al., 1997; Biersteker, 2001), a complete new approach to multiple removal was developed — The multiples could be predicted without any prior knowledge of the subsurface. Despite its major success over the years, there are still some limitations to the SRME approach. First, adaptive subtraction is usually based on minimum energy, which is not always a good assumption (Nekut and Verschuur, 1998). Second, it needs the reconstruction of missing near offsets because the data are used as a multiple predictor operator (Dragoset and Jeričević, 1998; Dragoset et al., 2010). Third, it requires dense source and receiver sampling, which often poses problems in the 3D or shallow-water-layer applications (Hargreaves, 2006; Moore and Bisley, 2006).

Recently, another approach to multiple removal was developed by van Groenestijn and Verschuur (2009a, 2009b) — estimation of primaries by sparse inversion (EPSI). The main difference with SRME is that the two-stage processing method, being prediction and adaptive subtraction, is replaced by a full-waveform inversion process. The primary reflection events are the unknowns in this algorithm. Although successful in many scenarios (Savels et al., 2011), the EPSI algorithm also brought some limitations. Due to the parameterization of the primary responses with spikes and wavelets, the algorithm showed difficulties in handling complexities in the subsurface such as wave dispersion, making refinements to the original algorithm a necessity (Baardman et al., 2010; Lin and Herrmann, 2011; Jumah and Herrmann, 2014). The challenges for EPSI were even larger in 3D, in which the coarse sampling required a method for combined multiple elimination and data interpolation. Due to the large number of iterations necessary for convergence and the artifacts observed when trying to extend EPSI to new parameterizations (due to the sparsity and causality constrains), the necessity of a new algorithm became evident.

To overcome the above-mentioned limitations, the closed-loop SRME method (CL-SRME) was proposed (Lopez and Verschuur, 2014). This novel method combines the robustness of SRME with the inversion approach of EPSI, to produce an inversion-based multiple removal method, in which the primaries are estimated directly via full-waveform inversion such that they, together with their multiples, explain the input data. The process can be efficiently carried out as an unconstrained least-squares problem, without the need of any precise subsurface information.

To use CL-SRME to accurately predict multiples in a coarsely sampled data set (e.g., for 3D application), a new parameterization

Manuscript received by the Editor 20 May 2015; published online 25 September 2015.
[1]Delft University of Technology, Lorentzweg, Delft, Netherlands. E-mail: g.a.lopezangarita@tudelft.nl; d.j.verschuur@tudelft.nl.

must be adopted to allow CL-SRME to efficiently reconstruct data over large gaps. In our new approach, we will use the focal transform (Berkhout and Verschuur, 2006; Kutscha et al., 2010; Kutscha and Verschuur, 2012), which aims at focusing primary reflections into localized events. The focal domain can compress the primary energy, a property that will be useful when separating primary signals from undersampling noise in the focal domain. A sparse norm regularization is used to remove the undersampling noise from the estimations and to eliminate the redundancy in the parameterization. Note that the multilevel focal transform is a redundant parameterization because every focal domain is as big as the input data. This implies that in a N-level focal transform, the number of inversion parameters is actually N times larger than the number of measurement points in the input data set. This redundancy can be removed by applying an extra constraint.

Other parameterizations using the curvelet transform (Candes et al., 2006; Herrmann et al., 2007, 2008; Herrmann and Verschuur, 2008) or the wavelet transform (Liu et al., 1995) are also possible candidates for a sparse representation of the seismic data. Our choice of focal transform is justified by the fact that, due to its focusing characteristics, the focal transform is able to compress the energy of highly curved events into localized events, making it ideal for shallow-water-layer applications, in which the events to reconstruct are strongly curved in the near offsets. For the deeper data and flatter events, other parameterizations such as curvelets or linear Radon might become more beneficial.

In the following sections, we will provide some theoretical explanations along with some practical 2D examples to explain the CL-SRME algorithm including the focal-domain parameterization. Special attention will be given to near-offset reconstruction in shallow-water environments.

CLOSED-LOOP SURFACE-RELATED MULTIPLE ELIMINATION

Following Lopez and Verschuur (2014), let $\mathbf{P}$ describe the upgoing wavefield at the surface, in which each column of the matrix contains the measurements for one (2D or 3D) shot record for a single frequency component. In this detail-hiding notation (Berkhout, 1982; Kinneging et al., 1989), we can write

$$\mathbf{P} = \mathbf{P}_0(\mathbf{I} + \mathbf{AP}), \tag{1}$$

where $\mathbf{P}_0$ represents the primary wavefield, and $\mathbf{A} = \mathbf{S}^{-1}\mathbf{R}^{\cap}$ represents the surface operator (Verschuur et al., 1992), which is dependent on the source signatures from all sources $\mathbf{S}$ and the reflectivity at the surface from below $\mathbf{R}^{\cap}$. In this expression, the total upgoing wavefield is represented as a product of a downgoing wavefield $(\mathbf{I} + \mathbf{AP})$ and an upgoing wavefield $\mathbf{P}_0$, in which the total wavefield is described as a sum of the primary wavefield $\mathbf{P}_0$ and the surface-related multiple wavefield $\mathbf{P}_0\mathbf{AP}$ (Berkhout and Verschuur, 1997). In this expression, the matrix products represent multidimensional convolutions in the space-frequency domain. By "primaries," we actually mean all possible events without reflection points at the surface. This category includes primaries, internal multiples, and all other upgoing events that are not related to the scattering at the surface (e.g., converted waves). Note that for practical applications, the directivity in $\mathbf{A}$ is often neglected and all sources are assumed to carry the same wavelet (Verschuur and Berkhout, 1997). Then, $\mathbf{A}$ can be written as $\mathbf{A} = A(\omega)\mathbf{I}$.

In traditional iterative SRME (Berkhout and Verschuur, 1997), the primaries are estimated by defining them as

$$\mathbf{P}_0 = \mathbf{P} - \mathbf{AP}_0\mathbf{P}, \tag{2}$$

where $\mathbf{A} = A(\omega)\mathbf{I}$ is estimated by minimizing the energy in the estimated primaries $\mathbf{P}_0$ and by parameterizing the inverse source wavelet $A(\omega)$ as a short filter in the time domain (Verschuur and Berkhout, 1997), with known problems for the minimum energy criterion (Nekut and Verschuur, 1998; Glutton and Verschuur, 2004).

CL-SRME proposes an inversion-based method to obtain the primary impulse response $\mathbf{P}_0$ from the measured data $\mathbf{P}$ in a fully data-driven manner. The strategy is to build an objective function $J = J(\hat{\mathbf{P}}_0, \hat{\mathbf{A}})$ that is minimized when the parameter estimates $\hat{\mathbf{P}}_0$ and $\hat{\mathbf{A}}$ accurately describe the input data set $\mathbf{P}$. Here, the symbols $\hat{\mathbf{P}}_0$ and $\hat{\mathbf{A}}$ represent the estimates of the true parameters $\mathbf{P}_0$ and $\mathbf{A}$. A sparsity constraint L_S on the primaries can be added to the process to ensure the minimum amount of multiple leakage in the estimates. Note that this constraint must be applied in the time domain. Using the Frobenius L2 norm, we can express J as

$$J = \sum_{\omega} \|\mathbf{P} - \hat{\mathbf{P}}_0(\mathbf{I} + \hat{\mathbf{A}}\mathbf{P})\|^2 + \lambda L_S(\hat{\mathbf{P}}_0), \tag{3}$$

where each data matrix contains one frequency component. Using J in a least-squares iterative inversion procedure, CL-SRME aims to estimate the primary wavefield $\mathbf{P}_0$ by a large-scale inversion in which the inversion parameters $\hat{\mathbf{A}}$ and $\hat{\mathbf{P}}_0$ are estimated such that the input data $\mathbf{P}$ is explained. In this way, the multiples help to estimate the primaries. Note that CL-SRME given by equation 3 will describe exactly the same problem as traditional SRME under the condition $L_S = L_2$, where the L_2 term represents the minimum-energy condition over the primaries. In that case, both algorithms tend to converge to the same solution.

Note the difference in parameterization with respect to EPSI. In CL-SRME, the input data are parameterized via the surface operator $\mathbf{A}$ and the primary wavefield $\mathbf{P}_0$, which is consistent with the parameterization in SRME (Verschuur et al., 1992). EPSI on the other hand describes the input data in terms of the source signature $\mathbf{S}$ and the primary impulse response $\mathbf{X}_0$. Mathematically, these two parameterizations are equivalent, but practically they involve very different wavefields, and they require different inversion strategies. The relationship between the inversion parameters $\mathbf{X}_0$ and $\mathbf{S}$ of EPSI and $\mathbf{P}_0$ and $\mathbf{A}$ of CL-SRME is given by $\mathbf{P}_0 = \mathbf{X}_0\mathbf{S}$ and $\mathbf{A} = \mathbf{S}^{-1}\mathbf{R}^{\cap}$.

CLOSED-LOOP SURFACE-RELATED MULTIPLE ELIMINATION INCLUDING DATA RECONSTRUCTION

To use CL-SRME to remove multiples from data sets with severe undersampling (which is typical for 3D geometries), we will follow the parameterization similar to the one described by Lopez and Verschuur (2013) such that data reconstruction and multiple removal are achieved. Our goal is to use CL-SRME to remove the multiples and to reconstruct big portions of missing data. To create a proper reconstruction algorithm, the following conditions must be met:

1) The primary wavefield must be represented in a transform domain with the smallest possible number of nonzero transform parameters, such that their information is compressed and effi-

ciently represented. Signals arising from multiple reflections do not need to be represented in this way because they simply follow any parameterization of the primaries $\mathbf{P}_0$ via $\mathbf{P}_0\mathbf{AP}$. Typically, the data gaps are mapped to artifacts in the transform domain. Due to their lack of coherency, these artifacts are often poorly compressed and are usually seen as aliasing noise in the transform domain.

2) A parameter selection method in the transform domain must be implemented to separate the parameters representing the primary signals from the parameters accounting for the aliasing noise. Depending on the geometry of the transform domain, these methods might vary, in which typical examples might be picking, thresholding, regularization, etc.

To fulfill the first condition, we propose the use of the focal transform (Berkhout and Verschuur, 2006; Kutscha et al., 2010; Kutscha and Verschuur, 2012; Lopez and Verschuur, 2013) applied to the primary data. In this description, we will make use of some rough normal moveout (NMO)-velocity information to create propagation operators $\mathbf{W}$ that will allow us to back-propagate the wavefields. If we select these operators to take the reflected energy back to where it came from (i.e., a strongly reflective layer), then the scattered energy will be focused in one small event in time. Focusing brings strong compression and coherency in the data, in such a way that it will easily allow reconstruction over large data gaps. The reasoning behind this idea is schematically depicted in Figure 1. As we can see from the figure, focusing generates compression of the events in the transform domain. The focal transformation can also be seen as a redatuming operation to a selected layer. As observed in Figure 1, the focal transform will concentrate the energy that originates from the level of interest. However, further away from this level, the compression decreases. Therefore, in Kutscha et al (2012), a simultaneous, multilevel focal transform is proposed in which many strategic levels are chosen for the focusing.

For the multilevel focal transform, we consider the M strongest reflectors in the data and extract their approximate root-mean-square (rms) velocities. Then, we construct a set of propagation operators from the surface to the selected reflectors. These operators have no need to be exact because they are expected to provide only an approximation of the traveltime information in our data. The set of operators for the whole seismic experiment is written as $\{\mathbf{W}_m\}_{m\in[1,M]}$ and is expected to extrapolate the data from the surface to a set of depth levels $\{z_m\}_{m\in[1,M]}$. These $\mathbf{W}_m$ will be later used to parameterize the primary estimates $\hat{\mathbf{P}}_0$ into a set of focal domains $\hat{\mathbf{X}}_m$ such that

$$\hat{\mathbf{P}}_0 = \sum_{m=1}^{M} \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m. \tag{4}$$

Note that each focal domain $\hat{\mathbf{X}}_m$ represents a full data volume (sources and receivers) for one frequency component. Thus, the focal transform is redundant. The relationship between the $\hat{\mathbf{X}}_m$ term of equation 4 and the $\hat{\mathbf{X}}_0$ in EPSI is given by $\hat{\mathbf{X}}_0 S = \sum_{m=1}^{M} \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m$. Hence, actually, $\hat{\mathbf{X}}_0$ can be interpreted as a special case of equation 4 in which $M = 1$, the source wavelet is a Dirac delta function and the propagation operators are chosen to be matrix identities (target propagation level at the surface).

Despite their similar notation, the $\hat{\mathbf{X}}_m$ term of equation 4 is in principle not directly related with the $\hat{\mathbf{X}}_0$ in EPSI (which represents the primary impulse response).

To fulfill the second condition above, we require a method to eliminate the aliasing noise from the transform domains $\hat{\mathbf{X}}_m$. In addition, with such a condition, the redundancy in the transform would be removed (e.g., nonsparse representations of events will be removed/attenuated to favor the sparser, and typically more unique, representations). We will use a sparsity-promoting regularization norm $\|\cdot\|_S$ applied to the focal domains $\hat{\mathbf{X}}_m$ to do this. With this extra constraint, we drive the algorithm toward a sparse representation of the focal domains. This condition will remove the energy from the aliasing artifacts to concentrate it in the main $\hat{\mathbf{X}}_m$ primary events, thus obtaining a fully sampled representation of $\hat{\mathbf{P}}_0$ via equation 4.

Using equations 3 and 4, the current algorithm can be described by minimizing the following objective function:

$$\begin{aligned} J &= \sum_{\omega} J_{\omega}^{(\mathrm{LS})} + \sum_{t} J_{t}^{(\mathrm{reg})} \\ &= \sum_{\omega} \left\| \mathbf{P} - \sum_{m=1}^{M} \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q} \right\|^2 + \lambda \sum_{t} \sum_{m} \|\hat{\mathbf{x}}_m\|_S, \end{aligned} \tag{5}$$

with

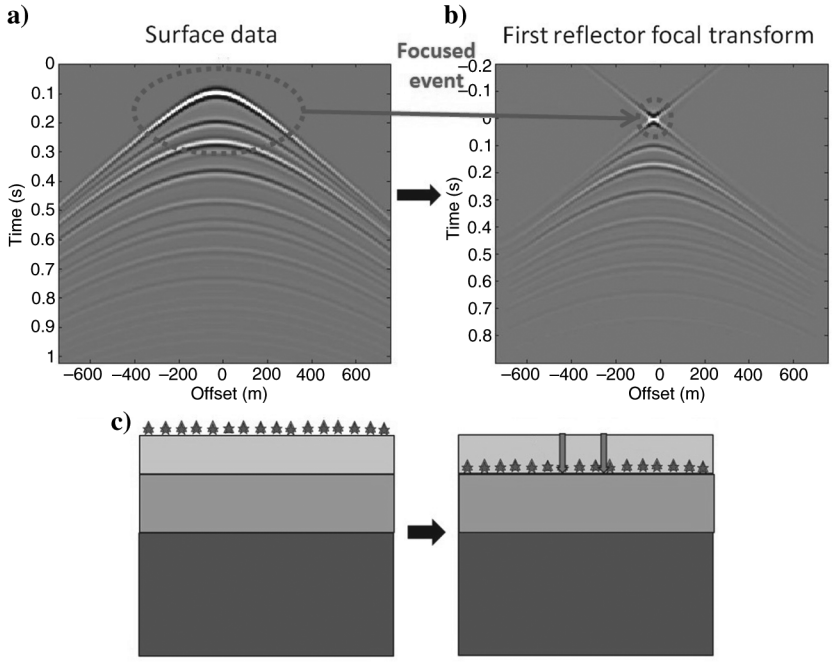

Figure 1. (a) The data acquired at the surface are focal transformed via the application of the propagation operators related to the first layer. (b) In the transformed data set, the primary event related with the first layer appears concentrated at $t = 0$. The focal transform has mapped the information of the first reflection into a compressed event in the transformed domain. (c) Application of the focal transform related to one particular layer is equivalent to redatuming the data set to that layer.

$$\mathbf{Q} = \mathbf{I} + \hat{\mathbf{A}}\mathbf{P}, \tag{6}$$

where $\hat{\mathbf{x}}_m$ is the inverse Fourier transform of the data cube of $\hat{\mathbf{X}}_m$, t representing a time slice, and λ is a user-defined regularization constant (typically $\lambda \approx 10^{-2}$). The norm $\|\cdot\|_S$ represents any sparsity-promoting norm of preference (e.g., L1), which applies to every time slice in $\hat{\mathbf{x}}_m$. Note that the regularization norm is applied in the time domain, so sparsity is enforced in that domain.

Missing data reconstruction is implicitly present in equation 6 by setting

$$\mathbf{P} = \mathbf{P}' + \mathbf{P}'', \tag{7}$$

where $\mathbf{P}$ are the total data (fully sampled), $\mathbf{P}'$ are the input data (undersampled), and $\mathbf{P}''$ are the data to reconstruct. Equations 5–7 represent the forward model for the unknown parameters $(\hat{\mathbf{X}}_m, \hat{\mathbf{A}})$, such that $J = J(\hat{\mathbf{X}}_m, \hat{\mathbf{A}})$. Using gradient-based inversion, we can find a set of model parameters $(\hat{\mathbf{X}}_m, \hat{\mathbf{A}})$ that minimizes our objective function J.

The inversion updates for the focal domains $\hat{\mathbf{X}}_m$ are given by $\Delta\hat{\mathbf{X}}_m^{(\mathrm{LS})} = -\nabla_{\hat{\mathbf{X}}_m} J_\omega^{(\mathrm{LS})}$ and $\Delta\hat{\mathbf{x}}_m^{(\mathrm{reg})} = -\nabla_{\hat{\mathbf{x}}_m} J_t^{(\mathrm{reg})}$, which leads to (outline of the proof in Appendix A)

$$\Delta\mathbf{X}_m = \Delta\mathbf{X}_m^{(\mathrm{LS})} + \Delta\mathbf{X}_m^{(\mathrm{reg})} \tag{8}$$

and

$$\begin{aligned} \Delta\mathbf{X}_m^{(\mathrm{LS})} &= 2\mathbf{W}_m^*(\mathbf{P} - \hat{\mathbf{P}}_0\mathbf{Q})\mathbf{Q}^H\mathbf{W}_m^H \quad \text{and} \\ \Delta\mathbf{x}_m^{(\mathrm{reg})} &= -\lambda\nabla_{\hat{\mathbf{x}}_m}\|\hat{\mathbf{x}}_m\|_S, \end{aligned} \tag{9}$$

where $\Delta\mathbf{X}_m$ is the total m-focal-domain update, $\Delta\mathbf{X}_m^{(\mathrm{LS})}$ is the portion of $\Delta\mathbf{X}_m$ associated with the least-squares part of J, and $\Delta\mathbf{X}_m^{(\mathrm{reg})}$ (the Fourier transform of the data cube of $\Delta\hat{\mathbf{x}}_m^{(\mathrm{reg})}$) is the portion of $\Delta\mathbf{X}_m$ associated with the regularization part of J. The superscripts $*$ and H refer to a complex conjugation and an adjoint operation, respectively. Note that in equations 8 and 9, $\Delta\mathbf{X}_m^{(\mathrm{LS})}$ and $\Delta\mathbf{X}_m^{(\mathrm{reg})}$ are related to $\Delta\mathbf{X}_m$ via an overall scaling factor λ, which controls the strength of the sparsity regularization.

The updates $\Delta\mathbf{X}_m$ are used to renew the estimate of $\mathbf{X}_m$ in every iteration i via the recursion formula:

$$\hat{\mathbf{X}}_m^{(i+1)} = \hat{\mathbf{X}}_m^{(i)} + \alpha\Delta\mathbf{X}_m^{(i)}, \tag{10}$$

where the scaling parameter α is chosen by a line search such that $J(\hat{\mathbf{X}}_m^{(i+1)}) = J(\hat{\mathbf{X}}_m^{(i)} + \alpha\Delta\mathbf{X}_m^{(i)})$ is minimized. The remaining parameters $\mathbf{P}$ and $\hat{\mathbf{A}}$ are strongly dependent on the estimate of $\hat{\mathbf{P}}_0$, so they can be calculated in every iteration directly from the $\hat{\mathbf{P}}_0$ estimate by

$$\hat{A} = \begin{cases} \hat{A}\ \text{s.t.}\ \|\mathbf{P} - \hat{\mathbf{P}}_0\hat{A}\mathbf{P}\|^2 \to \min & \text{in the early iterations} \\ \hat{A}\ \text{s.t.}\ \|\mathbf{P} - \hat{\mathbf{P}}_0 - \hat{\mathbf{P}}_0\hat{A}\mathbf{P}\|^2 \to \min & \text{in the rest of the iterations} \end{cases} \tag{11}$$

and

$$\mathbf{P}'' = \left(\sum_{m=1}^{M} \mathbf{W}_m^T\hat{\mathbf{X}}_m\mathbf{W}_m\mathbf{Q}\right)'', \tag{12}$$

where the double prime on the right side means selecting only the data points corresponding to $\mathbf{P}''$.

Equation 11 uses the minimum primary energy constraint (Verschuur and Berkhout, 1997) in the early iterations to estimate A (assuming $\hat{\mathbf{A}} = \hat{A}\mathbf{I}$). This is necessary to avoid the trivial solution given by $\hat{A} = 0$ and $\mathbf{P}_0 = \mathbf{P}$. It is also necessary to maintain the minimum energy constraint until the sparsity starts playing a relevant role in the updates. Typically, the constraint can be dropped after $\sqrt{J}$ has decayed to the 20% of its value at the initial guess. The matching of $\hat{A}$ is done in practice via a least-squares adaptive filter. The filter, which is restricted to be short in time, is found such that the amplitudes of $\mathbf{P}$ (or $\mathbf{P} - \mathbf{P}_0$) and $\hat{\mathbf{P}}_0\hat{A}\mathbf{P}$ are matched, driving their difference to a minimum (Verschuur and Berkhout, 1997).

Equation 12 uses the prediction equation 1 together with equation 7 to produce estimates of the missing data $\mathbf{P}''$, given the knowledge of the fully sampled primary wavefield $\hat{\mathbf{P}}_0$. Note that the relationship between $\mathbf{P}$ and $\mathbf{P}_0$ (equation 1) implies that the knowledge of $\mathbf{P}_0$ and $\mathbf{A}$ fully determines $\mathbf{P}$. This means that if we are able to reconstruct the missing data in $\hat{\mathbf{P}}_0$ (which we do via a sparsity constraint on the focal domain), then we can automatically reconstruct $\mathbf{P}$ via equation 12. In fact, it can be shown that $\mathbf{P}$ can be written fully in terms of $\mathbf{P}_0$ and $\mathbf{A}$ via a scattering series (Verschuur et al., 1992). Note that equation 12 refers only to the missing data positions in the input data set, such that the original data $\mathbf{P}'$ is kept intact.

Our initial guess for $\hat{\mathbf{P}}_0^{(0)}$, $\hat{\mathbf{A}}^{(0)}$, and $\mathbf{P}''^{(0)}$, together with equations 8–12, constitutes the essence of the proposed joint multiple prediction and interpolation method.

To obtain a first interferometric estimate of missing data (specially in the near offsets), it is important to start the inversion with an initial $\hat{\mathbf{A}}$ value such that the first $\mathbf{P}_0$ estimate contains dominant back-projections of the multiples into the primaries. In the near-offset regions, the back-projection mechanism of equation 9 is the main mechanism driving the solutions. We will assume $\hat{\mathbf{P}}_0^{(0)} = 0$, $\hat{\mathbf{A}}^{(0)} = A_{\mathrm{initial}}$, and $\mathbf{P}^{(0)} = 0$ at the beginning of the algorithm. The A_{initial} term is calculated using equation 11 and $\hat{\mathbf{P}}_0 = \mathbf{P}$.

RESULTS

We have designed two experiments to test the undersampling and the near-offset interpolation capabilities of the present algorithm in the 2D case.

Near-offset interpolation

To account for the near-offset reconstruction capabilities of the present algorithm, a synthetic 2D model with a shallow seafloor of 50 m (Figure 2) is used for the generation of the synthetic data via a finite-differences algorithm. In this example, we entirely remove the information of the shallow reflector, eliminating the information in all offsets. In practice, this is equivalent to not measuring the first reflector primary at all. Traditional methods based on NMO or Radon reconstruction are impossible to apply in these circumstances. The CL-SRME algorithm is used for primary estimation and near-offset reconstruction of this data set, and the results are shown in Figure 3.

Figure 3a depicts the input data set, and Figure 3b and 3c depicts the reconstructed data set and the estimated primaries. As we can

see, the shallow event has been reconstructed almost completely, and its associated multiples are largely removed from the estimations. Some multiple leakages are still observed in the estimates, but these effects are small if we consider the magnitude of the interpolation. Note, in particular, that the multiples in the deeper regions of the data are largely removed. Only one focal domain was used here for inversion (centered in the first reflector and based uniquely on NMO parameters). The ability of mapping the multiples into the primaries inherent in CL-SRME, together with the sparsity-promoting regularization in a compressed domain, is the keys for large-gap near-offset reconstruction in this example.

To better account for the differences in primary estimation of SRME (with initial interpolation) and CL-SRME, a new experiment is done for many models with several seafloor depths and near-offset gaps. However, this time, we will keep part of the input-offset range and we will interpolate only up to 0.32 s, such that we focus our interpolation efforts in the shallow part of the data. In each experiment, the difference between the true primaries and the estimated primaries is computed and the relative energy error E is calculated via

$$E = \frac{\sum \left(P_0^{(\text{true})} - P_0^{(\text{estimated})}\right)^2}{\sum \left(P_0^{(\text{true})}\right)^2} \times 100\%. \tag{13}$$

The results are shown in Figure 4. As expected, errors in both algorithms tend to become larger when the water bottom becomes shallower. The CL-SRME tends to produce always better primary estimations, with fewer residuals in all depth ranges. The difference between the primary estimates of the CL-SRME algorithm and the traditional SRME is larger for shallower models (where more interpolation power is needed) and becomes less important for deeper models (where the initial interpolation is more accurate). Overall, the CL-SRME algorithm seems to better handle shallow data. Small differences in parameter settings (of both algorithms) generate some of the variations observed in the shallower depths in Figure 4.

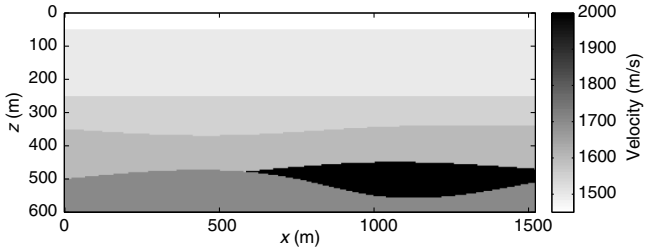

Figure 2. Velocity model used for the synthetic data generation in synthetic example. It has a shallow-water layer of 50-m depth.

Coarsely sampled data

In this section, we will show an example of the CL-SRME algorithm applied to a North Sea data set (courtesy of PGS). The data are taken from a 2D line from which a subset is selected with 201 receivers and 25 sources, with a 25-m receiver spacing. By applying reciprocity, a split-spread data set is obtained. Note that the same data set has also been used by Baardman et al. (2010). In this data

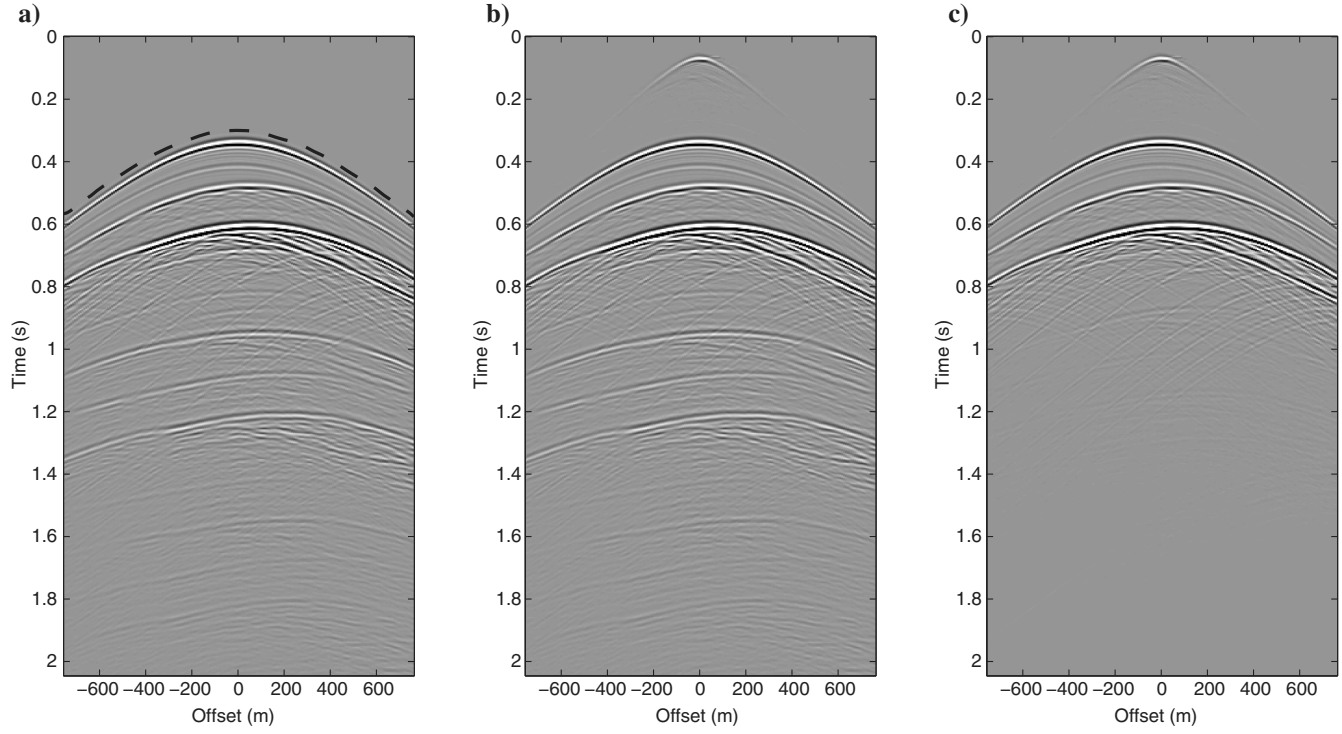

Figure 3. The CL-SRME in a shallow-water environment (50-m water depth) with complete near-offset reconstruction. (The entire offset range is reconstructed.) (a) Input data set with the first primary completely muted (muting done above the dashed line), (b) reconstructed data set, and (c) estimated primaries. Full near-offset reconstruction is achieved in the shallow region, with the multiples in the deep regions largely removed. Thirty iterations are needed for this example.

set, we have chosen sources to have a 1:5 irregular undersampling ratio in the sources. Then, a coarse NMO-based shot interpolation is applied to the data as a preprocessing step to reduce the aliasing artifacts in the focal domain. The initial interpolation does not have to be accurate because the interpolation errors will be mapped to the aliasing artifacts in the focal domain, which will be removed later on by the sparsity constraint. In this example, we will use three focal operators that are chosen to focus the energy of the three most prominent reflectors in the data. The operators are calculated using the stacking velocities of the reflectors of interest in the data, and we assume a homogeneous propagation medium for each operator. The NMO times and velocities used here are $t_{\mathrm{NMO}} = 0.16, 0.52, 0.96$ s and $v_{\mathrm{NMO}} = 1500, 1600, 1800$ m/s. Besides the stacking velocities and the apex times of the events of interest, no other model information is used to generate the operators.

Figure 5 shows the results of applying the CL-SRME algorithm to this data set. Because the interpolation is done along the source coordinate, results are presented in the common-receiver gather domain. The input data are shown in Figure 5a. The shot-interpolated data, primaries, and multiples (at the first iteration) are shown in Figure 5b–5d. The shot-interpolated data, primaries, and multiples (at the 30th iteration) are shown in Figure 5e–5g.

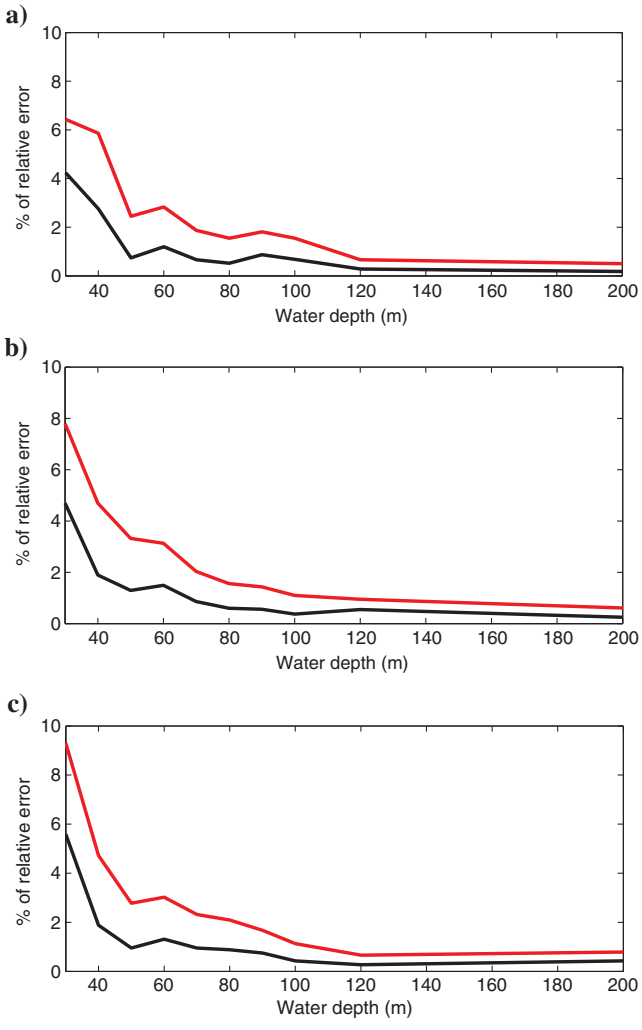

Figure 4. Relative energy error of SRME with interpolation (red) and CL-SRME (black) for different seafloor depths. With (a) a 64-m near-offset gap, (b) an 80-m near-offset gap, and (c) a 96-m near-offset gap.

As we can see from Figure 5, full data reconstruction has been achieved. By comparing the input (Figure 5a) and the output (Figure 5e) data sets, we can see the effect of the interpolation in filling up the missing data. By comparing the first and last iterations in the data, primaries, and multiples, we can see how the amplitude errors get healed, together with some phase errors. By comparing Figure 5c and 5f, we can also see the multiple energy eliminated from the primary recordings. Note that we have achieved a reasonably accurate primary-multiple separation, even in the case of a very coarsely sampled input data set. By comparing the input data set (Figure 5a) with the output primaries (Figure 5f), we can note the appearance of two weak primaries in the deep part of the recordings (between 1.4 and 1.8 s), which are hardly visible in the input data. However, due to the huge interpolation effort in this example, some minor aliasing/phase artifacts are still visible in the output data, especially in the top part.

Figure 6 shows the associated primary focal domains, used for obtaining the results in Figure 5. In this example, three operators are chosen to focus the main reflectors, thus compressing the largest amount of energy possible. The associated energy of these events maps at zero time, whereas the other events map to either negative or positive times. The following remarks can be deduced from Figure 6: (1) The initial redundancy in the focal domains has largely reduced. This is important for the uniqueness of the parameterization because initially one single focal domain contains the information of the entire data set. The former can be seen by events disappearing from one focal domain because they are more effectively represented in another focal domain. (2) The aliasing artifacts produced by the undersampling have been largely removed. The interpolation errors (due to unphysical data) in the initial primaries (Figure 5c) might be difficult to separate from the data in the original domain. However, they become easy to separate in the focal domain (e.g., Figure 6b) because they appear as aliasing artifacts.

The k_x-ω domain plots of the input and reconstructed data are presented in Figure 7. As we can see, the undersampling in the x-t domain is mapped to blending noise in the k_x-ω domain. After inversion, the full reconstruction of the k_x-ω response is obtained with no visible presence of any remaining aliasing artifacts.

DISCUSSION

To allow efficient reconstruction, the chosen parameterization must represent the wavefields with the smallest amount of nonzero parameters possible. Reconstruction is typically achieved after a thresholding/regularization process in which the undersampling artifacts are taken out of the estimates, leaving only the meaningful parameters to describe the data. As a result, the data then become fully sampled after the inverse transformation. Note that this type of reconstruction can be used in deep and shallow data reconstruction. The focal transform is especially suited for shallow reflectors because typically these events are strongly curved and can be efficiently represented in the focal domain. Other parameterizations such as curvelets (Lin and Herrmann, 2009) and Fourier (Zwartjes and Gisolf, 2006) or Radon (Hampson, 1986) transforms are also possible, but they provide poorer data com-

pression when applied to highly curved events (e.g., in situations with shallow-water data).

As discussed above, the focal operators do not need to be very accurate to provide a good description of the data. Figure 8 illustrates this by showing the reconstruction error in a typical example as a function of the velocity associated with the focal operator.

In terms of computation cost, the CL-SRME algorithm can be described as relatively fast, typically with 20 iterations needed

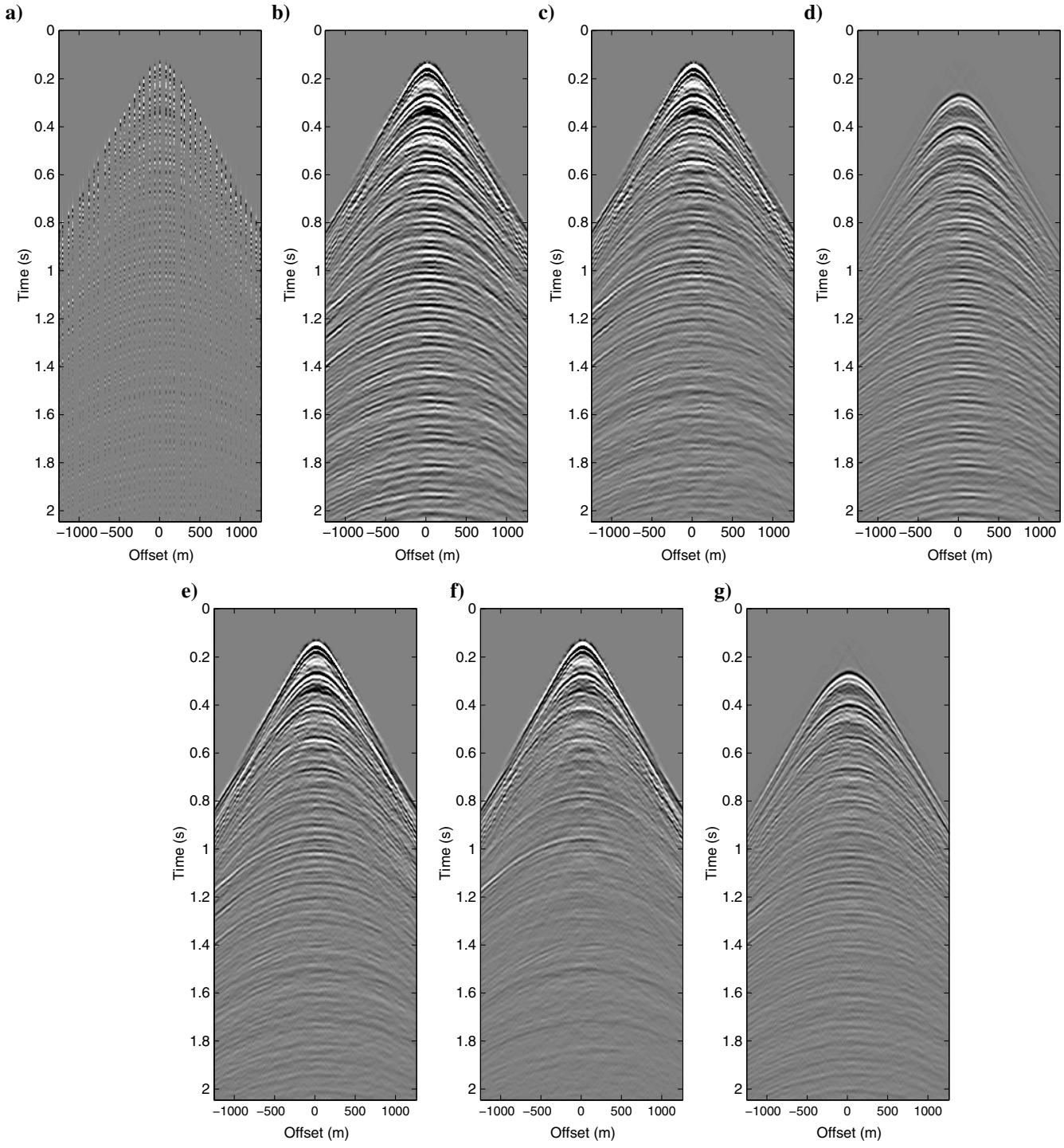

Figure 5. CL-SRME interpolation in a North Sea data set with a 1:5 shot undersampling ratio. A common receiver gather is shown. Here, we showed the (a) original undersampled data, together with the first iteration (b) reconstructed data, (c) primaries, and (d) multiples, and the 30th iteration, (e) reconstructed data, (f) primaries, and (g) multiples. The focal domains in the upper panel (Figure 5a–5c) contain the first estimates of the focal domains, still with aliasing artifacts and multiples, and with the events not sparsely represented over the focal domains. The focal domains in the lower panel (Figure 5d–5e) contain the final estimates of the focal domains, now with no aliasing artifacts, no multiples, and with events sparsely represented over the focal domains.

for convergence. The speed is, in principle, directly affected by the number of focal domains used, but this is relaxed by the fact that typically we do not require many focal domains to achieve a good result (one to three focal domains are enough in most cases). When a single focal domain is used, the current algorithm seems to be approximately two times slower than the SRME. This is due to a larger number of matrix products. With the same settings, the algorithm seems to be approximately 10 times faster than the EPSI.

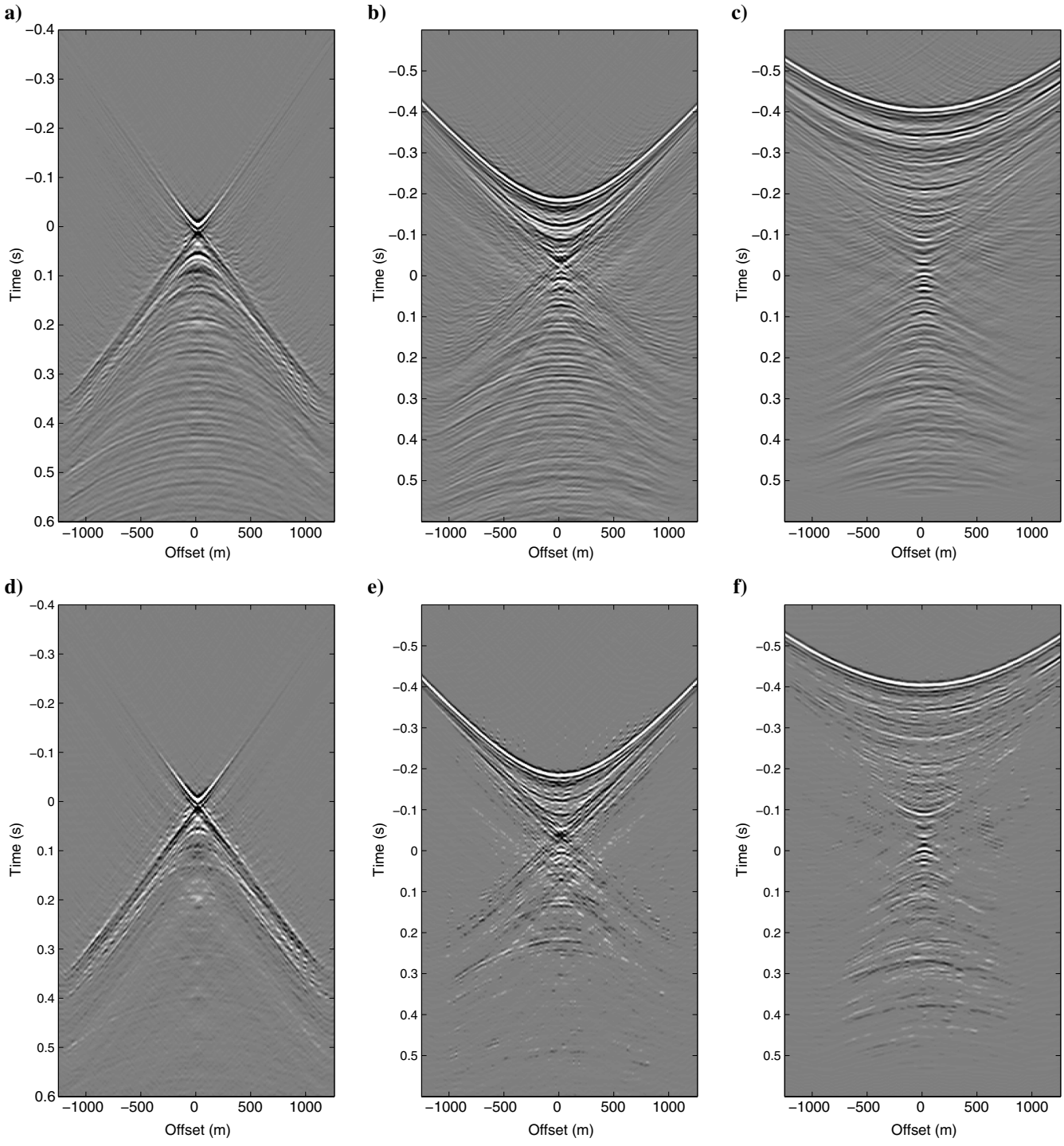

Figure 6. Focal domains of the primaries associated with the results in Figure 5 Panels (a-c) represent the first iteration primary focal domains. Panels (d-f) represent the 30th iteration primary focal domains. A common-receiver gather is shown. Some important differences can be appreciated between the upper panel (a-c) and the lower panel (d-f): (1) the elimination of the aliasing artifacts (e.g., diffractionlike events around $t = 0$), (2) the elimination of the multiples (events completely missing in the lower panel), and (3) the sparse representation of the primaries (events present in both panels but sparsely represented in the lower one).

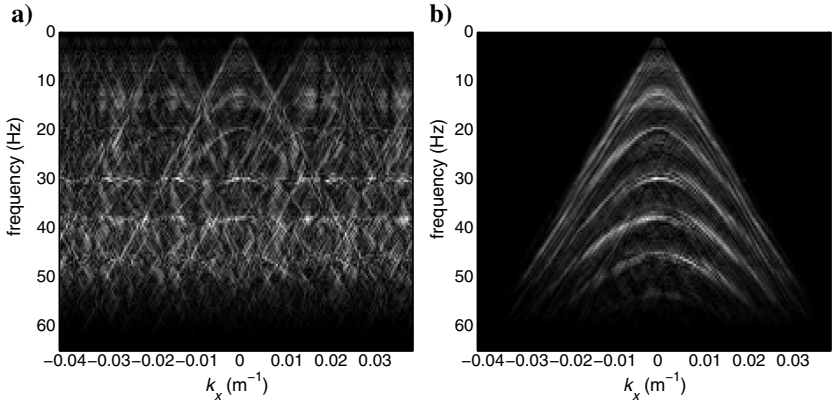

Figure 7. The k_x-ω domains of (a) the input undersampled data and (b) the output reconstructed data. Note that a 1:5 undersampling ratio in the sources was applied to the input. Both figure parts have the same color scale.

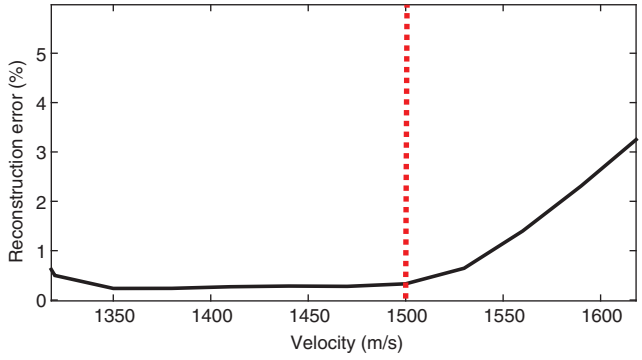

Figure 8. Reconstruction error in a typical example as a function of the associated focal operator velocity. The true velocity $c = 1500$ m/s is shown by the dotted line. As we can see, errors in the operator propagation velocity do not produce big errors in the reconstruction estimates. The higher velocities tend to be more sensitive than the lower velocities.

This is due to the fact that the EPSI requires a top-down approach in which the strongest primaries are explained first. This process typically involves many more iterations (∼100).

Though important in 2D, we expect the former method to have a stronger impact in 3D because data interpolation there is typically more challenging. However, practical problems must be faced when handling 3D data sets, the most important of them being the large data volumes present and the fact that in CL-SRME, we are required to store in memory several data sets with dense samplings (i.e., missing traces replaced by traces filled with zeros). A practical strategy to circumvent this issue would be to use a hybrid approach, such that the near offsets in the primaries are estimated by the accurate (but more expensive) CL-SRME, and the far offsets are estimated by a less accurate (but lighter) differential NMO interpolation method. This idea is left for future research beyond the scope of this paper.

Note that the CL-SRME algorithm is built to work in undersampling scenarios in which algorithms such as EPSI would not be applicable (due to the large amounts of missing data). Rather than competing with EPSI, we propose the CL-SRME algorithm as a method complementary to EPSI, to be used whenever EPSI fails to be properly applied.

CONCLUSIONS

We have introduced a new method to extend the current CL-SRME methodology such that it allows primary estimation on heavily undersampled data. This can be useful for 2D data, but it becomes mandatory for 3D data, where the data gaps tend to be much larger. To achieve an accurate reconstruction, we take the primary wavefield $\mathbf{P}_0$ to a transform domain in which the data gaps and/or interpolation errors are mapped to aliasing artifacts. The ultimate goal of the transform domain is to represent the primary energy with the smallest possible number of parameters (i.e., achieving a high information compression). We choose the focal domain, which is based on wavefield extrapolation operators, as it provides a high level of data compression. Being the product of a physics-based transform rather than a mathematical one, the focal domain also introduces additional physical constraints on the estimations, namely, wavefield continuity and wave-equation consistency. Sparseness regularization is chosen as a method to guide the estimates of the focal-domain inversion toward the elimination of the aliasing artifacts in these domains. Once all the artifacts have been cleared, the primary estimate becomes fully sampled.

Being a simultaneous reconstruction/multiple estimation algorithm, we expect the proposed method to perform better than other methods based on interpolation prior to the primary estimation. By allowing the missing data and the primary information to be simultaneously estimated, we allow their estimations to be simultaneously adjusted to optimally fit each other. Here, data reconstruction helps in obtaining better multiple estimates and the physical primary-multiple relationship (equation 1) helps in constraining the data interpolation.

Given an undersampled data set, the CL-SRME algorithm can perform simultaneous data reconstruction and primary-multiple separation. The required focal operators can be obtained via simple estimation of the rms velocities of the target reflectors.

The shown examples depict the capabilities of the current approach for 2D field data, in which heavy undersampling is overcome with data reconstruction. An important result is obtained for near-offset reconstruction in shallow-water environments, in which reconstruction over large gaps is achieved, reconstructing completely the missing shallow events. With a fast and efficient convergence, these results open the possibilities for accurate shallow-water demultiple in 2D and 3D geometries.

ACKNOWLEDGMENTS

The authors thank the sponsors of the Delphi Consortium for their support. We also thank PGS for providing the field data. Furthermore, we thank the associate editor and the reviewers of this work.

APPENDIX A

CALCULATING GRADIENTS FOR NONANALYTICAL COMPLEX-VALUED MATRIX FUNCTIONALS

GENERAL THEORY

Let

$$J(\mathbf{Z}) = Tr(\mathbf{V}(\mathbf{Z})\mathbf{V}(\mathbf{Z})^H) = \|\mathbf{V}(\mathbf{Z})\|_{Fr}^2 \tag{A-1}$$

be a matrix functional such that $J \in M_n(\mathbb{C}) \to \mathbb{C}$ where $M_n(\mathbb{C})$ is the space of $n \times n$ matrices over the complex field $\mathbb{C}$. We choose $\mathbf{V}: M_n(\mathbb{C}) \to M_n(\mathbb{C})$ to be a differentiable function of $\mathbf{Z}$ and $\|\cdot\|_{Fr}^2$ to be the Frobenius norm. Typically, $\mathbf{V}$ would correspond to the inversion residual in a descent algorithm. We are interested in calculating the gradient $\nabla_{\mathbf{Z}} J(\mathbf{Z})$ such that we can use it in an inversion algorithm. However, due to the appearance of $\mathbf{Z}$ and $\mathbf{Z}^H$ in the definition of $J(\mathbf{Z})$, the functional defined in equation A-1 is nonanalytical, thus nondifferentiable. This implies that we need to redefine the concept of gradient for this type of functional. We can resolve this problem by defining our gradient to be

$$\nabla_{\mathbf{Z}} J(\mathbf{Z}) \coloneqq 2\frac{dJ(\mathbf{Z})}{d\mathbf{Z}^*} = \frac{\partial J}{\partial \mathcal{R}\mathbf{Z}} + i\frac{\partial J}{\partial \mathcal{I}\mathbf{Z}}, \tag{A-2}$$

where $\mathcal{R}\mathbf{Z}$ and $\mathcal{I}\mathbf{Z}$ represent the real and imaginary parts of $\mathbf{Z}$, respectively. Equation A-2 is equal to zero when $J(\mathbf{Z})$ is an analytic function of $\mathbf{Z}$ and nonzero otherwise.

The gradient in equation A-2 applied to equation A-1 can be calculated using the following algorithm:

1) Take the exterior derivative of J: $J \to dJ$.
2) Use the linearity and the Leibniz rule to write dJ in $d\mathbf{Z}^*$ terms. Note that the gradient definition uses only $d\mathbf{Z}^*$, so any $d\mathbf{Z}$ variation is zero:

$$dJ = dTr(\mathbf{V}^H\mathbf{V}) = Tr(d\mathbf{V}^H\mathbf{V} + \mathbf{V}^H d\mathbf{V}). \tag{A-3}$$

3) Take dJ into a $dJ = Tr(f(\mathbf{Z})d\mathbf{Z}^*)$ form (with f a function of $\mathbf{Z}$).
4) Use the following property:

$$dJ = Tr(f(\mathbf{Z})d\mathbf{Z}^*) \to \frac{dJ(\mathbf{Z})}{d\mathbf{Z}^*} = f(\mathbf{Z}),$$

to calculate the required derivative.
5) Scale the result by two to obtain $\nabla_{\mathbf{Z}} J(\mathbf{Z})$.

Note that descent algorithms require the negative of the gradient $-\nabla_{\mathbf{Z}} J(\mathbf{Z})$ as descent direction. Also note that if $\mathbf{Z} = \mathbf{Z}^T$, $\mathbf{Z}^*$ can be replaced by $\mathbf{Z}^H$ in all the steps above.

PROOF OF EQUATION 9

In this section, we will calculate the least-squares gradient of a functional J (in equation 9) with respect to the variable $\hat{\mathbf{X}}_i(\omega')$. Start by introducing the variable $j^{(\mathrm{LS})} = \|\mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q}\|_{Fr}$, then $J^{(\mathrm{LS})}$ can be written as

$$J = \sum_{\omega} \left\| \mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q} \right\|_{Fr} = \sum_{\omega} j^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H; \omega). \tag{A-5}$$

Taking the gradient $\nabla_{\hat{\mathbf{X}}_i(\omega')}$ on this expression, we get

$$\begin{aligned}
\nabla_{\hat{\mathbf{X}}_i(\omega')} J^{(\mathrm{LS})} &= \frac{2dJ^{(\mathrm{LS})}}{d\hat{\mathbf{X}}_i(\omega')^H} \\
&= \sum_{\omega} \frac{2dj^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H; \omega)}{d\hat{\mathbf{X}}_i(\omega')^H} \\
&= \frac{2dj^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H; \omega')}{d\hat{\mathbf{X}}_i(\omega')^H} \\
&= \nabla_{\hat{\mathbf{X}}_i(\omega')} j^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H; \omega').
\end{aligned} \tag{A-6}$$

The elimination of $\sum_\omega$ in the above expression is due to the fact that all the ω frequencies are mutually independent, and then, the only nonzero contribution to the derivative comes when $\omega = \omega'$. The calculation of the derivative of $j^{(\mathrm{LS})}$ will now follow the steps outlined above. (For simplicity, we will now drop the explicit dependence of ω'.) Start rewriting $j^{(\mathrm{LS})}$ as

$$\begin{aligned}
j^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H) = Tr\Bigg[&\left(\mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q} \right)^H \\
&\times \left(\mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q} \right) \Bigg] = Tr[\mathbf{V}^H \mathbf{V}],
\end{aligned} \tag{A-7}$$

where $\mathbf{V} \coloneqq \mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q}$ is the data residual. Now taking the exterior derivative in $j^{(\mathrm{LS})}$, we get

$$dj^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H) = dTr[\mathbf{V}^H \mathbf{V}] = Tr[d\mathbf{V}^H \mathbf{V} + \mathbf{V}^H d\mathbf{V}], \tag{A-8}$$

where the last step corresponds to the Leibniz product rule. Here, the variables $\hat{\mathbf{X}}_i^H$ and $\hat{\mathbf{X}}_i$ can be taken as independent. Then, because we are considering variations over the $\hat{\mathbf{X}}_i^H$ variable only ($d\hat{\mathbf{X}}_m^H = 0$ if $m \neq i$), we can see that $d\mathbf{V}(\hat{\mathbf{X}}_i) = \mathbf{V}(d\hat{\mathbf{X}}_i) = \mathbf{0}$. Therefore, we can write

$$\begin{aligned}
dj^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H) &= Tr[d\mathbf{V}^H \mathbf{V}] \\
&= Tr\left[d\left(\mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q} \right)^H \mathbf{V} \right] \\
&= Tr[-(\mathbf{W}_i^T d\hat{\mathbf{X}}_i \mathbf{W}_i \mathbf{Q})^H \mathbf{V}] \quad = Tr[-\mathbf{Q}^H \mathbf{W}_i^H d\hat{\mathbf{X}}_i^H \mathbf{W}_i^* \mathbf{V}] \\
&= Tr[-\mathbf{W}_i^* \mathbf{V} \mathbf{Q}^H \mathbf{W}_i^H d\hat{\mathbf{X}}_i^H] \\
&= Tr\left[-\mathbf{W}_i^* \left(\mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q} \right) \mathbf{Q}^H \mathbf{W}_i^H d\hat{\mathbf{X}}_i^H \right],
\end{aligned} \tag{A-9}$$

so

$$\frac{dj^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H)}{d\hat{\mathbf{X}}_i^H} = \frac{dJ^{(\mathrm{LS})}(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_i^H)}{d\hat{\mathbf{X}}_i^H} = -\mathbf{W}_i^* \left(\mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q} \right) \mathbf{Q}^H \mathbf{W}_i^H, \tag{A-10}$$

which leads to

$$\nabla_{\hat{\mathbf{X}}_i} J^{(\mathrm{LS})} = -2\mathbf{W}_i^* \left(\mathbf{P} - \sum_m \mathbf{W}_m^T \hat{\mathbf{X}}_m \mathbf{W}_m \mathbf{Q} \right) \mathbf{Q}^H \mathbf{W}_i^H, \tag{A-11}$$

which is the desired gradient.

REFERENCES

Baardman, R. H., D. J. Verschuur, R. G. van Borselen, M. O. Frijlink, and R. F. Hegge, 2010, Estimation of primaries by sparse inversion using dual sensor data: 80th Annual International Meeting, SEG, Expanded Abstracts, 3468–3472.

Berkhout, A. J., 1982, Seismic migration, imaging of acoustic energy by wave field extrapolation, A: Theoretical aspects 2nd ed.: Elsevier.

Berkhout, A. J., and D. J. Verschuur, 1997, Estimation of multiple scattering by iterative inversion, part I: Theoretical considerations: Geophysics, **62**, 1586–1595, doi: 10.1190/1.1444261.

Berkhout, A. J., and D. J. Verschuur, 2006, Focal transformation, an imaging concept for signal restoration and noise removal: Geophysics, **71**, no. 6, A55–A59, doi: 10.1190/1.2356996.

Biersteker, J., 2001, Shell's surface multiple attenuation technique: 71st Annual International Meeting, SEG, Expanded Abstracts, 1301–1304.

Candes, E., L. Demanet, D. Donoho, and L. Ying, 2006, Fast discrete curvelet transforms: Multiscale Modeling and Simulation, **5**, 861–899, doi: 10.1137/05064182X.

Dragoset, W. H., and Z. Jeričević, 1998, Some remarks on surface multiple attenuation: Geophysics, **63**, 772–789, doi: 10.1190/1.1444377.

Dragoset, W. H., D. J. Verschuur, I. Moore, and R. Bisley, 2010, A perspective on 3D surface-related multiple elimination: Geophysics, **75**, no. 5, 75A245–75A261, doi: 10.1190/1.3475413.

Guitton, A., and D. J. Verschuur, 2004, Adaptive subtraction of multiples using the l1-norm: Geophysical Prospecting, **52**, 27–38, doi: 10.1046/j.1365-2478.2004.00401.x.

Hampson, D., 1986, Inverse velocity stacking for multiple elimination: Journal of the CSEG, **22**, 44–55.

Hargreaves, N., 2006, Surface multiple attenuation in shallow water and the construction of primaries from multiples: 76th Annual International Meeting, SEG, Expanded Abstracts, 2689–2693.

Herrmann, F. J., U. Böniger, and D. J. Verschuur, 2007, Non-linear primary-multiple separation with directional curvelet frames: Geophysical Journal International, **170**, 781–799, doi: 10.1111/j.1365-246X.2007.03360.x.

Herrmann, F. J., and D. J. Verschuur, 2008, Adaptive curvelet-domain primary-multiple separation: Geophysics, **73**, no. 3, A17–A21, doi: 10.1190/1.2904986.

Herrmann, F. J., D. Wang, G. Hennenfent, and P. P. Moghaddam, 2008, Curvelet-based seismic data processing: A multiscale and nonlinear approach: Geophysics, **73**, no. 1, A1–A5, doi: 10.1190/1.2799517.

Jumah, B., and F. J. Herrmann, 2014, Dimensionality-reduced estimation of primaries by sparse inversion: Geophysical Prospecting, **62**, 972–993, doi: 10.1111/1365-2478.12113.

Kinneging, N. K., V. Budejicky, C. P. A. Wapenaar, and A. J. Berkhout, 1989, Efficient 2D and 3D shot record redatuming: Geophysical Prospecting, **37**, 493–530, doi: 10.1111/j.1365-2478.1989.tb02220.x.

Kutscha, H., and D. J. Verschuur, 2012, Data reconstruction via sparse double focal transformation: An overview: IEEE Signal Processing Magazine, **29**, 53–60, doi: 10.1109/MSP.2012.2188209.

Kutscha, H., D. J. Verschuur, and A. J. Berkhout, 2010, High-resolution double focal transformation and its application to data reconstruction: International Geophysical Conference and Exhibition, SEG, Expanded Abstracts, 3589–3593.

Lin, T. Y., and F. J. Herrmann, 2009, Unified compressive sensing framework for simultaneous acquisition with primary estimation: 79th Annual International Meeting, SEG, Expanded Abstracts, 3113–3117.

Lin, T. Y., and F. J. Herrmann, 2013, Robust signature deconvolution and the estimation of primaries by sparse inversion: Geophysics, **78**, no. 3, R133–R150, doi: 10.1190/geo2012-0097.1.

Liu, F., M. M. N. Kabir, and D. J. Verschuur, 1995, Seismic processing using the wavelet and the radon transform: Journal of Seismic Exploration, **4**, 375–390.

Lopez, G. A., and D. J. Verschuur, 2013, 3D primary estimation by sparse inversion using the focal domain parameterization: 83rd Annual International Meeting, SEG, Expanded Abstracts, 4172–4177.

Lopez, G. A., and D. J. Verschuur, 2014, Closed-loop SRME — A new direction in surface multiple removal algorithms: 76th Annual International Meeting, EAGE, Extended Abstracts, E102.

Moore, I., and R. Bisley, 2006, Multiple attenuation in shallow-water situations: 73rd Annual International Meeting, EAGE, Extended Abstracts, 270–284.

Nekut, A. G., and D. J. Verschuur, 1998, Minimum energy adaptive subtraction in surface-related multiple elimination: 68th Annual International Meeting, SEG, Expanded Abstracts, 1507–1510.

Savels, T., K. de Vos, and J. W. de Maag, 2011, Surface multiple attenuation through sparse inversion; results for complex synthetics and real data: First Break, **29**, 55–64, doi: 10.3997/1365-2397.2011003.

van Groenestijn, G. J. A., and D. J. Verschuur, 2009a, Estimating primaries by sparse inversion and application to near-offset data reconstruction: Geophysics, **74**, no. 3, A23–A28, doi: 10.1190/1.3111115.

van Groenestijn, G. J. A., and D. J. Verschuur, 2009b, Estimation of primaries and near offsets by sparse inversion: Marine data applications: Geophysics, **74**, no. 6, R119–R128, doi: 10.1190/1.3213532.

Verschuur, D. J., and A. J. Berkhout, 1997, Estimation of multiple scattering by iterative inversion, part II: Practical aspects and examples: Geophysics, **62**, 1596–1611, doi: 10.1190/1.1444262.

Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177, doi: 10.1190/1.1443330.

Weglein, A. B., F. A. Gasparotto, P. M. Carvalho, and R. H. Stolt, 1997, An inverse scattering series method for attenuating multiples in seismic reflection data: Geophysics, **62**, 1975–1989, doi: 10.1190/1.1444298.

Zwartjes, P. M., and A. Gisolf, 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: Geophysics, **71**, no. 6, V171–V186, doi: 10.1190/1.2348633.