submitted to *Geophys. J. Int.*

Fast imaging with surface-related multiples by sparse inversion

Ning Tu[1] and Felix J. Herrmann[1]

[1] *Seismic Laboratory for Imaging and Modelling, Department of Earth and Ocean Sciences, the University of British Columbia, 6339 Stores Road, Vancouver, V6T 1Z4, BC, Canada,*

Email: tning@eos.ubc.ca, fherrmann@eos.ubc.ca

SUMMARY

In marine exploration seismology, surface-related multiples are usually treated as noise mainly because subsequent processing steps, such as migration velocity analysis and imaging, require multiple-free data. Failure to remove these wavefield components from the data may lead to erroneous estimates for migration velocity or result in strong coherent artifacts that interfere with the imaged reflectors. However, multiples can carry complementary information compared to primaries, as they interact with the free surface and are therefore exposed more to the subsurface. Recent work has shown that when processed correctly multiples can improve seismic illumination. Given a sufficiently accurate background velocity model and an estimate for the source signature, we propose a new and computationally efficient two-way wave-equation based linearized inversion procedure that produces accurate images of the subsurface from the total upgoing wavefield including surface-related multiples. Modelling of the surface-related multiples in the proposed method derives from the well-known surface-related multiple elimination method. We incur a minimal overhead from incorporating the multiples by having the wave-equation solver carry out

the multiple predictions via the inclusion of an areal source instead of expensive dense matrix-matrix multiplications. By using subsampling techniques, we obtain high-quality true-amplitude least-squares migrated images at computational costs of roughly a single reverse-time migration with all the data. These images are virtually free of coherent artifacts from multiples. Proper inversion of the multiples would be computationally infeasible without using these techniques that significantly brings down the cost. By promoting sparsity in the curvelet domain and using rerandomization, out method gains improved robustness to errors in the background velocity model, and errors incurred in the linearization of the wave-equation with respect to the model. We demonstrate the superior performance of the proposed method compared to the conventional reverse-time migration using realistic synthetic examples.

Key words: Inverse theory; Computational seismology; Wave propagation

1 INTRODUCTION

While the reverse-time migration (RTM) has been very successful, its application to data sets with strong surface-related multiples is challenging because this imaging technology generally hinges on the single-scattering approximation. Ignoring these multiples can lead to erroneous velocity models and images that are prone to coherent artifacts that can lead to misguided interpretation. Throughout this paper, we discuss surface-related multiples unless specified otherwise.

To overcome this problem, industry has been adopting workflows that are designed to remove surface-related multiples prior to velocity analysis and migration. Early methods employed periodicity of multiples, which lead to trace-based predictive deconvolution (Peacock & Treitel 1969), while later multi-trace techniques, such as NMO and the Radon transform, were used to separate primaries and multiples based on kinematic differences in the data space (Hampson 1986; Foster & Mosher 1992).

Unfortunately, these above techniques fail in cases where primaries and multiples overlap and where discrimination based on periodicity and kinematics alone is insufficient. In these cases, wave-equation based methods offer a viable, although computationally expensive alternative. Amongst these wave-equation based methods, we count Surface-Related Multiple Elimination (SRME) (Verschuur et al. 1992; Guitton & Verschuur 2004; Herrmann et al.

2008c; Wang et al. 2008), model-based multiple prediction methods as reported by Wiggins (1988) and Lu et al. (1999), the inverse scattering method by Weglein et al. (1997), and wave-equation based moveout discrimination by Nemeth et al. (2000) and Sava & Guitton (2005).

An alternative to multiple removal is to actively use them to benefit from the extra information they carry, and to altogether avoid the prediction-subtraction de-multiple paradigm that may hurt the primaries (e.g., SRME). In data space processing, we count Estimation of Primaries by Sparse Inversion (EPSI), which inverts the SRME relation by identifying the surface-free dipole Green's function and the source wavelet as unknowns (van Groenestijn & Verschuur 2009a; Lin & Herrmann 2013). Compared to the prediction-subtraction paradigm of SRME where multiples and therefore the information they carry are removed, EPSI maps the surface-related multiples back to the primary impulse response via multi-dimensional deconvolutions. As a result, primaries are better preserved in EPSI (see van Groenestijn & Verschuur 2009a Figure (4) for example). However, this progress comes at a price. Its iterative inversion procedure involves significantly more computational work (e.g., multiplications of dense matrices, the computational complexity of which is $\mathcal{O}(n^3)$ for $n \times n$ matrices) than SRME. This may partially explain the relatively slow uptake of EPSI by industry for 3D seismic, where the computation can be prohibitively expensive. In model space processing such as imaging, many works have shown that surface-related multiples are useful signals and can increase illumination coverage when used properly (see examples in Berkhout 1993; Guitton 2002; Verschuur 2011; Long et al. 2013). Among these works, the first school of methods treat the downgoing receiver wavefield as the generalized areal source for the surface-related multiples (Berkhout 1993; Guitton 2002; Shan 2003; Schuster et al. 2004; Berkhout & Verschuur 2006; Muijs et al. 2007; He et al. 2007; Whitmore et al. 2010; Liu et al. 2011; Verschuur 2011; Lu et al. 2011; Long et al. 2013, and counting). These methods can be derived from the same multiple prediction principle as in SRME, by identifying the surface-related multiples to be the multi-dimensional convolution of the Green's function and the downgoing receiver wavefield. Most of these methods apply the cross-correlation imaging condition to image these multiples, which results in acausal cross-correlation artifacts in the image (a more detailed explanation can be found in Muijs et al. 2007; Liu et al. 2011). Others propose to use the deconvolutional imaging condition to reduce these artifacts (Shan 2003; Muijs et al. 2007; Whitmore et al. 2010), but with limited success when complex geology is present (see Poole et al. 2010; Tu et al. 2013c for a detailed explanation). Methods exploiting least-squares migration to achieve this goal have also emerged in the literature (He & Schuster 2003; Lin et al. 2010; Verschuur

2011; Tu & Herrmann 2012a; Wong et al. 2012, 2014). However, wide adoption of these iterative methods is largely hindered by the expensive simulation cost (in terms of the number of wave-equation solves). The second school of imaging-with-multiple methods rely on including a free surface (and / or other reflecting boundaries) in the background model (e.g., Reiter et al. 1991; Brown & Guitton 2005; Wong et al. 2012). These methods do not fit into current single-scattering based migration paradigm, and are less discussed in the literature. Nevertheless, such a method that accounts for different orders of surface-related multiples still relies on a proper iterative inversion procedure (for example Brown & Guitton 2005; Wong et al. 2012) and therefore also faces the challenge of expensive computation.

1.1 Contributions of this work

Inspired by the computationally expensive but successful SRME-based iterative inversion approaches (e.g., EPSI and the least-squares type imaging-with-multiples methods), we present a new and computationally efficient inversion method to image data with surface-related multiples. As far as we know, our approach is the first instance where primaries and all orders of surface-related multiples are imaged jointly using a two-way wave-equation based linearized inversion method in a computationally efficient manner. As we will show, the method eliminates dense, and therefore prohibitively expensive, matrix-matrix multiplications by solving wave-equations w.r.t. an areal-source term that contains the total downgoing wavefield. By incorporating this areal source into recently proposed compressive imaging (Herrmann & Li 2012), we arrive at a formulation that creates high fidelity images at low computational costs. We control acausal cross-correlation artifacts and computational cost by promoting curvelet-domain sparsity in the model space.

1.2 Paper outline

We organize the paper as follows: first, we introduce the matrix equation that establishes the physical relationship between the up- and downgoing wavefields at the free surface. Next, we combine this relationship with linearize forward modelling using the solution operator of the time-harmonic Helmholtz system. This combination enables the use of all the techniques discussed in this paper to bring down the computation. With the linearized forward model, we introduce migration with multiples and sparsity-promoting imaging with multiples, using compressive imaging techniques to bring down the cost of our method on a par with that of a conventional RTM. We conclude by evaluating our algorithm using a stylized example and

a realistic synthetic case study that demonstrate the superior performance of the proposed method compared to conventional RTM.

2 IMAGING WITH MULTIPLES

Because we want to image the surface-related multiples, we start our treatise by first introducing the time-harmonic relationship between the up- and downgoing wavefields described by the SRME relation, and then propose to combine this relation with wave-equation based linearized inversion. This will allow us to directly relate the total data, including surface-related multiples, to the perturbations of the subsurface model.

2.1 Physics of the free surface

The SRME relation explicitly reveals the role of the surface-free Green's function that relates the total up- and downgoing pressure wavefields at the free surface. Using monochromatic matrices, it can be expressed as (Verschuur et al. 1992):

$$\mathbf{P}_i = \mathbf{X}_i(\mathbf{S}_i + \mathbf{R}_i\mathbf{P}_i), \tag{1}$$

where subscript $i = 1, \cdots, n_f$ indexes the frequencies where n_f is the total number of discretized frequencies, $\mathbf{P}_i$ is the discretized total upgoing receiver wavefield at the free surface, $\mathbf{X}_i$ is the surface-free dipole Green's function, and $\mathbf{S}_i$ is the downgoing point-source wavefield. The operator $\mathbf{R}_i$ models the reflectivity at the free surface, and as a result, $\mathbf{R}_i\mathbf{P}_i$ is the *down*going receiver wavefield at the surface that acts as a generalized "areal-source" wavefield for the surface-related multiples. The above relation reveals that the total upgoing wavefield can be represented by a frequency-domain (dense) matrix-matrix product of the surface-free dipole Green's function and the total downgoing wavefield (including the downgoing point-source wavefield and the downgoing receiver wavefield). Note that because direct waves do not reflect at the surface, they are excluded from both the data $\mathbf{P}_i$ and the dipole Green's function $\mathbf{X}_i$ (Verschuur 1991; Dragoset & Jeričević 1998). Internal multiples are part of the primary response in the SRME relation. To obtain the upgoing wavefield at the surface, receiver ghosts (i.e., the downgoing data) need to be removed (source ghosts should be kept untouched) (Verschuur 1991; Verschuur et al. 1992; Dragoset & Jeričević 1998), and then the upgoing data need to be extrapolated from the receiver level to the free surface, as only at the surface can the reflectivity operator $\mathbf{R}_i$ be approximated by $-\mathbf{I}$.

In Equation (1), the monochromatic upgoing wavefield is represented as a complex-valued matrix, $\mathbf{P}_i \in \mathbb{C}^{n_r \times n_s}$, with n_s common-shot gathers in its columns and n_r common-receiver

gathers in its rows. The dipole Green's function $\mathbf{X}_i$ is organized in the same way. Equation (1) assumes fixed-spread acquisition with co-located sources and receivers (i.e., $n_s = n_r$), and therefore the point-source wavefield $\mathbf{S}_i \in \mathbb{C}^{n_r \times n_s}$ can also be organized in this way so that the addition of $\mathbf{S}_i$ and $\mathbf{R}_i\mathbf{P}_i$ makes physical sense. For marine streamer acquisitions, one can fill in missing data using reciprocity. With these assumptions, the matrix product in the above expression corresponds to multi-dimensional convolutions consisting of a temporal convolution along the time dimension and a spatial convolution along the source / receiver dimension. Equation (1) describes 2D seismic lines but it can be extended to 3D seismic data where frequency slices can also be organized into matrices (Verschuur et al. 1992; Dragoset et al. 2010).

Here are more details about $\mathbf{X}_i$ and $\mathbf{S}_i$. In the continuous case, the dipole Green's function is the vertical derivative of the monopole Green's function, which can be alternatively represented using the particle-velocity wavefield (again see Fokkema & van den Berg 1993 Equation (12.25)). In the discretized case, we approximate $\mathbf{X}_i$ by the response of the subsurface due to the injection of a (closely placed) pair of monopole impulsive sources (with the opposite sign) above and below the free surface, properly scaled to include the subsurface response due to source ghost. The point-source wavefield $\mathbf{S}_i$ is a diagonal matrix, with the diagonal entries being the Fourier spectrum of the source wavelet at a given frequency, i.e., $\mathbf{S}_i = w_i\mathbf{I}$ with w_i the spectrum of the source wavelet at the i^{th} frequency, and $\mathbf{I}$ the identity matrix (Verschuur et al. 1992). Throughout this paper, we approximate the surface reflectivity to be -1, i.e., $\mathbf{R}_i = -\mathbf{I}$.

While the Equation (1) has been responsible for major breakthroughs in the mitigation of surface-related multiples, including SRME (Verschuur et al. 1992) and EPSI (van Groenestijn & Verschuur 2009b; Lin & Herrmann 2013), it poses the following challenges: (i) it relies on dense matrix-matrix multiplications that are computationally expensive in terms of CPU time, memory requirement, and computer input / output, and (ii) these dense matrix products require full data sampling, rendering expensive on-the-fly data interpolations necessary to compute these matrix products. While low-rank approximation techiniques can be used to alleviate the cost of performing matrix products (Jumah & Herrmann 2014), we propose to altogether avoid these dense matrix products by using the wave-equation solvers to implicitly carry them out.

2.2 Modelling the free surface via areal sources

To arrive at a formulation that is conducive to imaging data with surface-related multiples, we combine the physics of the free surface, described in the previous section, with the wave physics of the Earth interior, i.e., we incorporate the free surface as modelled in Equation (1) into the solution of the wave equation. As imaging is based on a linearization of the wave-equation at a smooth background model $\mathbf{m}_0$, we first approximate the direct-wave-free upgoing dipole Green's function $\mathbf{X}_i$ via Born modelling, assuming that the high-order reflections such as internal multiples are weak. This approximation can be expressed as

$$\begin{aligned}\mathbf{X}_i &\approx \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}, \mathbf{I}] \\ &= \omega^2 \mathbf{D}_r \mathbf{H}_i[\mathbf{m}_0]^{-1} \mathbf{H}_i[\mathbf{m}_0]^{-1} \mathrm{diag}(\delta\mathbf{m}) \mathbf{D}_s^* \mathbf{I} \\ &\doteq \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}](\mathbf{D}_s^* \mathbf{I}),\end{aligned} \tag{2}$$

where $\nabla\mathcal{F}_i$ denotes linearized forward modelling at the background model $\mathbf{m}_0$, which is linear in both the model perturbations $\delta\mathbf{m}$ and the source $\mathbf{I}$ (denoting the impulsive source array for the Green's function). The matrix $\mathbf{D}_r$ is a detection operator that collects data at receiver locations, and $\mathbf{D}_s^*$ injects the source wavefield. Again for the dipole-source effect, a pair of monopole sources are injected at each physical source location (with a positive sign) and the mirrored location w.r.t the free surface (with a negative sign). The matrix $\mathbf{H}_i$ represents the discretized monochromatic Helmholtz operator with absorbing boundaries (including at the surface, as the Green's function is surface free). Assuming constant density, the operator can be parameterized by the square of the slowness ($\mathrm{s}^2/\mathrm{m}^2$) discretized in $\mathbf{m}$. The symbol "$\doteq$" means "defined as". We include the third line of Equation (2), which explicitly reveals the linearity of the forward modelling w.r.t. the source, to facilitate the incorporation of surface-related multiples, which we will discuss in the next paragraph.

To incorporate surface-related multiples, a linearized modelling of the total upgoing wavefield can be derived by replacing $\mathbf{X}_i$ in Equations (1) by (2), yielding:

$$\mathbf{P}_i \approx \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}, \mathbf{I}](\mathbf{S}_i - \mathbf{P}_i). \tag{3}$$

The use of the symbol $\approx$ indicates that by linearized modelling of $\mathbf{X}_i$, we ignore higher-order scattering in the total upgoing data such as internal multiples. Compared to Equation (1), this linearized modelling involves the same dense matrix-matrix multiplications to predict the multiples, plus we need to simulate the full Green's function $\mathbf{X}_i$ by solving wave-equations for all (sequential) source experiments. To avoid the expensive dense matrix products, we simplify the forward modelling using the associativity of the matrix product, and arrive at

the following identity:

$$\begin{aligned}\mathbf{P}_i &\approx \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}, \mathbf{I}](\mathbf{S}_i - \mathbf{P}_i)\\ &= \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}](\mathbf{D}_s^*\mathbf{I})(\mathbf{S}_i - \mathbf{P}_i)\\ &= \nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}](\mathbf{D}_s^*(\mathbf{S}_i - \mathbf{P}_i))\\ &\doteq \nabla\mathcal{F}_i[\mathbf{m}_0, \mathbf{S}_i - \mathbf{P}_i]. \end{aligned} \tag{4}$$

This simplified expression forms the basis of the proposed fast linearized inversion procedure to image data with surface-related multiples. The seemingly trivial algebraic manipulations in Equation (4) have profound implications on the computations. First, the expensive matrix-matrix products between the Green's function and the total downgoing wavefield (the second line) are turned into wave-equations solves with the same downgoing wavefield injected as source (the third line). Second, the full source sampling required by the SRME relation can now be relaxed in seismic imaging: compared with the data-space inversion procedure, namely EPSI (van Groenestijn & Verschuur 2009a; Lin & Herrmann 2013), which identifies the entire Green's function $\mathbf{X}_i$ in Equation (1) as the unknown, we optimize in the model space and therefore have much less unknowns in the model parameters $\mathbf{m}$. This significant reduction in the number of unknowns leaves room for us to use less source experiments to recover the model parameters (however the underlying source grid should still co-locate with the receivers for the SRME relation to hold, although we do not have to put sources on every source grid point). Third, this simplified expression enables us to reduce the number of wave-equation solves by subsampling the source experiments (note that one source experiment corresponds to one column of the data matrices $\mathbf{P}_i$ and $\mathbf{S}_i - \mathbf{P}_i$) (Herrmann & Li 2012). According to the second line of Equation (4), we still need to solve wave-equations for all sequential sources to obtain the full Green's function $\mathbf{X}_i$ to perform matrix-matrix product, i.e., the subsampling does not bring down the simulation cost. According to the third line, we only need to solve wave-equations for the subsampled source wavefield and therefore have less source experiments to simulate. Without the simplified expression by Equation (4), significant reduction in simulation cost would be impossible.

To facilitate further discussion and problem formulation, we cast the imaging problem in the canonical form of solving a linear system of equations, i.e., $\mathbf{A}\mathbf{x} = \mathbf{b}$ where $\mathbf{A}$ is a linear operator and $\mathbf{x}$ and $\mathbf{b}$ are vectors. Throughout this paper, lower case symbols refer to vectors and upper-case symbols to matrices or operators. By vectorizing the wavefield $\mathbf{P}_i$ and rewriting the forward modelling in Equation (4) in a linear operator form that explicitly reveals the

model perturbations as the unknown, we have the following monochromatic expression

$$\begin{aligned}\mathbf{p}_i &= \mathrm{vec}(\mathbf{P}_i) \\ &\approx \mathrm{vec}(\nabla\mathcal{F}_i[\mathbf{m}_0, \delta\mathbf{m}, \mathbf{S}_i - \mathbf{P}_i]) \\ &\doteq \nabla\mathbf{F}_i[\mathbf{m}_0, \mathbf{S}_i - \mathbf{P}_i]\delta\mathbf{m},\end{aligned} \tag{5}$$

where "vec" means to vectorize a matrix, and $\nabla\mathbf{F}_i$ is the linearized demigration (i.e., Born modelling) operator. The above expression for multiple frequencies can be obtained by stacking over the rows (i.e., vertical concatenation along the columns). This expression relates the medium perturbations $\delta\mathbf{m}$ to the total upgoing wavefield in a linear manner. Remember that these multiples are modelled by injecting the downgoing receiver wavefield as an areal source, rather than by placing a free surface in the background model. With this linearized expression in place, we will now show how conventional RTM fails to correctly image the surface-related multiples, which motivates us to exploit the least-squares inversion approach with sparsity promotion.

2.3 Migraion with multiples by cross-correlation

Imaging with multiples by applying the cross-correlation imaging condition can be mathematically interpreted as applying the adjoint of the linearized demigration operator to the total upgoing data. In our case, we use two-way propagators and this amounts to the reverse-time migration (Baysal et al. 1983). Using Equation (5), we obtain the RTM image after summing over all frequencies:

$$\widetilde{\delta\mathbf{m}} = \sum_{i=1}^{n_f} \nabla\mathbf{F}_i^{\mathrm{H}}[\mathbf{m}_0, \mathbf{S}_i - \mathbf{P}_i]\mathbf{p}_i \tag{6}$$

where $\widetilde{\delta\mathbf{m}}$ denotes the RTM image, and superscript H denotes the adjoint. As there are primaries and different orders of surface-related multiples in the total up- and downgoing wavefields, cross-correlation between wrong pairs of up- and downgoing wavefields will lead to acausal artifacts in the image (see detailed explanations by Muijs et al. 2007; Liu et al. 2011).

We use a relatively simple velocity Earth model to illustrate these artifacts. The background model and the model perturbations are shown in Figure 1(a) and 1(b). The model dimensions are 1250 m deep and 2235 m wide with a five-meter grid distance in both dimensions. We make linearized data according to Equation (5) assuming a known source wavelet spectrum $\mathbf{w}$ for which we use a Ricker wavelet, and known areal-source wavefields $-\mathbf{P}_i$ that we simulate using finite-difference modelling with a free-surface boundary condition. The frequency band of the source wavelet goes up to 60 Hz. The total recording time is 2.044 s with

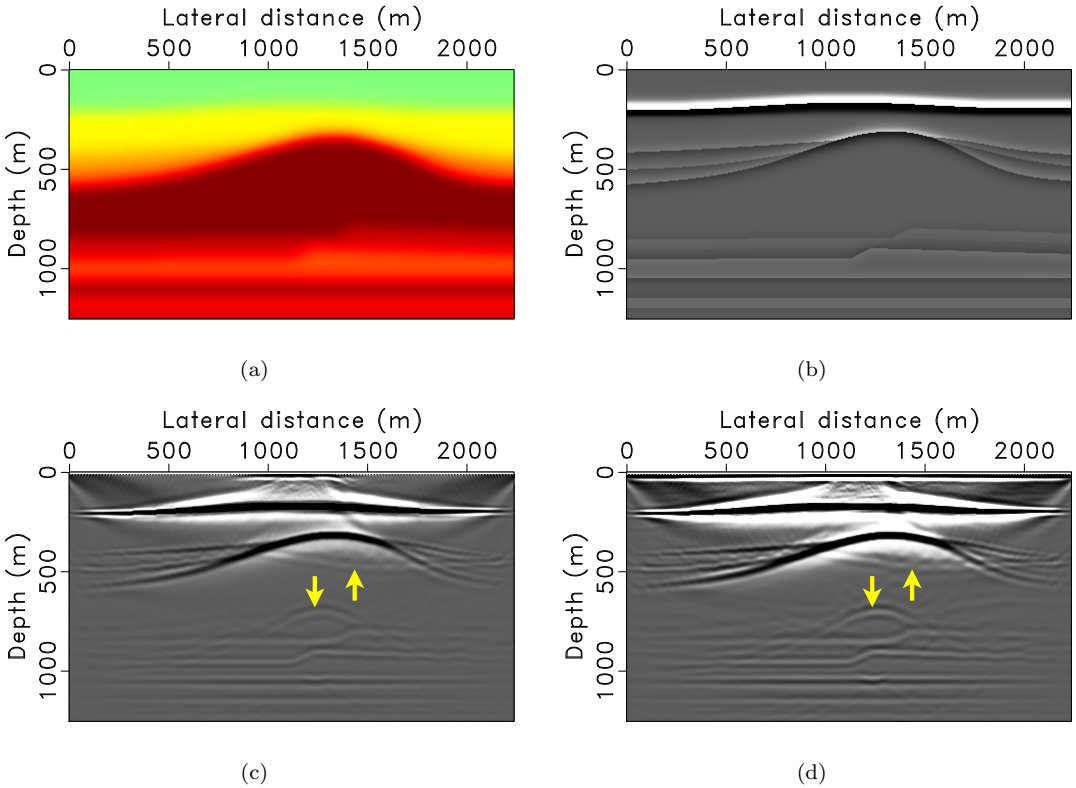


Figure 1. The synthetic salt dome model and two RTM images. **(a)** The background model. **(b)** Model perturbation. **(c)** RTM of total data but only with the primary imaging operator. **(d)** RTM of total data by including the total downgoing wavefield in the source wavefield.

4 ms sampling interval, which corresponds to a total number of 122 non-zero frequencies in the specified frequency range. There are 150 co-located sources and receivers with 15 m spacing.

We show two RTM images here. In both cases, the input data contain surface-related multiples. The first image is based on the conventional single-scattering forward model (i.e., the demigration operator only models the primaries). The result is shown in Figure 1(c). We can clearly see the phantom reflectors imaged from multiples (indicated by the arrows in the figure). This is understandable since the multiples are not taken into account by the modelling operator. The second image is by Equation (6), which models the surface-related multiples by using the total downgoing wavefield as the areal source wavefield. This image is shown in Figure 1(d). Although the multiples are modelled by the modelling operator, we still observe severe coherent artifacts (again indicated by the arrows in the figure) caused by acausal cross-correlations between wrong pairs of up- and downgoing wavefields that contain multiples. To overcome these harmful artifacts, we propose to properly invert the linearized modelling operator instead of applying the adjoint.

3 FAST SPARSITY-PROMOTING IMAGING WITH MULTIPLES

The computational cost of inverting Equation (5) for all the frequencies by the iterative least-squares migration is excessive, because each iteration involves one reverse-time migration and demigration (Nemeth et al. 1999). Without storing the wavefield, this amounts to solving four wave equations for each monochromatic source experiment. Therefore the total number of wave-equation solves roughly equals to $4 \times n_f \times n_s \times n_{iter}$, with n_{iter} the number of iterations of the inversion. We reduce this excessive simulation cost by subsampling the monochromatic source experiments and promoting curvelet-domain sparsity of the model parameters (Herrmann et al. 2009; Herrmann & Li 2012).

We use synthetic examples to help readers understand the different components of the method. In these examples, we use the same model as in Figure 1. To obtain a baseline image, we first run an inversion of all the data for 60 iterations. For a fair comparison with the following compressive imaging examples, we cast the inversion as the same Basis Pursuit De-Noise (BPDN) problem (Chen et al. 2001; van den Berg & Friedlander 2008), and use the same SPGℓ_1 solver (van den Berg & Friedlander 2008). Details about the BPDN problem and the SPGℓ_1 solver will follow. The result is shown in Figure 2(a), which is as expected more or less artifact free as all surface-related multiples are correctly mapped to locations of the true reflectors. The coherent artifacts in Figure 1(d) are largely removed by adopting the inversion approach. However, the computation for Figure 2(a) is expensive, involving roughly 4.4 million wave-equation solves.

3.1 Subsampling monochromatic source experiments

To reduce the number of wave-equation solves, which is proportional to the number of monochromatic source experiments, we follow Herrmann et al. (2009) and subsample over both frequencies and source experiments. First, we only use a randomized subset of frequencies, denoted by Ω. Second, we replace the monochromatic forward modelling in Equation (5) by

$$\begin{aligned} \underline{\mathbf{p}}_i &= \mathrm{vec}(\mathbf{P}_i\mathbf{E}) \\ &\approx \nabla\mathbf{F}_i[\mathbf{m}_0, (\mathbf{S}_i - \mathbf{P}_i)\mathbf{E}]\delta\mathbf{m} \\ &= \nabla\mathbf{F}_i[\mathbf{m}_0, \underline{\mathbf{S}}_i - \underline{\mathbf{P}}_i]\delta\mathbf{m}. \end{aligned} \tag{7}$$

In this expression, the underlined quantities refer to subsampled source and receiver wavefields by randomized source superimpositions (i.e., forming simultaneous sources) via the action of

a tall (i.e., more number of rows than columns) source-encoding matrix $\mathbf{E}$ on the right, which has Gaussian-distributed randomized entries. The number of simultaneous sources is defined as the number of columns of the encoding matrix $\mathbf{E}$. Since there are now fewer source experiments, we have fewer wave-equations to solve. However, this speed-up goes at the expense of introducing source crosstalks into an RTM image, caused by the randomized source superimposition (see Romero et al. (2000), Figure (1), and Herrmann & Li (2012), Figure (1b)). We also show such an example here. Instead of using all 150 sequential sources, we use 10 simultaneous sources. For this example we use all the frequencies. We then compute the RTM image of the 10 simultaneous shot-gathers and show it in Figure 2(b). In this image we can identify the noisy source crosstalks as well as the coherent acausal artifacts caused by the surface-related multiples. We propose to remove these crosstalk artifacts by promoting curvelet-domain sparsity of the model perturbations using the ℓ_1-norm.

3.2 Removal of source crosstalks by promoting sparsity

Following sparse-recovery and randomized sampling ideas from Compressive Sensing (CS) (Candès et al. 2006b; Donoho 2006), we remove these artifacts by promoting sparisty of the model perturbations using the ℓ_1-norm. To maximally exploit the sparsity, we incorporate the curvelet transform that affords a sparse representation of the model perturbations (Candès & Donoho 2004; Herrmann et al. 2008a,b; Herrmann & Hennenfent 2008; Neelamani et al. 2010). As a result, we aim to solve the following Basis Pursuit De-Noise (BPDN) problem:

$$\begin{aligned} \delta\tilde{\mathbf{m}} &= \mathbf{C}^{\mathrm{H}} \operatorname*{argmin}_{\mathbf{x}} \|\mathbf{x}\|_1 \\ &\text{subject to } \sum_{i\in\Omega} \|\underline{\mathbf{p}}_i - \nabla\mathbf{F}_i[\mathbf{m}_0, \underline{\mathbf{S}}_i - \underline{\mathbf{P}}_i]\mathbf{C}^{\mathrm{H}}\mathbf{x}\|_2^2 \le \sigma^2, \end{aligned} \tag{8}$$

where $\mathbf{C}^{\mathrm{H}}$ represent the curvelet synthesis operator, i.e., its action is to apply the inverse curvelet transform (Candès et al. 2006a), and $\mathbf{x}$ is the curvelet coefficient vector for the image. This optimization program finds, among all curvelet coefficient vectors, the one with the smallest ℓ_1-norm subject to the fact that its inverse curvelet transform yields demigrated (simultaneous) data that explain the input (simultaneous) data within a user-specified σ that allows for noise in the data and modelling errors (e.g., internal multiples). We solve the above BPDN problem using the SPGℓ_1 solver, which, instead of directly dealing with the BPDN formulation, solves a series of underlying Least Absolute Shrinkage and Selection Operator (LASSO) subproblems (Tibshirani 1996). We refer to Appendix A for a more detailed explanation of how SPGℓ_1 solves a general BPDN problem.

The CS theory states that solving such a convex sparsity program (cf. Equation (8)) can be used to invert underdetermined systems of equations as long as these systems behave approximately as Gaussian matrices, and $\delta\mathbf{m}$ permits a sufficiently sparse curvelet representation. The latter argument is made relatively easy by referring to the body of literature that has appeared on the compressibility of seismic images w.r.t. curvelets (Herrmann et al. 2008a,b). The first argument is a bit more difficult to make although we can argue that the combination of curvelets, randomized source superimposition, demigration, and free-surface-multiple generation leads to matrices that are well-mixed and not too far off from Gaussian matrices. However, the demigration operator has a null space caused by finite acquisition aperture and complex geologies, which may result in unresolvable steep dips and shadow zones (see e.g., Prucha & Biondi 2002; Herrmann et al. 2008a). Therefore we cannot expect perfect recoveries.

Using the same setup as for Figure 2(b), i.e., 10 simultaneous sources, we obtain an inversion result after 60 SPGℓ_1 iterations. The result is shown in Figure 2(c). Compared to Figure 2(b), the source crosstalks, as well as the acausal artifacts associated with the multiples, are both removed in Figure 2(c) by adopting the above optimization formulation that promotes sparsity in the curvelet domain (cf. Equation (8)). Compare to Figure 2(a), we reduce the number of wave-equation solves by 15 folds.

While a 15$\times$ speed-up is encouraging, Figure 2(c) still involves roughly 8$\times$ the simulation cost of a conventional RTM of all the data. We need to bring down the cost even further to make the method computationally feasible for large data sets. However, the SPGℓ_1 solver converges too slowly if we subsample too much. We show an example where we only use two simultaneous sources and 15 frequencies. The image after 305 iterations is shown in Figure 2(d), where we still see lots of remnant noisy artifacts. As explained in recent work on compressive imaging and Approximate Message Passing (AMP), these artifacts are related to the correlation built up between the model iterate and the randomized subsets of monochromatic source experiments (Herrmann & Li 2012; Herrmann 2012; Montanari 2010). We refer to Appendix B for more details about the cause of these artifacts. We will now explain how we remove these artifacts.

3.3 Further acceleration by rerandomization

Motivated by recent findings in AMP (Donoho et al. 2009; Montanari 2010), we proposed a simple rerandomization of the BPDN problem (cf. Equation (8)) to remove the remnant incoherent artifacts related to correlation build up shown in Figure 2(d). For our problem, this corresponds to selecting a new set of randomized source experiments and a new randomized

frequency subset for each LASSO subproblem (Herrmann & Li 2012; Tu & Herrmann 2012b). This practice misses a strict theoretical justification but is close in spirit to other randomized algorithms that have let to fast solutions of overdetermined problems, e.g., the randomized (block) Kaczmarz algorithm and the stochastic gradient descent method (Strohmer & Vershynin 2009; Needell & Tropp 2014; Needell et al. 2014).

To verify the efficacy of rerandomization by comparing with Figure 2(d) where no rerandomization is used, we rerun the above example using the same setup (i.e., two simultaneous sources, 15 frequencies, 305 SPGℓ_1 iterations) with rerandomization. The result is shown in Figure 2(e), and we can see that the image quality has drastically improved compared to Figure 2(d). Compared with Figure 2(a), we achieve a 120$\times$ speed-up without notable compromise on the image quality. In fact, the simulation cost to obtain Figure 2(e) roughly equals to that of a single RTM of all the data, i.e., the rerandomization technique enables us to obtain a high-fidelity true-amplitude least-squares migrated image when the computational budget only allows for one RTM with all the data.

3.4 Sparsity promoting by ℓ_1 or ℓ_2

While the success of the speed-up really hinges on the combination of the two contributing factors—sparsity promoting and rerandomization—the role that the ℓ_1-norm plays in promoting curvelet-domain sparsity should not be understated.

To demonstrate the importance of using the ℓ_1-norm as a measurement of sparsity (Mallat 2009), we solve a problem that is otherwise the same as Equation (8) except that we replace the ℓ_1-norm by the ℓ_2-norm in the objective function. Again we use two simultaneous sources, 15 frequencies, 305 iterations and rerandomization. The result is shown in Figure 2(f). We can see that it has more noise compared with Figure 2(e) where we use the ℓ_1-norm objective. Therefore the ℓ_1-norm should be used.

3.5 Putting it all together

Now we have explained each component of the proposed method, we summarize and explain the algorithm in Algorithm (1), which is an adapted version of the compressive imaging algorithm by Herrmann & Li (2012).

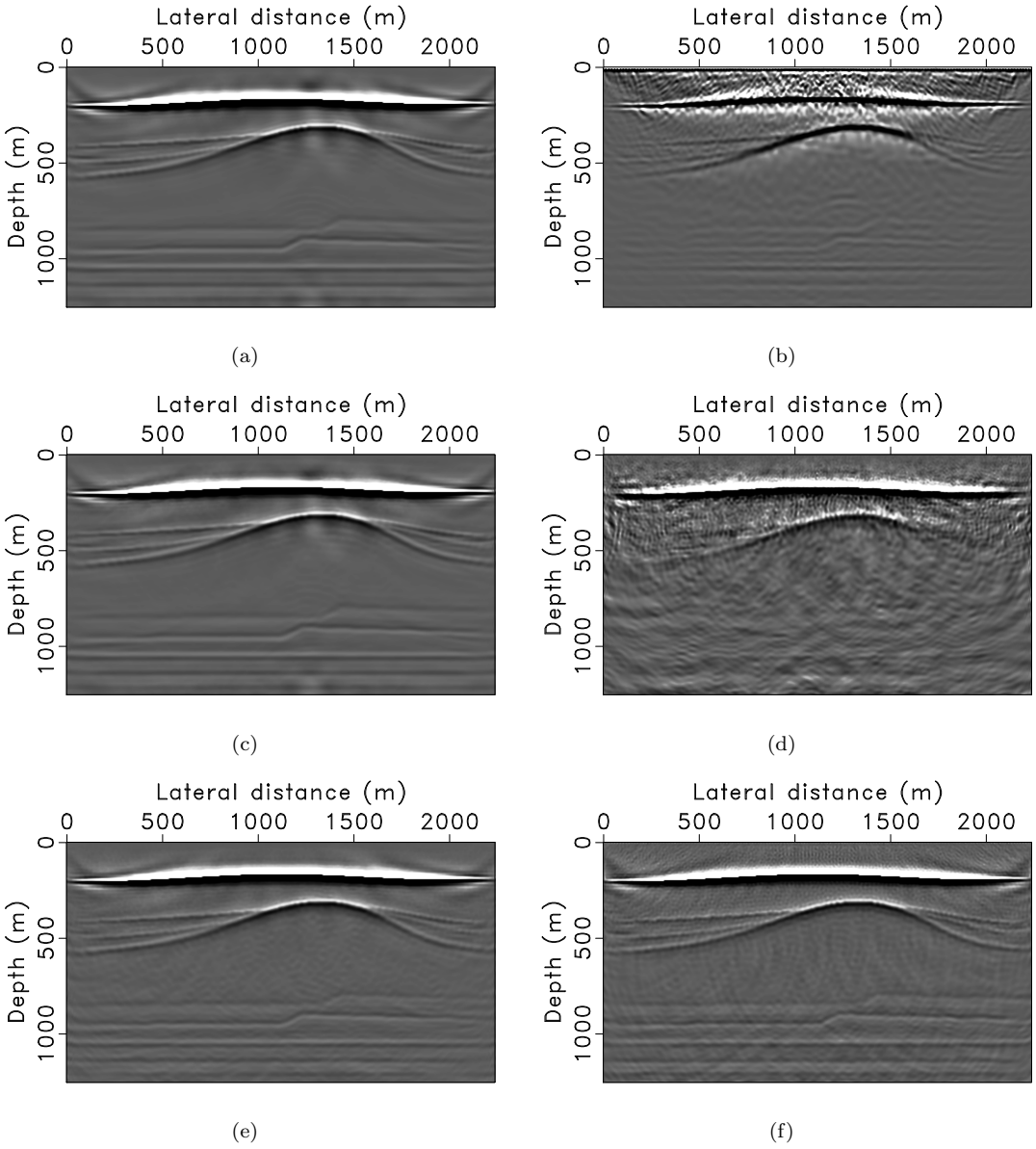


Figure 2. Examples using the salt dome model. **(a)** Inversion with all the data. **(b)** Migration with 10 simultaneous sources. **(c)** Inversion with 10 simultaneous sources. **(d)** Inversion with 2 simultaneous sources and 15 frequencies without rerandomization. **(e)** Same as **(d)** but with rerandomization. **(f)** Same as **(e)** but by ℓ_2 norm minimization.

3.5.1 Input and initialization [lines 1–4]

As we stated before, our inversion algorithm requires estimates of the source wavelet and the background velocity model $\mathbf{m}_0$. The solution vector $\mathbf{x}_0$ is initialized to be a zero vector. We select the tolerance parameter σ simply to be zero (i.e., we only stop when we reach the maximal number of iteration), and choose the number of simultaneous source experiments $n'_s \ll n_s$, the number of randomized frequencies $n'_f \ll n_f$ and the maximal number of it-

erations k_{max} based on our computational budget (Tu et al. 2013b). We use l to index the LASSO subproblems.

3.5.2 Main loop [lines 5–10]

Given the input and after the initialization, we enter into the main loop in which we solve the LASSO subproblems. For each subproblem, we draw new independent randomized subsets of frequencies Ω and source experiments by drawing new source encoding matrix $\mathbf{E}$ (line 6). Without this "redraw" step, the algorithm would be identical to the original SPGℓ_1. Following van den Berg & Friedlander (2008), we compute the sparsity constraint for the l-th LASSO subproblem τ_l using τ_{l-1} and σ using Newton's method on the Pareto curve (line 7), and solve the $\mathbf{LS}_\tau$ subproblem (line 8).

Algorithm 1 Fast imaging with multiples

1: **Input:**

2: total upgoing wavefield $\mathbf{P}_i$, point-source wavefield $\mathbf{S}_i$,
background velocity model $\mathbf{m}_0$, tolerance $\sigma = 0$, iteration limit k_{max}

3: **Initialization:**

4: $k \leftarrow 0, l \leftarrow 0, \mathbf{x}_l \leftarrow \mathbf{0}$

5: **while** $k < k_{max}$ **do**

6: $\Omega_l, \mathbf{E}_l \leftarrow$ new independent draw, $\underline{\mathbf{P}}_i = \mathbf{P}_i\mathbf{E}_l$, $\underline{\mathbf{S}}_i = \mathbf{S}_i\mathbf{E}_l$, $\underline{\mathbf{p}}_i = \mathrm{vec}(\underline{\mathbf{P}}_i)$

7: $\tau_l \leftarrow$ determine from τ_{l-1} and σ by root finding on the Pareto curve

8: $\mathbf{x}_l \leftarrow \begin{cases} \mathrm{argmin}_{\mathbf{x}} \sum_{i \in \Omega_l} \|\underline{\mathbf{p}}_i - \nabla\mathbf{F}_i[\mathbf{m}_0, \underline{\mathbf{S}}_i - \underline{\mathbf{P}}_i]\mathbf{C}^{\mathrm{H}}\mathbf{x}\|_2^2 \\ \text{subject to } \|\mathbf{x}\|_1 \leq \tau_l \end{cases}$ //warm start with $\mathbf{x}_{l-1}$, solved in k_l iterations

9: $k \leftarrow k + k_l, l \leftarrow l + 1$

10: **end while**

11: **Output:** Model perturbation estimate $\delta\mathbf{m} = \mathbf{C}^{\mathrm{H}}\mathbf{x}$

4 PERFORMANCE ANALYSIS

While the stylized example clearly showed qualitative improvements in the sparse recovery with rerandomization, we would like to study the behavior of the solver more quantitatively. In particular, we are interested in studying the solution path and the model error as a function of the number of wave-equation solves, as it can give us insights into the way that rerandomization works to speed up the convergence. We are also interested in the recovery error

as a function of the number of simultaneous sources, frequencies, and iterations, which will help us to choose the optimal inversion parameters for the proposed algorithm given a fixed computational budget.

4.1 Solution path and model error

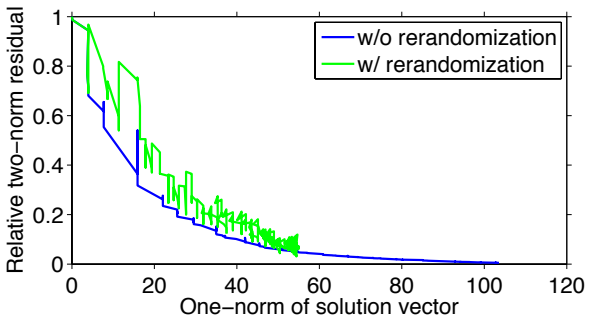
To understand the significant difference between optimizing with and without rerandomization (results in Figure 2(d) and 2(e)), we plot their respective solution path, i.e., the (relative) ℓ_2-norm of the data misfit as a function of the ℓ_1-norm of the solution vector. We also plot the decreases of the model errors (i.e., the relative ℓ_2-norm of the misfit between the model iterate and and true model perturbations) in the two cases as a function of the number of wave-equation solves.

Shown in Figure (3(a)), the solution path for the optimization with rerandomization behaves differently (expected, as we are solving a different problem with rerandomization) from that of the standard SPGℓ_1, where the ℓ_2-norm of the data residual decreased close to zero albeit slowly. Because the residual is larger for the case with rerandomization, one would normally expect a deterioration rather than an improvement in the image quality by using rerandomization. However, we can see in Figure (3(b)) that the model errors decay continuously with rerandomization that breaks down the correlation built up between the model iterate and the randomized subsets of monochromatic source experiments, but almost stall at an early stage of the optimization in the case of no rerandomization (i.e., no significant improvement by using more iterations).

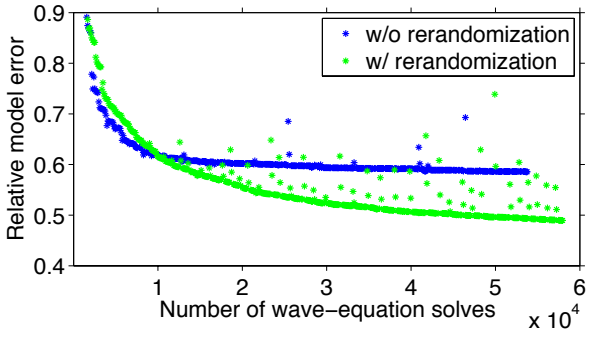
Because the artifacts in Figure 2(d) with wrongly identified support lie close to the nullspace of the compressive modelling operator (as the blue curve decreased close to zero in Figure 3(a)), they are difficult to remove even after extended number of iterations. However, by occasionally drawing new compressive modelling operators with independent subsets of frequencies and source experiments, these artifacts can be gradually removed. This observation was also made by Herrmann & Li (2012), where using rerandomization greatly promotes the convergence in terms of model error.

4.2 Computational considerations

As we use wave-equation solvers to carry out the SRME type multi-dimensional convolutions implicitly, which also enables us to reduce the simulation cost by subsampling the monochromatic source experiments, the costs of our imaging with multiple algorithm are dominated by the number of wave-equation solves. This raises the question of identifying the best inver-



(a)



(b)

Figure 3. Comparison between with (in green) and without (in blue) rerandomization. **(a)** Decrease of the relative data misfit. **(b)** Decrease of the relative model error.

sion strategy given a certain computational budget in terms of the number of wave-equation solves, which is proportional to the product of the number of monochromatic sources and the number of iterations, i.e., $n_f \times n_s \times n_{iter}$. In conventional inversion approaches, people usually work with all sources and frequencies, but can only afford a few iterations. With the proposed approach, we can work with significantly smaller data subsets and can therefore afford many more iterations because each iteration is now much cheaper. While both approaches are parallelizable over independent monochromatic source experiments in the data space or via domain decomposition in the model space, our approach has the advantage of partially removing the computer input / output bottleneck as the subsampling also reduces the data size.

Using the same synthetic model / data as in the previous section, we compare the performance of the proposed method with different combinations of randomized monochromatic source experiments and iterations. The signal-to-noise ratios (SNR) of the recoveries compared to the true model perturbation are shown in Table (1). From the table we can see that more iterations with cheaper per-iteration cost lead to better recoveries, which also offers flexibility in the design of parallel algorithms.

Another aspect is to understand the interplay between the number of simultaneous sources

No. of mono. sources.	15	30	60	120	240	480
Aspect ratio of $\nabla\mathbf{F}$	0.02	0.04	0.08	0.16	0.32	0.64
No. of iterations	610	305	152	76	39	20
SNR (dB)	6.5	6.2	5.8	5.1	4.2	3.4

Table 1. SNRs as a function of the number of monochromatic source experiments and the number of iterations.

and frequencies once we fix the number of monochromatic source experiments. We compare five different setups and show the results in Table (2). We can see that these combinations lead to almost equally good recoveries.

5 SYNTHETIC CASE STUDY

To evaluate the performance of the proposed method for a more realistic setup, we verify our method with a more complex Earth model. The model we use here is cropped from the Sigsbee 2B model that contains abundant sedimentary layers. The Sigsbee 2B model is designed to contain a thin high-velocity ocean bottom layer to generate strong surface-related multiples (The SMAART JV 2014). We reduced the thickness of the water column to allow for more orders of surface-related multiples. The modified true and background models are shown in Figure 4. The background model is obtained by smoothing the true model and is kinematically close to the true model. The grid spacing is 7.62m. There are 261 co-located sources and receivers with 22.86m spacing. We model the data using iWave with a free-surface boundary condition (Terentyev et al. 2014). We use a Ricker wavelet with 15Hz peak frequency, and record data for 8.184 seconds. We simulate both the pressure and particle velocity wavefields to separate the upgoing and downgoing wavefields (Wapenaar 1998). We then remove the direct arrivals by subtracting data modelled with the background model. Next, we extrapolate the upgoing wavefields to the surface level (see Verschuur et al. 1992 for more details). In the imaging procedure, we use our own frequency-domain modelling engine to

No. of sim. sources	2	3	6	10	15
No. of frequencies	15	10	5	3	2
SNR (dB)	6.2	6.2	6.2	6.3	6.7

Table 2. SNRs as a function of the number of simultaneous sources and frequencies. The number of monochromatic sources as well as the number of iterations are fixed.

mimic the discrepancy between the observed data and the modelled data in field seismic data processing (van Leeuwen & Herrmann 2012). For the same reason, we use forward modelling instead of linearized Born modelling to simulate the data. However, our proposed method is robust to this linearization error by virtue of using rerandomization (see Tu et al. 2013b for more details).

For comparison, we first computed the RTM image of data with multiples with all the data according to Equation (6). The result is shown in Figure 4(c). Again we see the coherent acausal artifacts caused by surface-related multiples, which are indicated by the arrow.

For the proposed method, we use 26 simultaneous sources from all 261 sequential sources (roughly 10× subsampling), 31 frequencies randomly chosen from all 311 discretized frequencies (roughly 10× subsampling), and run the inversion for 50 iterations. In this way the simulation cost remains comparable to a single RTM of all the data (e.g., Figure 4(c)). After the inversion is finished, we perform a simple additional curvelet thresholding to remove the remaining incoherent noise in the image (see Herrmann et al. 2008a for more details). We choose such a threshold level that the filtered-out noise does not contain noticeable coherent energy. The result is included in Figure 4(d). This thresholding procedure is also applied to the imaging results in the Discussion section. Again we obtain an image with minimal coherent acausal artifacts arising from the surface-related multiples by using the proposed inversion method. These weak remnant artifacts have the following sources: (i) we use different modelling engine for data simulation and inversion, (ii) the use of a pair of monopole sources to approximate the source-ghost effect is inexact, and (iii) the prediction of multiples is also inexact because of finite aperture. Despite these remnant artifacts, the resulting image is of higher spatial resolution compared to Figure 4(c), as we properly invert the demigration operator that includes the source wavelet.

To highlight the efficacy of the proposed method in reducing the coherent artifacts caused by the surface-related multiples, we run another inversion with the same setup, but ignore the presence of the surface-related multiples (i.e., the modelling operator only models primaries). The result is shown in Figure 4(e), where we can see strong coherent artifacts that mimic those in Figure 4(c).

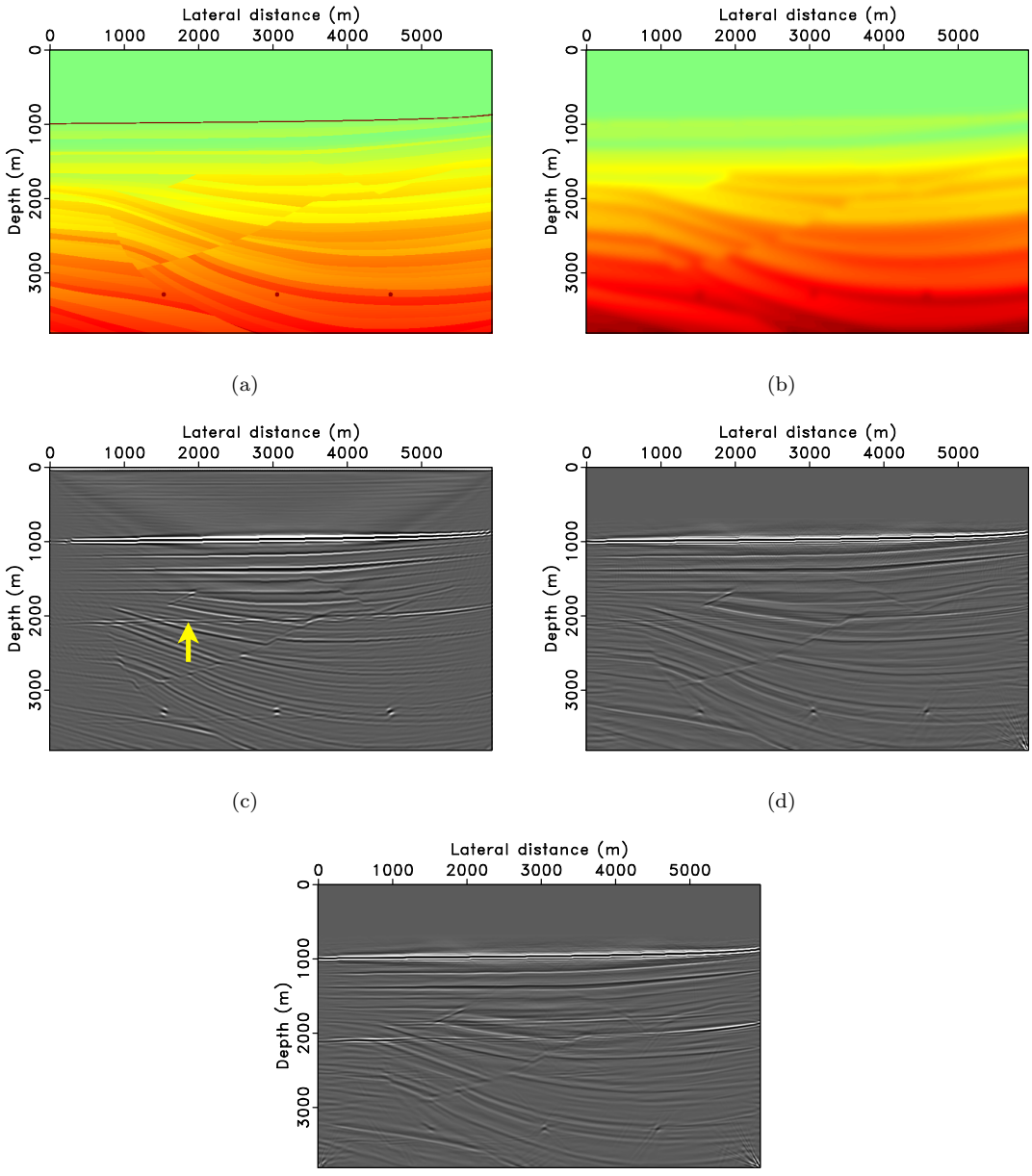


Figure 4. Examples using the cropped Sigsbee 2B model. **(a)** The true model. **(b)** The background model. **(c)** RTM image of data with multiples. **(d)** Fast inversion result using the proposed method after curvelet thresholding. **(e)** Fast inversion result without accounting for the multiples, after curvelet thresholding.

6 DISCUSSION

6.1 The source wavelet

We assumed that we have access to the true source wavelet. In practice, however, source estimation is never a trivial task because of the amplitude / phase ambiguity in blind decon-

volution (Ulrych & Sacchi 2005, chapter 6). The EPSI algorithm successfully eliminates these ambiguities by using information from surface-related multiples (van Groenestijn & Verschuur 2009a; Lin & Herrmann 2013). However, it is computationally expensive because it uses an iterative algorithm with dense matrix-matrix multiplies.

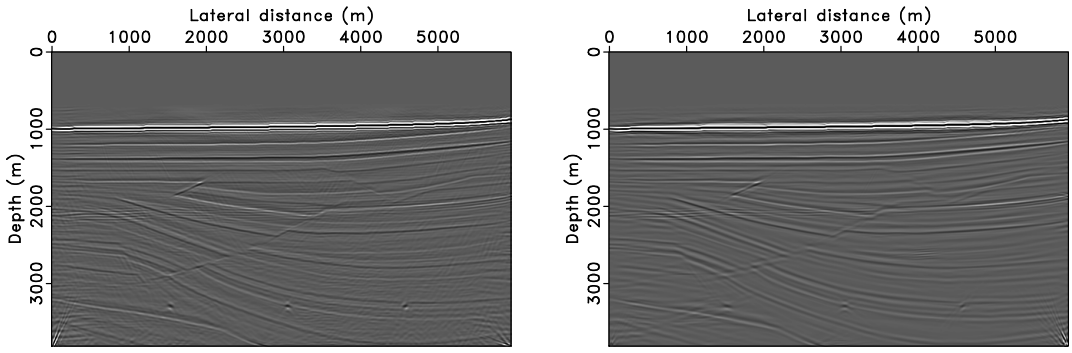
Alternatively, source signatures can be estimated during the imaging itself using an approach known as variable projection (Golub & Pereyra 2003; Aravkin & van Leeuwen 2012). Recent work by Aravkin et al. (2013) and Tu et al. (2013a) successfully applied this alternating optimization method to recover the image as well as an estimate for the source function. Figure 5(a) contains an example of this method obtained using the same inversion parameters as in the previous section, i.e., we use 26 simultaneous sources, 31 randomized frequencies, and run the inversion for 50 iterations. Compared to Figure 4(d), a 50% computational overhead is incurred by the variable-projection step within each iteration, as the primary and multiple wavefields have to be modelled separately to update the source estimate (see Tu et al. 2013a for more details). This extra computation, which is about half of a single RTM of all the data, is still insignificant compared with a conventional least-squares migration of all the data. We plan to report this source-estimation method in more details in another paper.

6.2 Imaging with multiples only

Our proposed method can also be used in cases where a reliable primary / multiple separation is available, e.g. after a conventional multiple prediction step. In that situation, we can image multiples separately and exploit their extra illumination properties without explicit knowledge of the source wavelet. (However, multiple predictions by the SRME-type multi-dimensional convolutions typically need to apply the inverse of the wavelet.) As before, we obtain these images by including an areal source, which now consists of the downgoing receiver wavefield only. Compared to earlier work (e.g. Guitton 2002; Muijs et al. 2007; Whitmore et al. 2010; Liu et al. 2011), our approach leads to a true inversion result while others merely apply the deconvolutional imaging condition or the cross-correlation imaging condition, which is known to lead to undesirable acausal artifacts. We use the same inversion parameters as for Figure 5(a), and the result is included in Figure 5(b).

6.3 Sensitivity to the background velocity model

Our method hinges, as all other wave-equation based imaging procedures, on the availability of a smooth background velocity model that accurately captures the kinematics of reflected waves. We have also seen that incorporating surface-related multiples in the formulation calls



(a) (b)

Figure 5. Dealing with the unknown source signature. Both images are after curvelet thresholding. **(a)** Fast imaging of the total upgoing wavefield with source estimation. **(b)** Fast imaging of multiples only.

for an iterative sparsity-promoting inversion procedure and the question arises how sensitive this more elaborate imaging procedure is to errors in the background velocity model. This question is particularly pertinent because migration velocity analysis tools also often suffer from surface-related multiples. Moreover, iterative inversion procedures such as (sparsity-promoting) least-squares migration may be more sensitive to velocity-model errors than conventional RTMs and it is not clear how the inclusion of surface-related multiples plays into this.

To address these concerns we conducted an experiment where we deliberately introduce errors in the velocity model that cumulate with depth (see Figure 6(a) and 6(b)). The results of this exercise are included in Figure 6(c) to 6(f) and show that for this example least-squares imaging with multiples is relatively well behaved with imaging results that degrade gracefully. As expected, imaging results with curvelet-domain sparsity promotion are more coherent by virtue of that fact that curvelets allow for small kinematic errors as reported in the literature (Herrmann et al. 2008c). The results without curvelet-domain sparsity were obtained by putting the sparsity-contraint parameter τ in each $\mathbf{LS}_\tau$ subproblem to infinity (it is essentially the block Kaczmarz method (see e.g. Needell & Tropp 2014)), and eliminating the use of the curvelet transform. While more work is needed to fully understand the influence of the background velocity model on our iterative inversion algorithm, these results show that our method is relatively well-behaved w.r.t. possible errors in the background velocity model.

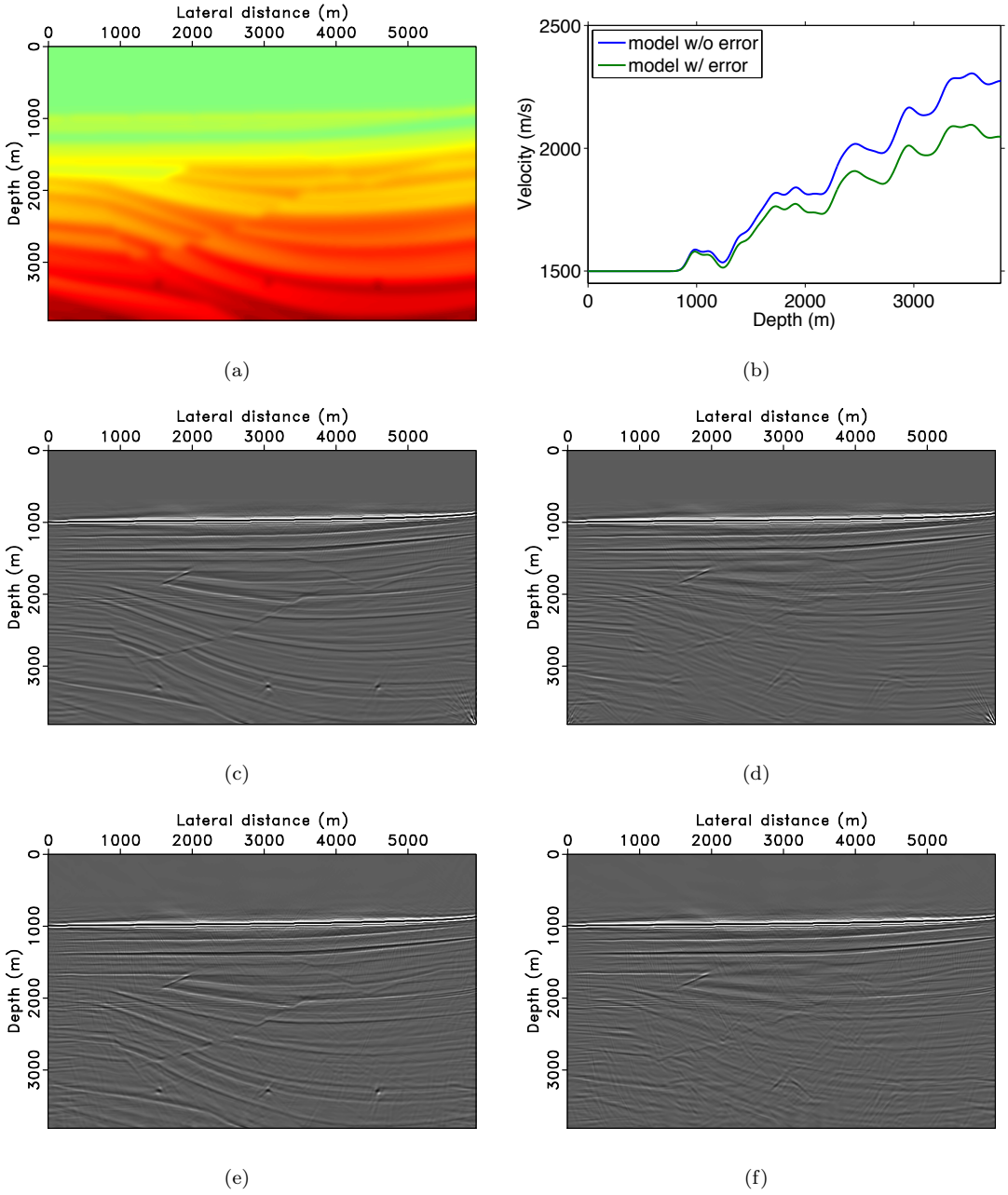


Figure 6. Example with an inaccurate background model. All images were after curvelet thresholding to remove remaining incoherent artifacts. **(a)** The wrong background velocity model. **(b)** One vertical trace of **(a)**. **(c)** and **(d)** Image using the proposed method, without and with the velocity error. **(e)** and **(f)** Image without curvelet-domain sparsity promotion, without and with the velocity error.

6.4 Other applications

In essence, the proposed method entails multi-dimensional deconvolutions of wavefields combined with a linearized imaging procedure. While we specialized this approach to image with surface-related multiples, there is nothing that prevents the application of our proposed me-

thod to other areas as long as the receivers are sampled adequately. For instance, the method could relatively easily be adapted to imaging with receiver functions. In that case, the observed converted shear waves are deconvolved with the pressure waves to image the pressure-to-shear transmission coefficients (Ryberg & Weber 2000; Shang et al. 2012). Another example would be the deconvolutional interferometric imaging as recently proposed by van der Neut & Herrmann (2013).

7 CONCLUSIONS

We proposed a new and computationally efficient method to image seismic data with surface-related multiples. Our contributions are threefold. First, our formulation combines the wave equation with surface-related-multiple prediction by using a generalized areal-source term that contains the downgoing receiver wavefield. By incorporating this source term, we arrive at a linearized formulation that predicts, and therefore correctly images, data with surface-related multiples. Instead of predicting the multiples via explicit dense matrix-matrix multiplications as performed in surface-related multiple elimination, which is prohibitively expensive for large seismic data sets, our method predicts multiples implicitly by solving wave equations for the areal-source wavefields, and therefore eliminate the cost of the matrix product altogether. Second, our combination of multiple prediction and linearized forward modelling with the Born approximation allows us to come up with a proper curvelet-domain sparsity-promoting inversion framework where primaries and surface-related multiples are jointly inverted to produce a high-resolution true-amplitude image of the subsurface medium perturbations. Compared to non-iterative correlation-based imaging methods with multiples, this approach leads to high-resolution images with minimal artifacts. Third, we overcome prohibitive computational cost of sparsity-promoting inversion by choosing to work with small independent randomized subsets of data. This subsampling and rerandomization strategy leads to a simulation cost that is comparable to the cost of a single reverse-time migration with all the data, which makes our method feasible in practice. More importantly, our method produces by virtue of the inversion high-fidelity true-amplitude images, an accomplishment that is difficult, if at all possible, to achieve with conventional methods. Because of its superior performance on complex synthetics, in conjunction with a reduced sensitivity to modelling errors and errors in the velocity model, we expect our method to be suitable to field data.

8 ACKNOWLEDGEMENTS

The authors would like to thank Dr. Eric Verschuur and Dr. Joost van der Neut for the beneficial discussions. We thank Dr. Tristan van Leeuwen for the frequency-domain modelling code that we use in our inversion algorithm. We also would like to thank the authors of SPGℓ_1, Curvlab and iWave, as well as the SMAART JV for providing the Sigsbee 2B model.

This work was financially supported in part by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (RGPIN 261641-06) and the Collaborative Research and Development Grant DNOISE II (CDRP J 375142-08). This research was carried out as part of the SINBAD II project with support from the following organizations: BG Group, BGP, BP, Chevron, ConocoPhillips, CGG, ION GXT, Petrobras, PGS, Statoil, Total SA, WesternGeco, Woodside. Ning Tu was also partially financially supported by the Chinese Scholarship Council.

REFERENCES

Aravkin, A. Y. & van Leeuwen, T., 2012. Estimating nuisance parameters in inverse problems, *Inverse Problems*, **28**(11).

Aravkin, A. Y., van Leeuwen, T., & Tu, N., 2013. Sparse seismic imaging using variable projection, in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, pp. 2065–2069.

Baysal, E., Kosloff, D., & Sherwood, J., 1983. Reverse time migration, *GEOPHYSICS*, **48**(11), 1514–1524.

Berkhout, A. J., 1993. Migration of multiple reflections, in *SEG Technical Program Expanded Abstracts*, vol. 12, pp. 1022–1025, SEG.

Berkhout, A. J. & Verschuur, D. J., 2006. Imaging of multiple reflections, *Geophysics*, **71**(4), SI209–SI220.

Brown, M. & Guitton, A., 2005. Least-squares joint imaging of multiples and primaries, *GEOPHYSICS*, **70**(5), S79–S89.

Candès, E. J. & Donoho, D. L., 2004. New tight frames of curvelets and optimal representations of objects with piecewise c2 singularities, *Communications on Pure and Applied Mathematics*, **57**(2), 219–266.

Candès, E. J., Demanet, L., Donoho, D. L., & Ying, L., 2006a. Fast discrete curvelet transforms, *Multiscale Modeling and Simulation*, **5**(3), 861–899.

Candès, E. J., Romberg, J. K., & Tao, T., 2006b. Stable signal recovery from incomplete and inaccurate measurements, *Communications on Pure and Applied Mathematics*, **59**(8), 1207–1223.

Chen, S. S., Donoho, D. L., & Saunders, M. A., 2001. Atomic decomposition by basis pursuit, *SIAM Review*, **43**(1), 129.

Donoho, D., 2006. Compressed sensing, *Information Theory, IEEE Transactions on*, **52**(4), 1289 –1306.

Donoho, D., Maleki, A., & Montanari, A., 2009. Message passing algorithms for compressed sensing, *Proceedings of the National Academy of Sciences*, **106**(45), 18914–18919.

Dragoset, B., Verschuur, E., Moore, I., & Bisley, R., 2010. A perspective on 3d surface-related multiple elimination, *Geophysics*, **75**(5), 75A245–75A261.

Dragoset, W. H. & Jeričević, Ž., 1998. Some remarks on surface multiple attenuation, *Geophysics*, **63**(2), 772–789.

Fokkema, J. T. & van den Berg, P. M., 1993. *Seismic applications of acoustic reciprocity*, vol. 552, Elsevier.

Foster, D. & Mosher, C., 1992. Suppression of multiple reflections using the radon transform, *Geophysics*, **57**(3), 386–395.

Golub, G. & Pereyra, V., 2003. Separable nonlinear least squares: the variable projection method and its applications, *Inverse problems*, **19**(2), R1–R26.

Guitton, A., 2002. Shot-profile migration of multiple reflections, in *SEG Technical Program Expanded Abstracts*, vol. 21, pp. 1296–1299.

Guitton, A. & Verschuur, D. J., 2004. Adaptive subtraction of multiples using the ℓ_1-norm, *Geophysical prospecting*, **52**(1), 27–38.

Hampson, D., 1986. Inverse velocity stacking for multiple elimination, in *SEG Technical Program Expanded Abstracts*, pp. 422–424.

He, R. & Schuster, G., 2003. Least-squares migration of both primaries and multiples, in *SEG Technical Program Expanded Abstracts*, vol. 22, pp. 1035–1038, SEG.

He, R., Hornby, B., & Schuster, G. T., 2007. 3d wave-equation interferometric migration of vsp free-surface multiples, *Geophysics*, **72**(5), S195–S203.

Herrmann, F. J., 2012. Pass on the message: recent insights in large-scale sparse recovery, in *74th EAGE Conference & Exhibition*, EAGE.

Herrmann, F. J. & Hennenfent, G., 2008. Non-parametric seismic data recovery with curvelet frames, *Geophysical Journal International*, **173**, 233–248.

Herrmann, F. J. & Li, X., 2012. Efficient least-squares imaging with sparsity promotion and compressive sensing, *Geophysical Prospecting*, **60**(4), 696–712.

Herrmann, F. J., Moghaddam, P. P., & Stolk, C., 2008a. Sparsity- and continuity-promoting seismic image recovery with curvelet frames, *Applied and Computational Harmonic Analysis*, **24**(2), 150–173.

Herrmann, F. J., Wang, D., Hennenfent, G., & Moghaddam, P. P., 2008b. Curvelet-based seismic data processing: a multiscale and nonlinear approach, *Geophysics*, **73**(1), A1–A5.

Herrmann, F. J., Wang, D., & Verschuur, D. J., 2008c. Adaptive curvelet-domain primary-multiple separation, *Geophysics*, **73**(3), A17–A21.

Herrmann, F. J., Erlangga, Y. A., & Lin, T. T., 2009. Compressive simultaneous full-waveform simulation, *Geophysics*, **74**, A35.

Jumah, B. & Herrmann, F. J., 2014. Dimensionality-reduced estimation of primaries by sparse inversion, *Geophysical Prospecting*, **62**(5), 972–993.

Lin, T. T. & Herrmann, F. J., 2013. Robust estimation of primaries by sparse inversion via one-norm minimization, *Geophysics*, **78**(3), R133–R150.

Lin, T. T., Tu, N., & Herrmann, F. J., 2010. Sparsity-promoting migration from surface-related multiples, in *SEG Technical Program Expanded Abstracts*, vol. 29, pp. 3333–3337, SEG.

Liu, Y., Chang, X., Jin, D., He, R., Sun, H., & Zheng, Y., 2011. Reverse time migration of multiples for subsalt imaging, *Geophysics*, **76**(5), WB209–WB216.

Long, A. S., Lu, S., Whitmore, D., LeGleut, H., Jones, R., Chemingui, N., & Farouki, M., 2013. Mitigation of the 3d cross-line acquisition footprint using separated wavefield imaging of dual-sensor streamer seismic, in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*, EAGE.

Lu, G., Ursin, B., & Lutro, J., 1999. Model-based removal of water-layer multiple reflections, *Geophysics*, **64**(6), 1816–1827.

Lu, S., Whitmore, N. D., & Valenciano, A. A., 2011. Imaging of primaries and multiples with 3d seam synthetic, in *SEG Technical Program Expanded Abstracts*, vol. 30, pp. 3217–3221.

Mallat, S. G., 2009. *A wavelet tour of signal processing*, Academic Press.

Montanari, A., 2010. Graphical models concepts in compressed sensing, arXiv:1011.4328v3 [cs.IT].

Muijs, R., Robertsson, J. O. A., & Holliger, K., 2007. Prestack depth migration of primary and surface-related multiple reflections: Part i—imaging, *Geophysics*, **72**, S59–S69.

Needell, D. & Tropp, J. A., 2014. Paved with good intentions: Analysis of a randomized block kaczmarz method, *Linear Algebra and its Applications*, **441**(0), 199 – 221, Special Issue on Sparse Approximate Solution of Linear Systems.

Needell, D., Srebro, N., & Ward, R., 2014. Stochastic gradient descent and the randomized kaczmarz algorithm, arXiv:1310.5715v2 [math.NA].

Neelamani, R., Baumstein, A., & Ross, W., 2010. Adaptive subtraction using complex-valued curvelet transforms, *Geophysics*, **75**(4), V51–V60.

Nemeth, T., Wu, C., & Schuster, G. T., 1999. Least-squares migration of incomplete reflection data, *Geophysics*, **64**(1), 208–221.

Nemeth, T., Sun, H., & Schuster, G. T., 2000. Separation of signal and coherent noise by migration filtering, *Geophysics*, **65**(2), 574–583.

Peacock, K. L. & Treitel, S., 1969. Predictive deconvolution: theory and practice, *Geophysics*, **34**(2), 155–169.

Poole, T., Curtis, A., Robertsson, J., & van Manen, D.-J., 2010. Deconvolution imaging conditions and cross-talk suppression, *Geophysics*, **75**(6), W1–W12.

Prucha, M. & Biondi, B., 2002. Subsalt event regularization with steering filters, in *SEG Technical Program Expanded Abstracts*, pp. 1176–1179.

Reiter, E. C., Toksoz, M. N., Keho, T. H., & Purdy, G. M., 1991. Imaging with deep-water multiples, *Geophysics*, **56**(7), 1081–1086.

Romero, L., Ghiglia, D., Ober, C., & Morton, S., 2000. Phase encoding of shot records in prestack migration, *GEOPHYSICS*, **65**(2), 426–436.

Ryberg, T. & Weber, M., 2000. Receiver function arrays: a reflection seismic approach, *Geophysical Journal International*, **141**(1), 1–11.

Sava, P. & Guitton, A., 2005. Multiple attenuation in the image space, *Geophysics*, **70**(1), V10–V20.

Schuster, G. T., Yu, J., Sheng, J., & Rickett, J., 2004. Interferometric/daylight seismic imaging, *Geophysical Journal International*, **157**(2), 838–852.

Shan, G., 2003. Source-receiver migration of multiple reflections, in *SEG Technical Program Expanded Abstracts*, vol. 22, pp. 1008–1011, SEG.

Shang, X., de Hoop, M. V., & van der Hilst, R. D., 2012. Beyond receiver functions: Passive source reverse time migration and inverse scattering of converted waves, *Geophysical Research Letters*, **39**(15), n/a–n/a.

Strohmer, T. & Vershynin, R., 2009. A randomized kaczmarz algorithm with exponential convergence, *Journal of Fourier Analysis and Applications*, **15**(2), 262–278.

Terentyev, I. S., Vdovina, T., Symes, W. W., Wang, X., & Sun., D., 2014. iWave: a framework for wave simulation, `http://www.trip.caam.rice.edu/software/iwave/doc/html/index.html`, Online; accessed 16-January-2014.

The SMAART JV, 2014. Sigsbee2B FS & NFS 2D synthetic datasets, `http://www.delphi.tudelft.nl/SMAART/sigsbee2b.htm`, Online; accessed 24-February-2014.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society Series B Methodological*, **58**(1), 267–288.

Tu, N. & Herrmann, F. J., 2012a. Least-squares migration of full wavefield with source encoding, in *EAGE*.

Tu, N. & Herrmann, F. J., 2012b. Imaging with multiples accelerated by message passing, in *SEG Technical Program Expanded Abstracts*, pp. 1–6, Chapter 682, SEG.

Tu, N., Aravkin, A. Y., van Leeuwen, T., & Herrmann, F. J., 2013a. Fast least-squares migration with multiples and source estimation, in *75th EAGE Conference & Exhibition incorporating SPE EUROPEC 2013*, EAGE.

Tu, N., Li, X., & Herrmann, F. J., 2013b. Controlling linearization errors in ℓ_1 regularized inversion by rerandomization, in *SEG Technical Program Expanded Abstracts*, pp. 4640–4644, SEG.

Tu, N., van Leeuwen, T., & Herrmann, F. J., 2013c. Limitations of the deconvolutional imaging

condition for two-way propagators, in *SEG Technical Program Expanded Abstracts 2013*, pp. 3916–3920.

Ulrych, T. J. & Sacchi, M. D., 2005. *Information-Based Inversion and Processing with Applications*, Elsevier Science.

van den Berg, E. & Friedlander, M. P., 2008. Probing the pareto frontier for basis pursuit solutions, *SIAM Journal on Scientific Computing*, **31**(2), 890–912.

van der Neut, J. & Herrmann, F. J., 2013. Interferometric redatuming by sparse inversion, *Geophysical Journal International*, **192**(2), 666–670.

van Groenestijn, G. J. A. & Verschuur, D. J., 2009a. Estimation of primaries and near-offset reconstruction by sparse inversion: Marine data applications, *Geophysics*, **74**(6), R119–R128.

van Groenestijn, G. J. A. & Verschuur, D. J., 2009b. Estimating primaries by sparse inversion and application to near-offset data reconstruction, *Geophysics*, **74**(3), A23–A28.

van Leeuwen, T. & Herrmann, F. J., 2012. A parallel, object-oriented framework for frequency-domain wavefield imaging and inversion., Tech. Rep. TR-2012-03, Department of Earth and Ocean Sciences, University of British Columbia, Vancouver.

Verschuur, D. J., 1991. *Surface-related multiple elimination, an inverse approach*, Ph.D. thesis, Delft University of Technology.

Verschuur, D. J., 2011. Seismic migration of blended shot records with surface-related multiple scattering, *Geophysics*, **76**(1), A7–A13.

Verschuur, D. J., Berkhout, A. J., & Wapenaar, C. P. A., 1992. Adaptive surface-related multiple elimination, *Geophysics*, **57**(9), 1166–1177.

Wang, D., Saab, R., Yilmaz, O., & Herrmann, F. J., 2008. Bayesian wavefield separation by transform-domain sparsity promotion, *Geophysics*, **73**(5), 1–6.

Wapenaar, K., 1998. Reciprocity properties of one-way propagators, *Geophysics*, **63**(4), 1795–1798.

Weglein, A., Gasparotto, F., Carvalho, P., & Stolt, R., 1997. An inversescattering series method for attenuating multiples in seismic reflection data, *GEOPHYSICS*, **62**(6), 1975–1989.

Whitmore, N. D., Valenciano, A. A., & Sollner, W., 2010. Imaging of primaries and multiples using a dual-sensor towed streamer, in *SEG Technical Program Expanded Abstracts*, vol. 29, pp. 3187–3192.

Wiggins, J. W., 1988. Attenuation of complex water-bottom multiples by wave-equation-based prediction and subtraction, *Geophysics*, **53**(12), 1527–1539.

Wong, M., Biondi, B., & Ronen, S., 2012. Imaging with multiples using linearized full-wave inversion, in *SEG Technical Program Expanded Abstracts*, pp. 1–5, Chapter 706, SEG.

Wong, M., Biondi, B., & Ronen, S., 2014. Imaging with multiples using least-squares reverse time migration, *The Leading Edge*, **33**(9), 970–976.

APPENDIX A: SOLVING THE BPDN PROBLEM USING SPGℓ_1

Using canonical linear algebra notations, the BPDN formulation to invert a linear system of equations $\mathbf{Ax} = \mathbf{b}$ with sparsity promoting using the ℓ_1-norm follows

$$\mathbf{BP}_\sigma : \quad \min_{\mathbf{x}} \|\mathbf{x}\|_1 \quad \text{subject to} \quad \|\mathbf{b} - \mathbf{Ax}\|_2 \leq \sigma, \tag{A.1}$$

where τ is an estimate for the noise level in the data.

To efficiently solve the above $\mathbf{BP}_\sigma$ problem, SPGℓ_1 leverages the Spectrum Projected Gradient (SPG) method that efficiently solves the ℓ_1-constrained least-squares problem, aka the LASSO problem (van den Berg & Friedlander 2008):

$$\mathbf{LS}_\tau : \quad \min_{\mathbf{x}} \|\mathbf{b} - \mathbf{Ax}\|_2 \quad \text{subject to.} \quad \|\mathbf{x}\|_1 \leq \tau, \tag{A.2}$$

where τ is a sparsity constraint on the solution vector $\mathbf{x}$.

van den Berg & Friedlander (2008) found that problem $\mathbf{BP}_\sigma$ and $\mathbf{LS}_\tau$ share the same solution for certain choices of σ and τ, which are related by the Pareto trade-off curve between the ℓ_2-norm of the data residual (i.e., the objective function of problem $\mathbf{LS}_\tau$) and the ℓ_1-norm of the solution vector (i.e., the objective function of problem $\mathbf{BP}_\sigma$). As the Pareto curve is continuously differentiable, SPGℓ_1 uses the Newton's method on the Pareto curve to determine the optimal value of τ given an estimate of σ. As a result, SPGℓ_1 effectively solves a series of warm-started LASSO subproblems for gradually increasing τ's. By "warm-started", we mean that the solution for the k^{th} subproblem with $\tau^k < \tau^{k+1}$ serves as the initial guess for the next subproblem. Solving each subproblem $\mathbf{LS}_\tau$ itself using the SPG method involves several gradient updates. In each update, we need to evaluate the (expensive) action of $\mathbf{A}$ and $\mathbf{A}^{\mathrm{H}}$, perform a line search, followed by a (cheap) ℓ_1 projection that promotes sparsity.

APPENDIX B: SUBSAMPLING NOISE AND APPROXIMATE MESSAGE PASSING

In simple terms, the key idea of compressive sensing is to work with a flat sampling matrix $\mathbf{A}$ that creates noisy interferences, i.e., the action of $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ on a sparse vector yield a "noisy" sparse vector where the "noise" actually is incoherent crosstalks with a noise level that depends on degree of subsampling (aspect ratio of $\mathbf{A}$, defined as the ratio of the number of rows to the number of columns) and the sparsity of $\mathbf{x}$. In principle, this interference noise should be easier to remove than coherent noise. As explained by Donoho et al. (2009) and Herrmann (2012), this statement holds true for the first iteration where some noises are removed by projection onto the ℓ_1 ball (which corresponds to a single soft thresholding with the appropriate threshold

level). However, later iterations suffer from the correlation built up between the model iterate and the sampling matrix $\mathbf{A}$, which makes the denoising step less efficient and may result in coherent artifacts.

In the context of our compressive imaging problem, this means that when we reduce the number of simultaneous sources and frequencies, which corresponds to a reduced aspect ratio of $\mathbf{A}$, we may get more coherent artifacts in the image that cannot be removed by simply using more iterations.

Recent insights into approximate message passing finds that with a correction term in the iterative soft thresholding, this denoising step can continuously bring down the model errors in later iterations (Donoho et al. 2009; Montanari 2010). While this approach relies on careful analysis, we found that rerandomization of source experiments has statistically the same effect (Herrmann 2012) in breaking down this correlation buildup, and fits well in the optimization framework of the proposed method.