Geophysical Journal International

Geophys. J. Int. (2016) **205,** 1492–1511
Advance Access publication 2016 March 14
GJI Seismology

doi: 10.1093/gji/ggw097

Source estimation with surface-related multiples—fast ambiguity-resolved seismic imaging

Ning Tu,[1,2] Aleksandr Aravkin,[1,3] Tristan van Leeuwen,[1,4] Tim Lin[1,5] and Felix J. Herrmann[1]

[1]*Earth, Ocean and Atmospherical Sciences, University of British Columbia, Vancouver, Canada*
[2]*Ocean and Earth Sciences, Tongji University, Shanghai, China. Email: ntu@tongji.edu.cn*
[3]*Department of Applied Mathematics, University of Washington, Seattle, WA, USA*
[4]*Mathematical Institute, Utrecht University, P.O. Box* 80010, 3508 *TA Utrecht, The Netherlands*
[5]*Sub-Salt Solutions Ltd,* 52 *Princes Gate, Exhibition Road, London SW*7 2*PG, United Kingdom*

Accepted 2016 March 7. Received 2016 February 21; in original form 2015 May 14

SUMMARY

We address the problem of obtaining a reliable seismic image without prior knowledge of the source wavelet, especially from data that contain strong surface-related multiples. Conventional reverse-time migration requires prior knowledge of the source wavelet, which is either technically or computationally challenging to accurately determine; inaccurate estimates of the source wavelet can result in seriously degraded reverse-time migrated images, and therefore wrong geological interpretations. To solve this problem, we present a 'wavelet-free' imaging procedure that simultaneously inverts for the source wavelet and the seismic image, by tightly integrating source estimation into a fast least-squares imaging framework, namely *compressive imaging*, given a reasonably accurate background velocity model. However, this joint inversion problem is difficult to solve as it is plagued with local minima and the ambiguity with respect to amplitude scalings because of the multiplicative, and therefore nonlinear, appearance of the source wavelet in the otherwise linear formalism. We have found a way to solve this nonlinear joint-inversion problem using a technique called variable projection, and a way to overcome the scaling ambiguity by including surface-related multiples in our imaging procedure following recent developments in surface-related multiple prediction by sparse inversion. As a result, we obtain without prior knowledge of the source wavelet high-resolution seismic images, comparable in quality to images obtained assuming the true source wavelet is known. By leveraging the computationally efficient compressive-imaging methodology, these results are obtained at affordable computational costs compared with conventional processing work flows that include surface-related multiple removal and reverse-time migration.

Key words: Numerical solutions; Inverse theory; Wave propagation.

1 INTRODUCTION

Conventional reverse-time migration (RTM; Baysal *et al.* 1983) requires the knowledge of the source wavelet as prior information, which is used during the forward propagation of the source wavefield. Unfortunately, for field seismic data, this knowledge is either technically (e.g. relying on statistical assumptions, Ulrych *et al.* 1995) or computationally (e.g. as used in full-waveform inversion, Virieux & Operto 2009) challenging to accurately determine during the pre-processing procedure. Inaccurate estimates of the source wavelet will introduce errors to the forward propagated source wavefield, and later to the seismic image during the cross-correlation of the source and receiver wavefields. For example, wrong estimates in the phase of the source wavelet can result in misplaced structures in the seismic image, and therefore lead to wrong geological interpretations.

To eliminate the dependence of seismic imaging on the knowledge of the source wavelet, we are motivated to incorporate source estimation into seismic imaging with minimal computational overhead. We achieve this objective via the joint inversion of the seismic image and the source wavelet, by extending the least-squares-migration formalism (Nemeth *et al.* 1999). Because the forward modelling of the seismic data is linear with respect to the source wavelet, this type of joint inversion problem is known as the separate least-squares problem (Golub & Pereyra 1973; Kaufman 1975; Golub & Pereyra 2003; Aravkin & van Leeuwen 2012). To reduce the excessive simulation cost of conventional least-squares migrations, we leverage curvelet-domain sparsity of the seismic image (Herrmann *et al.* 2008a), and

© The Authors 2016. Published by Oxford University Press on behalf of The Royal Astronomical Society.

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

tightly integrate source estimation into a computationally efficient least-squares imaging formalism (Herrmann & Li 2012), also known as the *compressive imaging* method. As a result, we obtain high-resolution least-squares migrated images with simulation costs (in terms of wave-equation solves, not the overall computing time) that roughly equal conventional RTM, whereas the computational overhead introduced by source estimation is negligible compared to the wave-equation solves.

Although a separable least-squares problem can usually be effectively solved using variable projection (Golub & Pereyra 2003), applying variable projection to the proposed joint inversion procedure is complicated by the following two issues. First, the compressive imaging formalism adopts a sparsity-promoting objective function, which is different from typical separable least-squares problems (Golub & Pereyra 2003). We follow van den Berg & Friedlander (2008) and Aravkin *et al.* (2013), and derive an alternative problem formulation that turns the sparsity-promoting objective into a sparse constraint. Second, there is an ambiguity in the amplitude scaling between the model perturbations and the estimated source wavelet, which cannot be resolved using variable projection alone, as in any blind-deconvolution type of problem (Stockham *et al.* 1975). Inspired by recent developments in Primary Estimation by Sparse Inversion (EPSI; van Groenestijn & Verschuur 2009; Lin & Herrmann 2013), we propose to resolve the ambiguity by incorporating surface-related multiples in the inversion. In this way we leverage the self-consistency between the primary events and their corresponding surface-related multiples, and obtain a properly scaled source wavelet that leads to the separation of primaries and multiples. Within the scope of this paper, we only consider surface-related multiples (therefore internal multiples are treated as part of the primaries, as in EPSI), and refer to Tu & Herrmann (2015a) for more discussions on the possibility of imaging internal multiples using a nonlinear extension of the proposed approach. As any other imaging algorithm, our proposed method relies on the knowledge of a reasonably accurate background model.

1.1 Related work and our contribution

The variable projection method has found its applications in a variety of geophysical problems in recent years, for example, in characterization of *P–S* wave conversion (Fomel *et al.* 2003), in velocity model building (van Leeuwen & Mulder 2009; Peters & Herrmann 2014; Peters *et al.* 2014), in impedance reconstruction (Mtivier 2011; Mtivier *et al.* 2011) and notably in source estimation during full-waveform inversion (Aravkin *et al.* 2012; Li *et al.* 2013; Rickett 2013). In all these applications, an objective function that measures the data misfit is used for the variable projection technique to be applicable. The misfit is usually measured using the ℓ_2-norm (i.e. a least-squares formulation), but other differentiable misfit functions can also be used (Aravkin *et al.* 2012). In this paper, however, we minimize the ℓ_1-norm of the solution vector to promote sparsity, with the least-squares data-fitting term acting as a constraint. To the authors' knowledge, this work is the first instance of applying the variable projection technique to an optimization problem with a sparsity-promoting objective.

Regarding the use of multiples in source estimation, successful data space applications have been demonstrated in the literature of Surface-Related Multiple Elimination (SRME; Verschuur *et al.* 1992) and EPSI (van Groenestijn & Verschuur 2009; Lin & Herrmann 2013; Esser *et al.* 2015). Based on the same physical principles as described by the SRME relation (Berkhout 1993; Guitton 2002; Muijs *et al.* 2007; Whitmore *et al.* 2010; Liu *et al.* 2011; Verschuur 2011; Wong *et al.* 2014; Zhang & Schuster 2014; Davydenko *et al.* 2015; Tu & Herrmann 2015a), we incorporate surface-related multiples in sparsity-promoting seismic imaging including source estimation, and demonstrate why incorporation of the surface-related multiples mitigates the scaling ambiguity that plagues source estimation in seismic imaging. By optimizing in the model space rather than in the data space, we gain the following benefits: (i) we have access to the physical locations of the entire subsurface model, and can therefore overcome some limitations that SRME/EPSI imposes on data acquisition, such as (approximate) source/receiver co-location; (ii) the unknown model parameters are of much smaller dimensions compared to seismic data, which enables us to subsample the source wavefields to relax the dense data-acquisition requirement or the computation cost (Tu & Herrmann 2015a). Aside from mitigating the scaling ambiguity, multiples also provide extra illumination coverage that complements primaries, especially along the cross-line direction in 3-D seismic surveys (Long *et al.* 2013; Lu *et al.* 2014). However, properly imaging the surface-related multiples that contain many different orders of surface reflections calls for an inversion approach to suppress spurious artefacts from the multiples (Verschuur 2011; Wong *et al.* 2014; Zhang & Schuster 2014; Tu & Herrmann 2015a), which can be computationally challenging. Tu & Herrmann (2015a) extended the compressive imaging approach by Herrmann & Li (2012) to account for surface-related multiples in a computationally efficient way, assuming prior knowledge of a properly scaled source wavelet. Verschuur (2011) also showed that simultaneous imaging of the primary and multiple wavefields requires the exact knowledge of the source wavelet. Davydenko *et al.* (2015) proposed a two-step approach to first image the multiples and then the primaries by source estimation. In this work, we are motivated to obviate the need for the exact knowledge of source wavelet, and jointly image the primary and multiple wavefields, by integrating source estimation into the method by Tu & Herrmann (2015a) using variable projection (Golub & Pereyra 2003; Aravkin & van Leeuwen 2012).

1.2 Paper outline

The paper is organized as follows. First, we formulate the joint sparsity-promoting source- and image-estimation problem, followed by a technical discussion on how to solve ℓ_1-norm minimization problems with source estimation via variable projection. Next, we explain why including surface-related multiples into the formulation helps mitigate the scaling ambiguity during imaging and source estimation. We conclude by demonstrating the advantages of the proposed method over conventional RTM, using carefully elaborated synthetic examples.

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

2 PROBLEM FORMULATION AND METHOD

As we propose the joint inversion approach in the compressive-imaging framework, we will first briefly review the compressive-imaging formalism and formulate the source estimation problem. Afterwards, we will explain how to solve the problem using variable projection. Within the scope of this paper, we present the proposed method in the frequency domain, although the methodology carries over to the time domain with proper modifications in the implementation.

2.1 Compressive imaging

Given the knowledge of the source wavelet, compressive imaging aims to obtain least-squares migrated images in a computationally efficient manner, by solving the following Basis Pursuit De-Noise (BPDN) problem (Herrmann & Li 2012):

$$\begin{aligned} \text{BPDN}(\mathbf{w}, \sigma): \quad & \underset{\mathbf{x}}{\text{minimize}} \ \|\mathbf{x}\|_1 \\ & \text{subject to} \quad \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{d}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2 \le \sigma^2. \end{aligned} \tag{1}$$

In this formulation, vector $\mathbf{w}$ represents the Fourier spectrum of the source wavelet, and the tolerance parameter σ allows for data mismatch due to noise in the data and modelling errors. Vector $\mathbf{x}$ is the curvelet coefficients of the image $\delta\mathbf{m}$, that is, $\delta\mathbf{m} = \mathbf{C}^*\mathbf{x}$ with $\mathbf{C}$ the curvelet synthesis operator (Candès & Donoho 2004), and $\delta\mathbf{m}$ itself being the perturbations over a given gridded background velocity model $\mathbf{m}_0$ parametrized by the square of the P-wave slowness (multipamameter/elastic inversion is doable but is beyond the scope of this paper). In the data-fitting constraint of the above optimization program, i indexes the discretized frequencies, and j indexes the different source experiments. Instead of using all n_f discretized frequencies and n_s source experiments, Herrmann & Li (2012) proposed to subsample the monochromatic source experiments to reduce the simulation cost in forward modelling: notations Ω and Σ denote the randomly selected $n'_f \ll n_f$ frequencies and $n'_s \ll n_s$ source experiments. By subsampling, the number of wave-equation solves in each iteration is reduced by a factor of $(n_f * n_s)/(n'_f * n'_s)$. While a brief explanation of drawing the simultaneous source experiments is to multiply a tall (i.e. more rows than columns: the number of rows is the number of sequential sources and the number of columns is the number of simultaneous sources) source encoding matrix on the right-hand-side of the source matrix, we refer to Herrmann & Li (2012) and Tu & Herrmann (2015a) on the details of the subsampling procedure. The underlined variables $\underline{\mathbf{d}}_{i,j}$ and $w_i\underline{\mathbf{s}}_j$ represent the jth column of the subsampled *primary-only* data and *point*-sources respectively. We assume that all sources share the same source signature $\mathbf{w}$, while it is possible to extend our work to allow for source estimation for each shot separately, by leveraging the fact that the unknown source wavelets still have a much smaller dimension compared to the model parameters and the seismic data. The operator $\nabla\mathbf{F}$ represents the linearized Born scattering operator. Note that the source weight w_i is separable from the source vector $\underline{\mathbf{s}}_j$, that is, $\nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j]\delta\mathbf{m} = w_i\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\delta\mathbf{m}$. Throughout this paper, we form the operator $\nabla\mathbf{F}$ based on the two-way wave equation, therefore the application of $\nabla\mathbf{F}^*$ to data residuals produces reverse-time migrated images (Lailly 1983). The above formulation means that among all possible solutions, the optimization program looks for the sparsest curvelet coefficients that, after curvelet synthesis and linearized modelling with the simultaneous sources, predict the simultaneous observed data up to a noise threshold specified by σ. With the above compressive-imaging formalism in place, we now continue to formulate the source estimation problem.

2.2 Source estimation

Now we identify the source wavelet w_i, $i \in \Omega$ (collected in the vector $\mathbf{w}$) as an additional unknown in the compressive imaging formulation, that is, we have

$$\begin{aligned} \text{BPDN}(\sigma): \quad & \underset{\mathbf{x},\mathbf{w}}{\text{minimize}} \ \|\mathbf{x}\|_1 \\ & \text{subject to} \quad \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{d}}_{i,j} - w_i\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2 \le \sigma^2. \end{aligned} \tag{2}$$

While conceptually attractive, BPDN(σ) does not lend itself to develop tractable algorithms. Following early work by van den Berg & Friedlander (2008), we interchange the objective and constraint to arrive at an extension—remember the formulation now includes the source as an unknown, which appears nonlinearly in the BPDN(σ) program, of the Least Absolute Shrinkage and Selection Operator (LASSO; Tibshirani 1996) formulation:

$$\begin{aligned} \text{LASSO}(\tau): \quad & \underset{\mathbf{x},\mathbf{w}}{\text{minimize}} \ f(\mathbf{x}, \mathbf{w}) \doteq \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{d}}_{i,j} - w_i\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2 \\ & \text{subject to} \quad \|\mathbf{x}\|_1 \le \tau. \end{aligned} \tag{3}$$

As shown by Aravkin *et al.* (2013), the set of minimizers of problem BPDN(σ) and LASSO(τ) coincide, given the condition that each minimizer of problem BPDN(σ) satisfies its constraint to equality (i.e. $f(\mathbf{x}, \mathbf{w}) = \sigma^2$). Not coincidently, when this condition is met, each minimizer also satisfies the constraint of the LASSO(τ) problem to equality, meaning that we can look for such a τ that the infimum of $f(\mathbf{x}, \mathbf{w})$

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

with $|\mathbf{x}|_1 \leq \tau$ has a value that equals σ^2. Because the objective function of problem LASSO(τ) adopts the canonical form of a separable least-squares problem (Golub & Pereyra 2003), we can now solve the problem using variable projection combined with projection onto a convex set. In the next few sections, we will first discuss how we solve problem LASSO(τ) with a given τ, and then discuss how we compute the right τ so that problem LASSO(τ) converges to the same solution as problem BPDN(σ).

2.3 Variable projection

To incorporate variable projections for the source wavelet into ℓ_1-norm sparsity promoting imaging, let us first consider iterations that involve soft thresholding that undergirds this type of ℓ_1-norm optimization. Model parameters at the kth iteration are in this case projected onto the ℓ_1-norm ball $\|\mathbf{x}\|_1 \leq \tau$, and the involved projected-gradient step is given by

$$\mathbf{x}^{k+1} = \mathrm{P}_{\mathcal{X}}\left[\mathbf{x}^k + \lambda \nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{w})|_{\mathbf{x}=\mathbf{x}^k, \mathbf{w}=\mathbf{w}^k}\right]. \tag{4}$$

In this expression, λ is a line-search parameter, $\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{w})|_{\mathbf{x}=\mathbf{x}^k, \mathbf{w}=\mathbf{w}^k}$ is the gradient of the objective function in eq. (3) with respect to $\mathbf{x}$ at the kth iteration, and $\mathrm{P}_{\mathcal{X}}$ denotes the projection onto the feasible set $\mathcal{X} = \{\mathbf{x} : \|\mathbf{x}\|_1 \leq \tau\}$.

Contrary to conventional ℓ_1-norm promoting imaging where $\mathbf{w}$ is assumed to be known, we need to evaluate this gradient $\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{w})|_{\mathbf{x}=\mathbf{x}^k, \mathbf{w}=\mathbf{w}^k}$ with variable projection at each iteration for unknown $\mathbf{w}^k$'s. During each iteration, given a solution of $\mathbf{x}$, we look for the least-squares solution of $\mathbf{w}$:

$$\widetilde{\mathbf{w}}(\mathbf{x}) = \underset{\mathbf{w}}{\operatorname{argmin}}\, f(\mathbf{x}, \mathbf{w}), \tag{5}$$

which permits the following closed-form solution for the source wavelet at the ith frequency (Pratt 1999):

$$\widetilde{w}_i(\mathbf{x}) = \frac{\sum_{j\in\Sigma} \langle \underline{\mathbf{d}}_{i,j}, \nabla \mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}{\sum_{j\in\Sigma} \langle \nabla \mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}, \nabla \mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}. \tag{6}$$

Here, the angular brackets $\langle\cdot, \cdot\rangle$ denote the inner product of two vectors, and $\widetilde{w}_i(\mathbf{x}), i \in \Omega$ denote complex-valued estimates of the source wavelet at different angular frequencies. The above solution enables us to eliminate the unknown source wavelet term in problem LASSO(τ), rendering the problem to solely and nonlinearly depend on $\mathbf{x}$:

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}}\ \overline{f}(\mathbf{x}) \doteq f(\mathbf{x}, \widetilde{\mathbf{w}}(\mathbf{x})) \\ &\text{subject to} \quad \|\mathbf{x}\|_1 \leq \tau. \end{aligned} \tag{7}$$

At first inspection, the above problem becomes more complicated than LASSO(τ) as we need to evaluate the derivative of $\widetilde{w}_i(\mathbf{x}), i \in \Omega$ with respect to $\mathbf{x}$. However, Aravkin & van Leeuwen (2012, corollary 2.3) have proved that a stationary point of the above problem remains a stationary point of problem LASSO(τ), and we have (Aravkin & van Leeuwen 2012, theorem 2.1)

$$\nabla_{\mathbf{x}} \overline{f}(\mathbf{x}) = \partial_{\mathbf{x}} f(\mathbf{x}, \widetilde{\mathbf{w}}(\mathbf{x})) + \partial_{\widetilde{\mathbf{w}}} f(\mathbf{x}, \widetilde{\mathbf{w}}(\mathbf{x}))\partial_{\mathbf{x}}\widetilde{\mathbf{w}}(\mathbf{x}) = \nabla_{\mathbf{x}} f(\mathbf{x}, \widetilde{\mathbf{w}}(\mathbf{x})), \tag{8}$$

because the term $\partial_{\mathbf{x}}\widetilde{\mathbf{w}}(\mathbf{x})$ vanishes as $\widetilde{\mathbf{w}}(\mathbf{x})$ is a minimizer. The above equation means that in each iteration, after vector $\mathbf{x}$ is modified according to eq. (4), we can compute $\nabla_{\mathbf{x}} \overline{f}(\mathbf{x})$ by evaluating $\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{w})$ of problem LASSO(τ) at $\mathbf{w} = \widetilde{\mathbf{w}}(\mathbf{x})$ via eq. (6).

While there is no guarantee the variable projection approach will yield the globally optimal solution as the problem remains non-convex, numerical experiments show that in most cases it enables the optimization program to converge within less iterations (Golub & Pereyra 2003). Rickett (2013) also showed its superior performance over the simultaneous descent method in source estimation for full-waveform inversion applications. Furthermore, the projections themselves (*cf.* eq. 6) are computationally affordable as they do not involve additional expensive operations such as wave-equation solves.

2.4 Relaxing the ℓ_1-norm constraint

In this section, we describe our strategy to select the right τ so we actually solve problem BPDN(σ) by solving problem LASSO(τ). We follow the methodology proposed by van den Berg & Friedlander (2008) in SPGℓ_1, where a series of τ's are found by solving a root-finding problem on the Pareto trade-off curve. With the unknown source wavelet, we find the τ's by solving the value function $\nu(\tau) \doteq \inf f(\mathbf{x}, \mathbf{w})|_{|\mathbf{x}|_1 \leq \tau} = \sigma^2$, which yields the following ℓ_1-norm constraint for the $(l+1)$th LASSO subproblem using the Newton's method with an initial guess of $\tau^0 = 0$:

$$\tau^{l+1} = \tau^l - \frac{\nu(\tau^l) - \sigma^2}{\nu'(\tau^l)}. \tag{9}$$

Using this approach, we arrive at the solution of BPDN(σ) by solving a series of LASSO(τ) subproblems for gradually increasing τ's. While $\nu(\tau^l)$ in the above equation can be evaluated straightforwardly with the previous τ estimate, the evaluation of $\nu'(\tau^l)$ is more difficult, because the nonlinearity introduced by the unknown source wavelet violates the linearity assumption on the forward model, which is needed

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

to compute this value (Aravkin *et al.* 2013). However, by treating $\mathbf{w}$ as fixed, we approximate $\nu'(\tau^l)$ by (van den Berg & Friedlander 2008; Aravkin *et al.* 2013):

$$\nu'(\tau^l) \approx -\left\| \sum_{i\in\Omega}\sum_{j\in\Sigma} \mathbf{C}\nabla\mathbf{F}^*[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j](\underline{\mathbf{d}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x})\right\|_\infty . \tag{10}$$

With numerical examples, we will show that this approximation works reasonably well.

2.5 Acceleration by redrawing random subsets

Except for the extra variable projection step, solving LASSO(τ) is similar to compressive imaging (Herrmann & Li 2012), during which computational costs can be reduced by working with randomized subsets of data. As shown by Herrmann & Li (2012) and Tu & Herrmann (2015a), solutions of BPDN($\mathbf{w};\sigma$) for a given source wavelet can be accelerated significantly by selecting new independent randomized subsets of frequencies and simultaneous sources, after each corresponding LASSO subproblem is solved. We found empirically that working with these new independent subsets leads to faster decay of the model error and to improved robustness with respect to possible linearization errors (Tu *et al.* 2013)—that is, $\underline{\mathbf{d}}_{i,j} - \mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j] \approx \nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j]\delta\mathbf{m}$. Both findings can be explained because independent subsets tend to break correlations between errors in the solution and the randomly selected subsets of data (Herrmann 2012).

While these empirical findings lead to a computationally feasible and robust compressive imaging scheme, we need to justify that working with different randomized subsets does not interact adversely with the proposed source estimation approach. Aside from additional empirical evidence from various examples included below, which suggests that there is no measurable adverse effect, we argue that (i) numerical evaluation of the value function, $\nu'(\tau)$, and its derivative $\nu'(\tau)$ remain similar as long as we keep the number of frequencies and simultaneous sources the same (evidenced by Herrmann & Li 2012, fig. 2c); and (ii) we only draw new subsets after each LASSO(τ^l) is solved. The latter argument assumes that estimates of the source at the *l*th subproblem have converged, which is reasonable to make because the number of knowns in eq. (6) for each frequency far exceeds the dimensionality of the single unknown complex-value for the source.

3 RESOLVING THE SCALING AMBIGUITY WITH MULTIPLES

In the algorithm presented above, fundamental issues related to the ambiguity with respect to amplitude scalings remain, which are intrinsic to any blind-deconvolution type of problems where both the source wavelet and the seismic image are unknown (Stockham *et al.* 1975). However, compared to source estimation in data space using the convolution model (Ulrych *et al.* 1995), source estimation in the image space does not suffer from ambiguities with respect to global phase shifts, which can be explained by the multiplicity of data and moveout characteristics that are specific to the background-velocity model being used. As stated in the introduction, we assume this background model to be given and to be relatively accurate.

Unfortunately, the same observation does not apply to the ambiguity with respect to amplitude scalings, which correspond mathematically to an invariance of our objective functions with respect to amplitude scalings by $\alpha \in \mathbb{R}^+$—that is, we have

$$f(\mathbf{x}, \mathbf{w}) = f\left(\frac{1}{\alpha}\mathbf{x}, \alpha\mathbf{w}\right). \tag{11}$$

This invariance results in amplitude ambiguity in the source estimation, which originates from the fact that the forward model in LASSO(τ), that is, $w_i\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}$, is bi-linear (i.e. when one variable is fixed, the forward model is linear w.r.t. the second variable) w.r.t. the source wavelet $\mathbf{w}$ and the solution vector $\mathbf{x}$.

Unfortunately, including sparsity constraints via the ℓ_1-norm or even via the non-convex ℓ_1/ℓ_2-norm (Esser *et al.* 2015), are inadequate to resolve the above amplitude ambiguity. We therefore alternatively resort to the physics of the free surface. By incorporating the free surface, our problem becomes nonlinear because the forward model now includes a 'feedback loop' (Verschuur *et al.* 1992), which generates surface-related multiples from primaries scaled by the (unknown) source wavelet. Extending the physics in this way allows us to mitigate the scaling ambiguity by using observed surface-related multiples to estimate the source wavelet as proposed during EPSI (van Groenestijn & Verschuur 2009; Lin & Herrmann 2013). Following recent work by Tu & Herrmann (2015a), we propose to do this by incorporating predictions of surface-related multiples into our formulation for compressive imaging.

3.1 Compressive imaging with total upgoing wavefields

If we ignore internal multiples, the total upgoing wavefield, including primaries and surface-related multiples, can be modelled by including the observed upgoing wavefields $\mathbf{u}_{i,j}$ at the water surface as areal sources in the linearized Born scattering operator (Tu & Herrmann 2015a), yielding

$$\mathbf{u}_{i,j} \approx \nabla\mathbf{F}[\mathbf{m}_0, w_i\mathbf{s}_j - \mathbf{u}_{i,j}]. \tag{12}$$

As the water surface has a reflection coefficient of approximately -1, we replace the downgoing spatially impulsive source wavefields $w_i\mathbf{s}_j$ by 'areal' sources $w_i\mathbf{s}_j - \mathbf{u}_{i,j}$, that is, we include the downgoing receiver wavefields at the water surface $-\mathbf{u}_{i,j}$ into the source wavefields. We

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

obtain $\mathbf{u}_{i,j}$ after careful pre-processing to the observed seismic data, such as applying source-receiver reciprocity (to fill in missing traces), up-down decompositions (for receiver-side deghosting), and extrapolation of the upgoing wavefield from the receiver level to the free surface (see e.g. Verschuur *et al.* 1992). Note that source ghosts should be kept intact in the data (Verschuur *et al.* 1992). As a result, we arrive at a formulation that remains conducive to sparsity-promoting imaging with source estimation via variable projection. Because prediction for the multiples are carried out by the wave-equation solver, this forward model is also computationally viable compared to processing work flows that involve separate multiple-prediction/separation and RTM-imaging procedures.

To arrive at our final formulation, we first replace our objective by

$$f(\mathbf{x}, \mathbf{w}) = \sum_{i\in\Omega}\sum_{j\in\Sigma} \|\underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, w_i\underline{\mathbf{s}}_j - \underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}\|_2^2, \tag{13}$$

which we obtain by substituting the primary wavefield with the total upgoing wavefield and the spatially impulsive sources with areal sources containing the total downgoing wavefield at the surface. Given this new definition for the objective, we proceed by solving the LASSO(τ) subproblems with variable projections. For this end, we solve for all (simultaneous) sources and for each frequency the following problem:

$$\underset{w_i}{\text{minimize}} \quad \sum_{j\in\Sigma} \|\underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, -\underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x} - w_i\nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\|_2^2, \quad i \in \Omega,$$

which in turn permits the following analytic solution:

$$\tilde{w}_i(\mathbf{x}) = \frac{\sum_{j\in\Sigma}\langle \underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, -\underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}, \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}{\sum_{j\in\Sigma}\langle \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}, \nabla\mathbf{F}[\mathbf{m}_0, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}. \tag{14}$$

To arrive at the modified solution for the source wavelet above, we substitute the primary-only wavefield by an estimate for the primaries given by the total upgoing wavefield minus the current prediction for the surface-related multiples—that is, $\underline{\mathbf{d}}_{i,j} \mapsto \underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m}_0, -\underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}$. The multiple predictions do not require information on the source wavelet; rather, they are carried out by solving wave-equations with the areal sources given by $-\underline{\mathbf{u}}_{i,j}$. With these substitutions, we are able to estimate the source wavelet by variable projection at additional costs of a single linearized forward modelling operation to predict the multiples.

In return of the increased computational costs, including surface-related multiples during imaging with source estimation has several key advantages, namely (i) including the multiples mitigates the ambiguity in the source estimation, which leads to a scaling of the source wavelet (and the image) that leads to the prediction and separation of the primary and multiple wavefields, which makes it possible to map multiple energy onto the true image; (ii) in seismic imaging, multiples provide extra illumination coverage from primaries (as each receiver acts as a virtual source), especially along the cross-line direction in 3-D seismic data (Long *et al.* 2013; Tu & Herrmann 2015b).

3.2 Putting it all together

With the building blocks of our sparsity-promoting imaging with source estimation and multiples in place, we summarize our proposed imaging scheme in Algorithm 1.

Algorithm 1 Fast imaging with source estimation and multiples

Input:
total upgoing wavefield $\mathbf{u}$, background velocity model $\mathbf{m_0}$,
tolerance $\sigma = 0$, iteration limit $k_{\max}$
Initialization:
iteration index $k \leftarrow 0$, LASSO subproblem index $l \leftarrow 0, \tau^0 \leftarrow 0, \mathbf{x^0} \leftarrow \mathbf{0}$
$w_i = 1$ for all $i \in 1, \ldots, n_f$
while $k < k_{\max}$ **do**
  $\Omega_l, \Sigma_l, \underline{u}_{i,j}, \underline{s}_j \leftarrow$ new independent draw
  $\tau^l \leftarrow$ determined from τ^{l-1} and σ using Newton's method
  $\mathbf{x}^l \leftarrow \begin{cases} \text{argmin}_{\mathbf{x}} \sum_{i\in\Omega_l, j\in\Sigma_l} \|\underline{u}_{i,j} - \nabla\mathbf{F}_i[\mathbf{m}_0, w_i(\mathbf{x})\underline{s}_j - \underline{u}_{i,j}]\mathbf{C}^*\mathbf{x}\|_2^2 \\ \text{subject to } \|\mathbf{x}\|_1 \le \tau^l \end{cases}$
    //warm start with $\mathbf{x}^{l-1}$, solved in k^l iterations
    //in each iteration, update the source by
    $w_i(\mathbf{x}) = \frac{\sum_{j\in\Sigma}\langle \underline{\mathbf{u}}_{i,j} - \nabla\mathbf{F}[\mathbf{m_0}, -\underline{\mathbf{u}}_{i,j}]\mathbf{C}^*\mathbf{x}, \nabla\mathbf{F}[\mathbf{m_0}, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}{\sum_{j\in\Sigma}\langle \nabla\mathbf{F}[\mathbf{m_0}, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}, \nabla\mathbf{F}[\mathbf{m_0}, \underline{\mathbf{s}}_j]\mathbf{C}^*\mathbf{x}\rangle}$
  $k \leftarrow k + k^l, l \leftarrow l + 1$
end while
Output: model perturbations $\delta\mathbf{x} = \mathbf{C}^*\mathbf{x}$

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

3.2.1 Input and initialization [Lines 1–6]

As discussed above, our proposed inversion approach requires the knowledge of a reasonably accurate background velocity model $\mathbf{m}_0$. As discussed in Tu *et al.* (2013), we choose a zero σ to prevent the optimization algorithm from being terminated prematurely (Line 3). The initial guess of the source wavelet is simply an impulse at zero time, which corresponds to a flat Fourier spectrum with unit amplitude and zero phase (Line 6). Both the solution vector $\mathbf{x}$ and the sparsity level τ are simply initialized as zeros.

3.2.2 Main loop [Lines 7–16]

In the main loop we solve a series of LASSO(τ^l) subproblems. For each subproblem, we draw new independent subsets of randomized frequencies and simultaneous sources (Line 8). We update the sparsity parameter τ^l using Newton's method (Line 9), and solve for the model parameters collected in vector $\mathbf{x}$ (Line 10) as well as the source wavelet collected in vector $\mathbf{w}$ (Line 13) by variable projection. Note that we do not impose any assumption on the phase of the source wavelet (e.g. minimum phase assumption as used in Robinson 1967) during source estimation. We terminate the optimization program when the pre-specified maximal number of iterations $k_{\max}$ is reached.

4 EXAMPLES

To validate the efficacy of the proposed imaging scheme with source estimation, we conduct a series of synthetic examples designed to illustrate key aspects of our algorithm including robustness to noise or modelling errors, resolving scaling ambiguity with multiples, and sensitivity to the initial guess of the source wavelet.

4.1 Imaging with primaries only

Most conventional seismic data processing work flows use the primary wavefield as the input for migration, that is, after a de-multiple procedure to the observed data, for example using SRME (Verschuur *et al.* 1992). We therefore first apply our method to primary-only data to demonstrate its advantages over RTM when used in conventional data work flows.

4.1.1 Experiment setup

We use a 2-D slice of the SEG/EAGE salt model. We pad 10 grid points at the top of the model so that the water layer becomes slightly thicker to better retain the water velocity near the surface when we smooth the true model to get the background model. In this way, we can model and therefore remove the direct waves more accurately. The true and the background-velocity models are shown in Fig. 1. The model is 3.9 km deep and 15.7 km long, with a grid spacing of 24.38 m (80 ft). We use a fixed-spread configuration, and put 323 co-located sources and receivers with a 48.77 m lateral grid spacing (i.e. every other grid point) at a depth of 24.38 m (i.e. the second vertical grid point). We use a Ricker wavelet with a peak frequency of 5 Hz, and the peak of the wavelet is at 0.25 s. We record for 8 s, which in the frequency domain yields 96 discretized frequencies up to 12 Hz (roughly the highest frequency that can be used for a 80-ft grid spacing to avoid numerical dispersion).

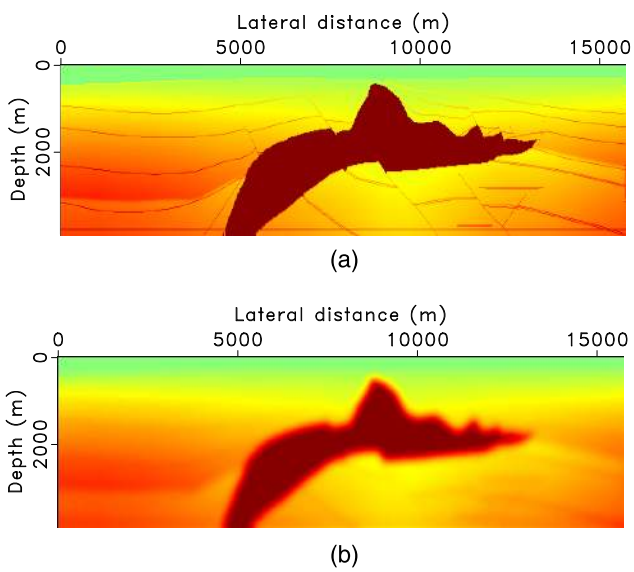

Figure 1. A slightly modified 2-D slice of the SEG/EAGE salt model. (a) True velocity model. (b) Smoothed background velocity model.

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

4.1.2 Ideal case

To rule out the influence of any imperfection in the data, we first test our method with an idealized setup, that is, we synthesize and invert the primary-only data using the same linearized modelling engine. To obtain a baseline image, we run a conventional RTM by assuming that we know the true source wavelet (Fig. 2a). To remove the low-wavenumber artefacts in the RTM image, we apply a high-pass filtering along the depth dimension (Mulder & Plessix 2003). Using the true source wavelet, we also obtain a least-squares migrated image using the compressive-imaging approach (Herrmann & Li 2012), shown in Fig. 2(b). To demonstrate the importance of using an accurate source wavelet, we obtain the second least-squares migrated image using compressive imaging with a wrong source wavelet, shown in Fig. 2(c). The wrong source wavelet has a time advance of 0.1 s (i.e. a phase shift) compared with the true source wavelet. We obtain the third least-squares migrated image using the proposed method, shown in Fig. 2(d).

We can see that compared with conventional RTM (Fig. 2a), the proposed approach produces an image of higher spatial resolution given the true source wavelet (Fig. 2b). The higher spatial resolution is achieved because solving the least-squares formulation, through many iterations, accounts for the effect of the Gauss–Newton Hessian on the seismic image (we refer to the 'basic equations' section of Nemeth *et al.* 2000, for more details). However, the image quality can be significantly compromised when the wavelet is wrong. In this case the shifted

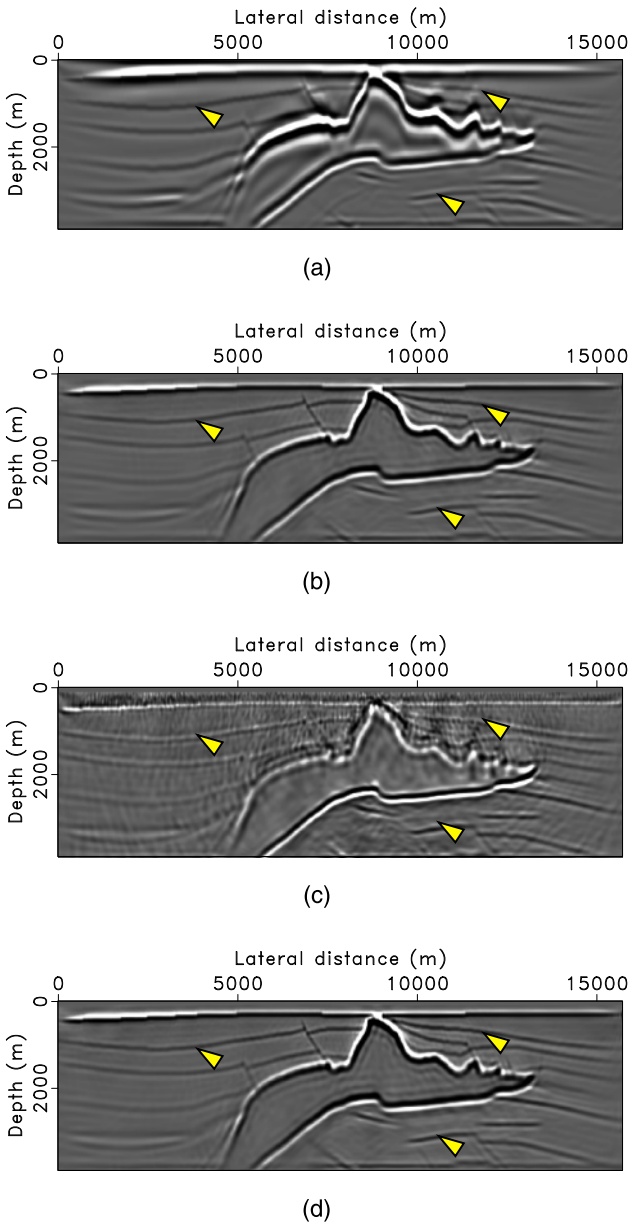

Figure 2. Imaging results of the ideal primary-only data set. The arrows point to the same subsurface locations. (a) Conventional RTM with the true source wavelet. Compressive imaging results (b) with the true source wavelet, (c) with the wrong source wavelet that has a 0.1 s time advance and (d) with the proposed source-estimation method. The images in panels (b) and (d) are comparable, and are of much higher spatial resolution than the conventional RTM image in panel (a).

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

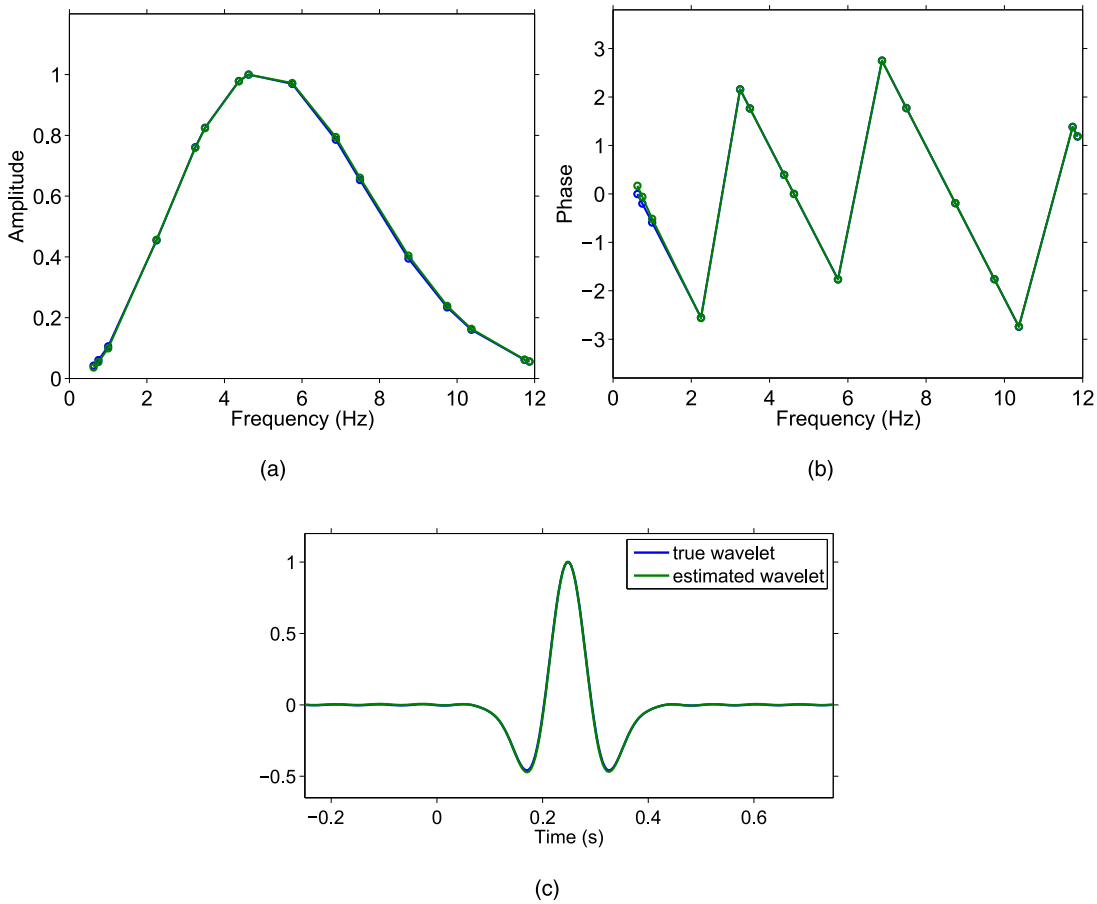

Figure 3. Comparison of the true (in blue) and the estimated (in green) source wavelets from the ideal primary-only data set. Amplitude normalization is applied. (a) The amplitude spectrum. (b) The phase spectrum. (c) The waveform.

phase of the wavelet not only produces an image contaminated with subsampling-related noises, but also leads to a shift of the subsurface structures in depth (Fig. 2c), which can result in erroneous interpretations. With the proposed source-estimation approach, we obtain a faithful image (Fig. 2d), which is comparable to the image obtained using the true source wavelet (Fig. 2b).

Remarks on the experiment. In terms of computational cost, we use 16 randomly selected frequencies and 16 simultaneous sources (about 120× subsampling), and 60 iterations for all inversion results (remember one iteration involves one forward and one adjoint evaluations of the demigration operator); as a result, the number of wave-equation solves through inversion roughly equals that of a single RTM. Note that source estimation by evaluating eq. (6) does not involve extra wave-equation solves other than the ones needed for linearized forward modelling required by the imaging step. For all inversion results in this and later sections, after the inversion is finished, we apply a curvelet thresholding to remove residual incoherent noises in the image (Herrmann *et al.* 2008b). The threshold is chosen in such a way that the thresholded noise does not contain noticeable coherent energy. As our solution vector $\mathbf{x}$ is already in the curvelet domain, extra computation incurred in this thresholding step is minimal.

As a quality-control (QC) procedure, we also plot in Figs 3(a) and (b) the spectra of the estimated source wavelet. For an intuitive interpretation, we also obtain the waveform of the wavelet and show it in Fig. 3(c). Because reliable source estimates output by the algorithm only exist at frequencies that are sampled in the last LASSO(τ) subproblem, we obtain the waveform of the source wavelet by first computing the spectrum of the source wavelet at all frequencies using eq. (6) with the inverted seismic image, and then applying the inverse Fourier transform to the spectrum (this procedure is also applied to all later examples). Because of the scaling ambiguity (of the image as well as the source wavelet) we discussed above, we normalize the amplitude (i.e. the ℓ_{inf} norm) of both the true and the estimated source wavelets to compare them. The source estimates are highly accurate except for the absolute amplitude scaling.

4.1.3 More realistic case

Compared with data synthesized by linearized modelling, field seismic data are contaminated with all sorts of incoherent and coherent noises. While most incoherent noises can be removed in data pre-processing, coherent noises, such as internal multiples, are difficult to identify and remove and can therefore degrade the quality of the final seismic image. In least-squares migrations, as the objective is to match the predicted linearized data and the observed data, other sources of errors also include linearization errors (i.e. as the name suggests, the errors incurred

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

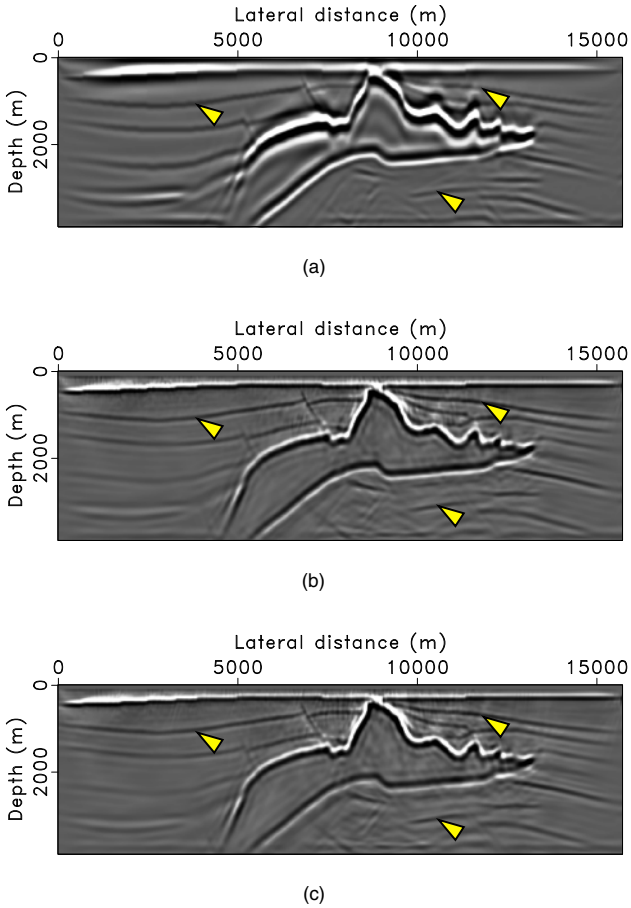

Figure 4. Imaging results of the primary-only data set modelled using iWave. The arrows point to the same subsurface locations. (a) Conventional RTM image. Compressive imaging results using (b) the true source wavelet and (c) with the proposed source-estimation method.

during linearization of the wave equation) and errors by approximating earth physics using, for example, the acoustic wave-equation. To mimic some of these noises and errors, which we will generally refer to as 'modelling errors' hereafter, we choose to use iWave, a 2-D time-domain finite-difference acoustic modelling engine (Terentyev *et al.* 2014), to simulate the observed data, and use our in-house frequency-domain finite-difference acoustic modelling engine to invert the data.

To test the robustness of our method to these 'modelling errors' as defined above, we compare the following examples. We again get the baseline image using conventional RTM with the true wavelet (Fig. 4a). Next we obtain two least-squares migrated images: one with the true source wavelet, shown in Fig. 4(b); the second one with source estimation, shown in Fig. 4(c).

We observe that, in the presence of these modelling errors in the data, (i) the proposed source-estimation method still yields an image that is comparable to the one obtained with the true source wavelet; (ii) both images are still of higher spatial resolution than the conventional RTM image. The modelling errors (especially internal multiples in this case because of the high velocity contrast around the salt boundary) do result in some imaging artefacts beneath the salt structure, but they do not particularly cause problems for the source-estimation approach. The above observations lead to the conclusion that the proposed source-estimation approach is relatively robust with respect to coherent modelling errors in the primary-only data.

Remarks on the experiment. In generating the data with iWave, we use an absorbing boundary condition at the surface to avoid generating surface-related multiples. The observed data are the difference between the data modelled with the true model and the background model, and therefore contain internal multiples. We use the same inversion parameters as the above set of examples, and as a result, the computational costs of the three images in Fig. 4 remain roughly the same.

In Fig. 5, we again compare the true wavelet and the estimated wavelet as a QC procedure. We can see that in this case the accuracy of the estimated source wavelet degrades gracefully compared with the idealized case, although there exists some difference, mainly in the phase (see Figs 5b and c), between the two wavelets. To understand the cause of this difference, we compare in Fig. 5(d) traces that contain strong top-salt reflections from two data sets: one synthesized using our linearized frequency-domain modelling and another one using iWave. We can see that the slight phase shift can also be observed from the data, most likely due to linearization errors especially by smoothing the salt structure and the difference between iWave and our in-house modelling engine. As the wavelet is estimated to fit the observed data, it is understandable that it contains such a phase error.

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

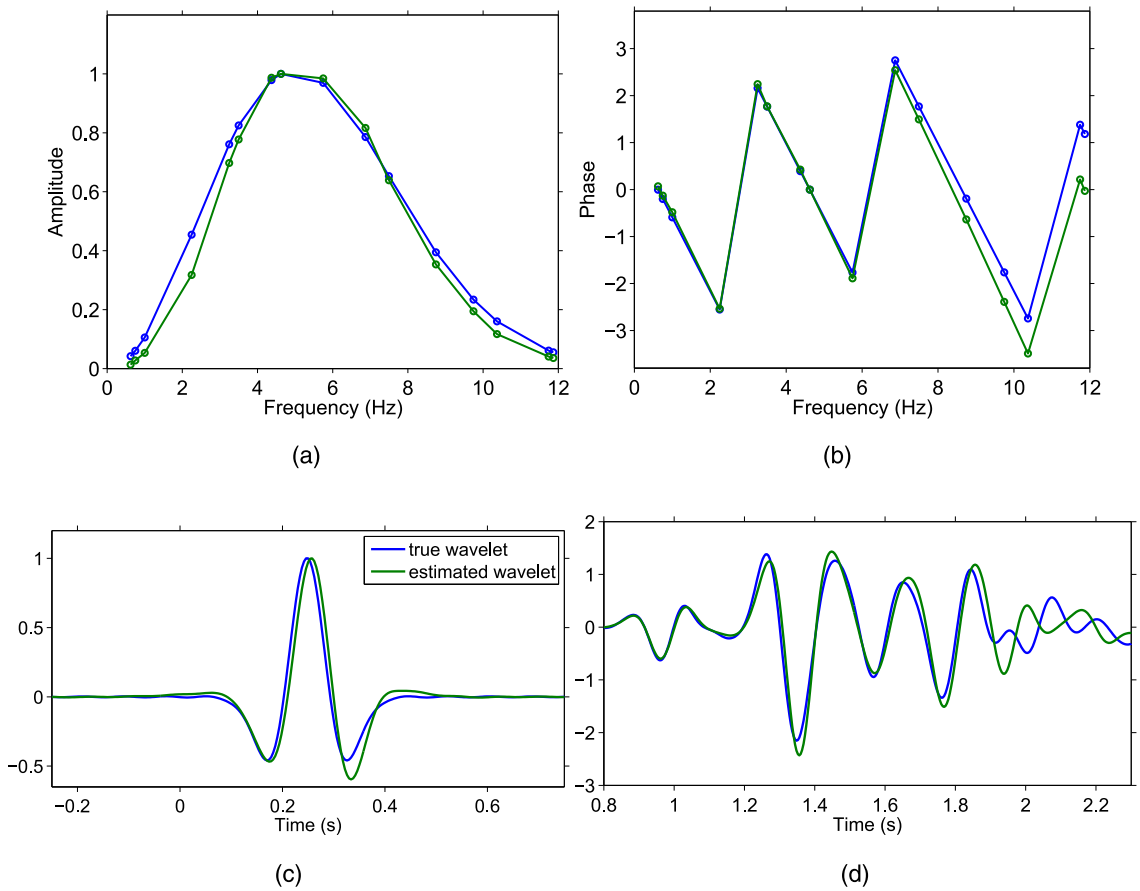

Figure 5. (a) The amplitude spectrum, (b) the phase spectrum and (c) the waveform of the estimated source wavelet from the primary-only data set modelled using iWave, compared to that of the true source wavelet. Amplitude normalization is applied. (d) Comparison of two traces modelled using our in-house frequency-domain linearized modelling (used during inversion, in blue) and iWave (used during data simulation, in green). Amplitude scaling is applied to compare the two traces.

As demonstrated by the two sets of examples above, our proposed source estimation method, when applied to conventional primary-only data, produces seismic images that are comparable to images obtained using true source wavelets. The method is also relatively robust to modelling errors such as linearization errors and difference between different modelling engines. However, as we discussed above, the scaling ambiguity cannot be solved by imaging with primaries alone because of the blind-deconvolutional nature of the problem. Next we will demonstrate how we resolve the ambiguity using surface-related multiples.

4.2 Scaling images with multiples

To validate our proposal to incorporate surface-related multiples in the imaging procedure to resolve the amplitude ambiguity, we apply the proposed method to data that contain surface-related multiples. As above, we study the performance of the proposed approach both with an ideal data set and with a more realistic data set.

4.2.1 Experiment setup

We use a sedimentary part of the Sigsbee 2B model (The SMAART JV 2014), which by design has a strong ocean-bottom reflector, to carry out examples using surface-related multiples. We reduce the water depth of the original model, to allow for more orders of multiples to be recorded. The true and the background velocity models are shown in Fig. 6. The model is 3.8 km deep and 6 km long, with 7.62 m (25 ft) grid spacing. We again use a fixed-spread configuration, and put 261 co-located sources and receivers with 22.86 m lateral grid spacing (i.e. every three grid point) at a depth of 15.24 meters (i.e. the third vertical grid point). We use a zero-phase Ricker wavelet with a peak frequency of 15 Hz, and record for 8.184 s (i.e. 1024 time samples with 8 ms sampling interval). This setup yields 311 discretized frequencies up to 38 Hz.

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

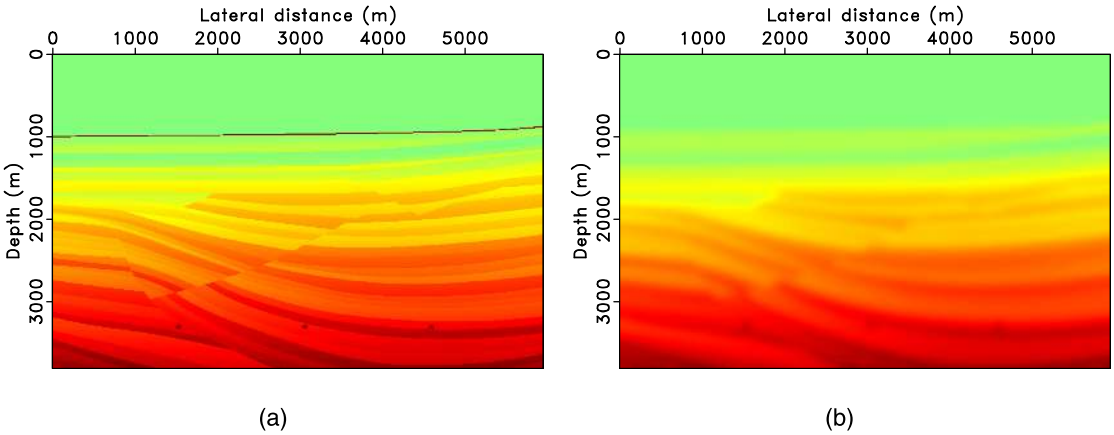

Figure 6. A sedimentary part of the Sigsbee 2B model. (a) True velocity. (b) Smooth background velocity.

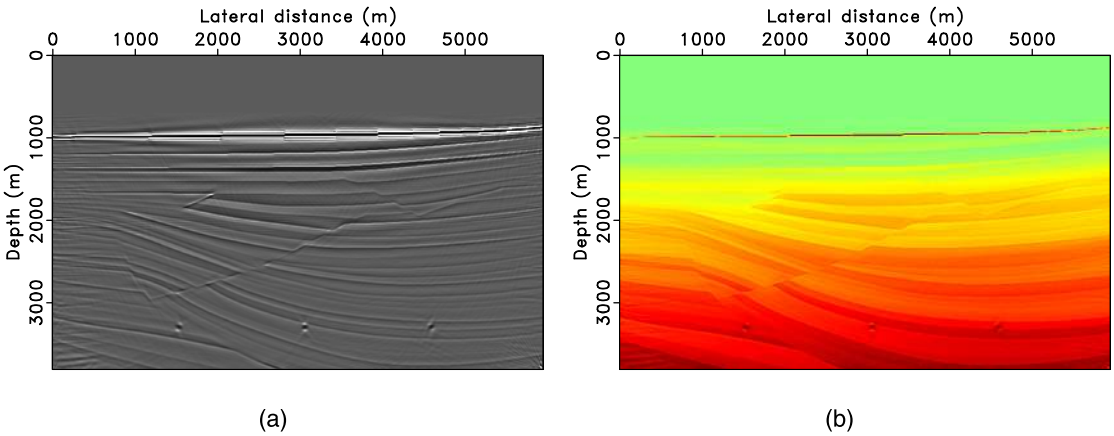

Figure 7. Compressive imaging-with-multiple result of the ideal data set containing multiples. (a) The inverted model perturbations. (b) Estimate of the true velocity model by adding the inverted model perturbations back to the background model.

4.2.2 Ideal case

We synthesize the ideal data set including surface-related multiples using linearized modelling following eq. (12). Note that in this case, the multiples are not generated by including an explicit free surface in the model, but by injecting the downgoing receiver wavefield as areal sources (which equates to SRME-type multiple prediction, Tu & Herrmann 2015a). To show the removal of amplitude ambiguity, we compare the imaging results obtained with source estimation, including both the image and estimates of the source wavelet during the last LASSO(τ) problem, against the true model and source wavelet. The inverted model perturbations are shown in Fig. 7(a). To show that it is indeed of the true amplitude and phase, we add it back to the background model to estimate the true model. The result is shown in Fig. 7(b), and is comparable to the true velocity model shown in Fig. 6(a). We compare the true and estimated source wavelets in Fig. 8, without applying *any* normalization. This comparison shows that in this idealized case, the proposed imaging-with-multiple approach can retrieve the original amplitude (and phase) of both the model perturbations and the source wavelet with high accuracy.

Remarks on the experiment. In the inversion, we use 31 randomly selected frequencies (out of all 311 frequencies), 26 simultaneous sources (out of all 261 sequential sources), and 50 iterations. As a result, the number of wave-equation solves is roughly 1.5× of one RTM with all the data (as discussed above, source estimation with multiples introduces a 50 per cent increase in simulation cost).

4.2.3 More realistic case

As above, we model a more realistic data set that contains multiples using iWave with a free-surface boundary condition (Terentyev *et al.* 2014). The observed data is the difference between the data modelled with the true model and the background model. Note that the 'modelling errors' in this data also arise from: (i) approximating the free-surface related source ghosts using a pair of mirrored sources without the free-surface, that is, we place an extra virtual source with the opposite sign of the true source at a position that is symmetric with the true source w.r.t. the surface level; (ii) approximating the surface-related multiples in the observed data using the SRME-type multiple prediction (although the prediction is carried out implicitly using wave-equation solvers, Tu & Herrmann 2015a). We discuss these two types of errors in more detail in the discussion section. Because the modelling engines used in data simulation and inversion have different scalings, we can

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

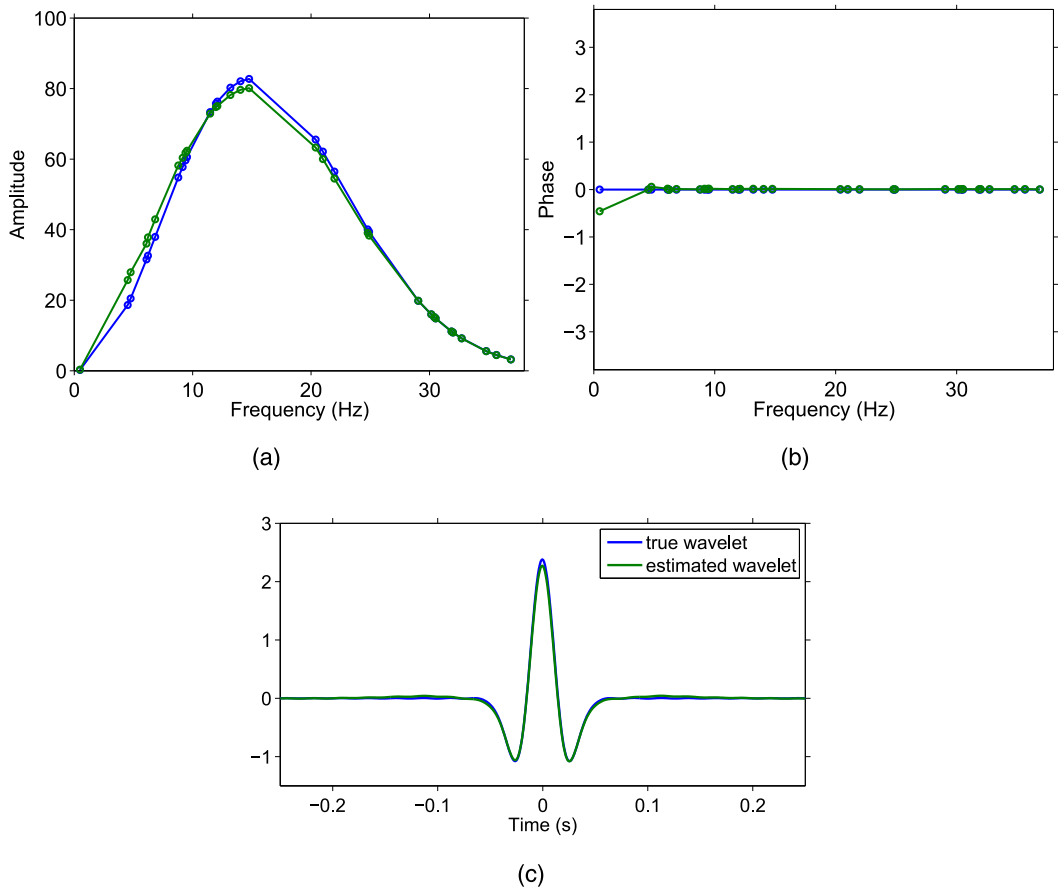

Figure 8. The true (in blue) and the estimated (in green) source wavelet from the ideal data set containing multiples. No normalization is applied. (a) The amplitude spectrum. (b) The phase spectrum. (c) The waveform.

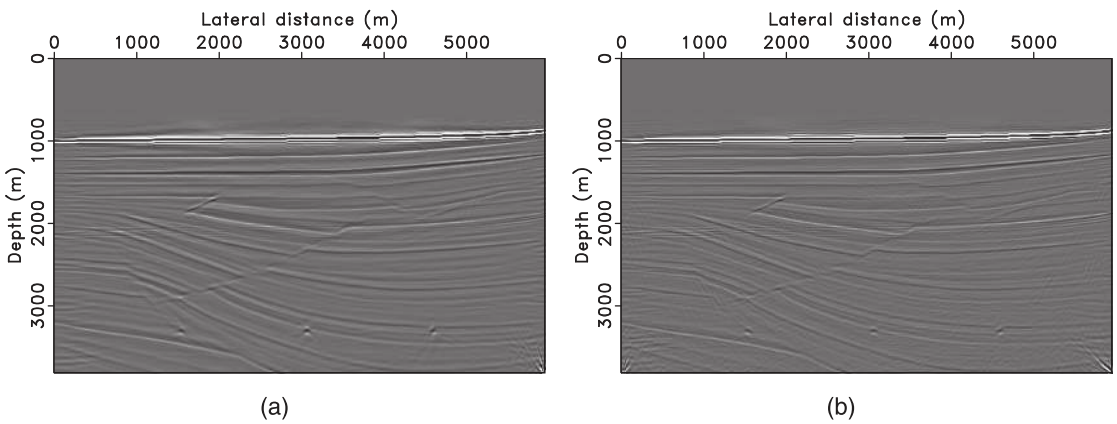

Figure 9. Compressive imaging-with-multiple results of the data set modelled using iWave. Images obtained (a) with the true source wavelet and (b) with source estimation.

no longer recover the original amplitudes of the image and the wavelet. However, the resulting scaling of the wavelet has a new meaning: it is the scaling that leads to the separation (in a minimum ℓ-2 norm sense) of primaries and surface-related multiples based on the SRME relation. We compare the images obtained with the true source wavelet and with our source estimation in Fig. 9. The inversion parameters and therefore the computational costs remain the same as the above set of examples. The two images are comparable, both containing minimal coherent artefacts from surface-related multiples (e.g. periodic copies of the ocean bottom at a deeper section of the image, see figs 4c and e in Tu & Herrmann 2015a).

We compare the true and the estimated source wavelets in Fig. 10. Because we do not recover the original amplitude, we normalize their amplitudes to compare. We observe that compared to the idealized case, the estimated source wavelet has a significant bias towards the low frequencies. Our numerical examples indicate that this bias is related to the extremely strong and geologically unrealistic thin (one

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

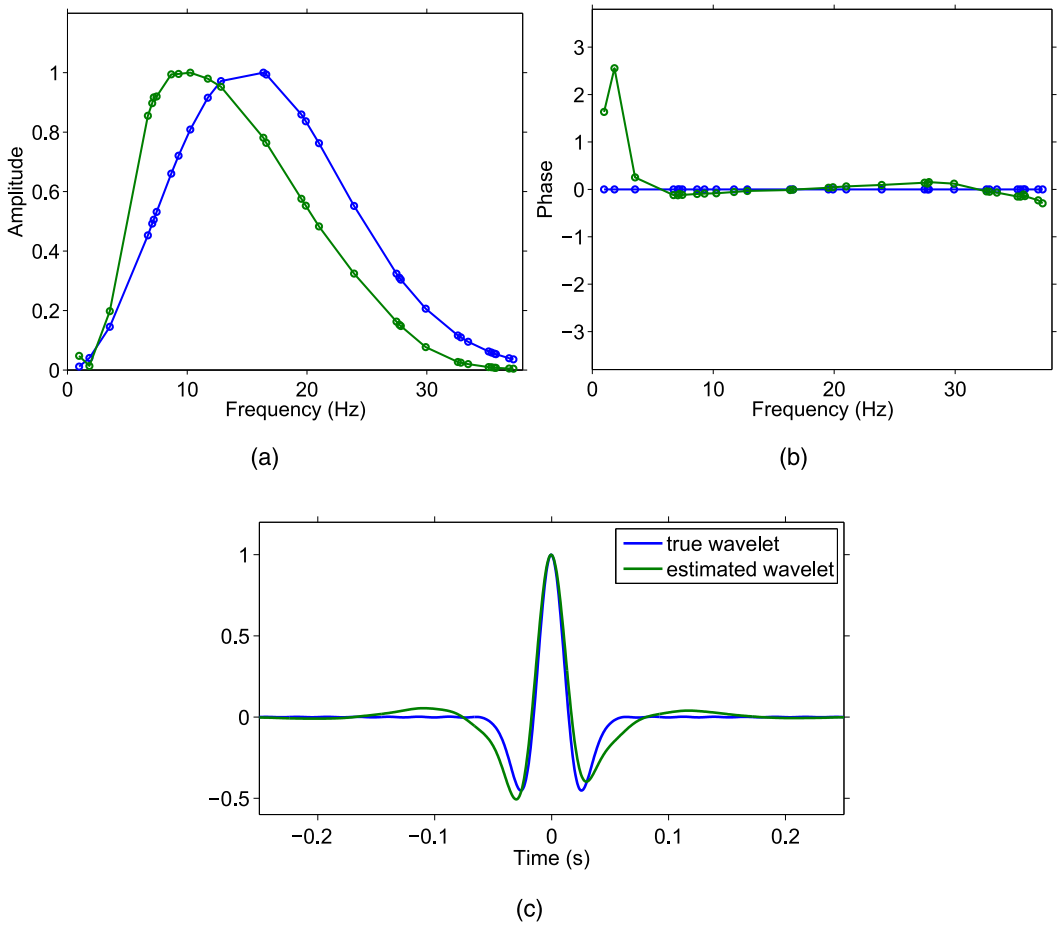

Figure 10. (a) The amplitude spectrum, (b) phase spectrum and (c) the waveform of the estimated (in green) source wavelet from the data set containing multiples, compared with the true source wavelet (in blue). The data set is modelled using iWave with a free-surface boundary condition.

grid-distance) high-velocity layer located at the ocean bottom. This layer was included in this Sigsbee 2B benchmark model to generate strong surface-related multiples in the water column (The SMAART JV 2014). While inclusion of this layer, makes sense to generate multiples it evidently causes issues with wave-equation based RTM, which is based on a linearization that assumes small perturbations while finite-difference solvers can generally not be expected to yield accurate results for such a thin and high-velocity layer. When we include this layer into the background model, we see in Fig. 11 that part of the bias is removed because there is no longer a severe linearization error associated with this layer. However, the results are still not perfect since the afore-mentioned forward modelling errors remain. The corresponding source estimates included in Fig. 11(a) is significantly improved and reasonably accurate at the high end of the available spectrum as expected. For completeness, we also include the estimate for the image in Fig. 11(b), which now no longer includes the unnatural velocity high at the ocean bottom.

4.3 Convergence analysis

To obtain a more quantitative understanding of the proposed method, we study and compare the convergence behaviours of the recovery algorithm with the true wavelet and with source estimation. We use the normalized cross-correlation (NCC) to measure how well the inverted results approximate true model perturbations. For two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ (in our case, we vectorize the true and inverted model perturbations first), the NCC is the inner product of two vectors divided by the product of the ℓ_2 norms of the two vectors, that is, normalized inner product:

$$\mathrm{NCC}(\mathbf{v}_1, \mathbf{v}_2) = \frac{\langle \mathbf{v}_1, \mathbf{v}_2 \rangle}{\|\mathbf{v}_1\|_2 \|\mathbf{v}_2\|_2}. \tag{15}$$

Geometrically the NCC is the cosine of the angle between two vectors and is not sensitive to the amplitude of either vector. The smaller the angle, the closer the two vectors are, and the closer the NCC is to one. We plot the NCC between the model iterates (we save a model iterate whenever a LASSO subproblem is finished) and the true model perturbation.

We plot the NCC curves in Fig. 12 for both the imaging-with-primary and the imaging-with-multiple results in the presence of modelling errors (see definition in the previous text). In both cases, we can see that the convergence behaviours of using the true source wavelet and source estimation are comparable, demonstrating the efficacy of the proposed source-estimation method.

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

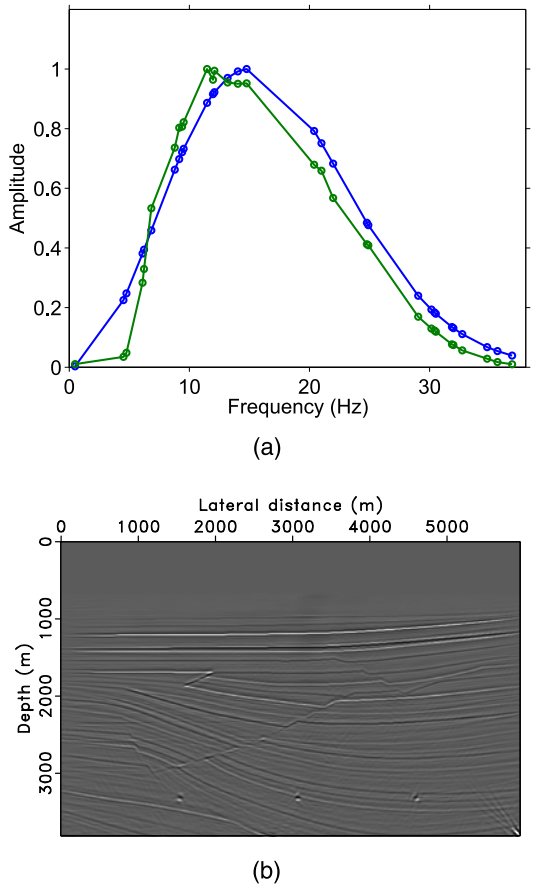

Figure 11. (a) Amplitude spectrum of the estimated source wavelet and (b) the inverted model perturbations, obtained using the background model that includes the strong ocean-bottom layer.

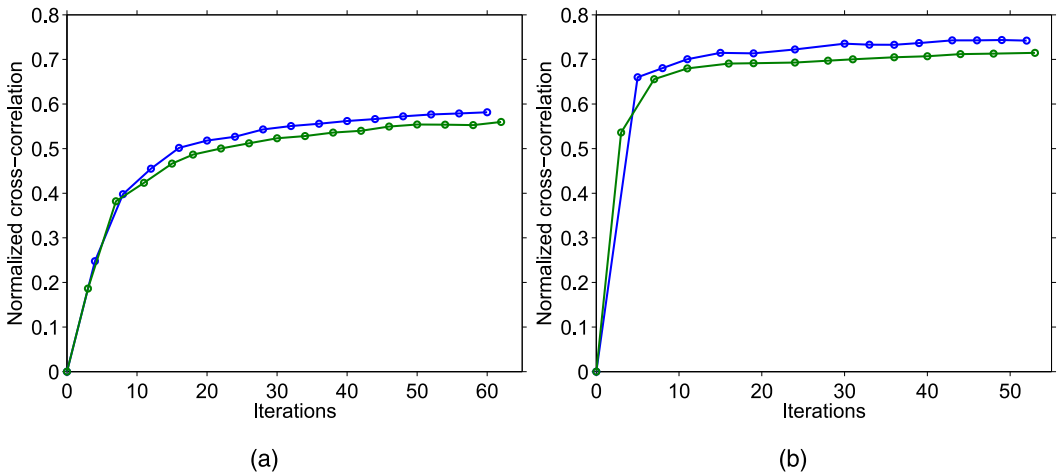

Figure 12. Model misfit represented by the normalized cross-correlation (NCC) as a function of the number of iterations. Each dot indicates the finish of one LASSO subproblem. (a) NCC curves for the imaging-with-primary examples in Fig. 4, using the true wavelet (in blue), and with source estimation (in green). (b) NCC curves for the imaging-with-multiple examples in Fig. 9, using the true wavelet (in blue) and with source estimation (in green).

4.4 Robustness to wavelet initial guess

We use an impulsive source as the initial guess of the source wavelet in all examples shown in this paper and get reasonable recoveries of the seismic images as well as the source wavelets. In this section, we would like to study the performance of the proposed method when the initial guess of the source wavelet contains notable phase errors. Robustness of the proposed method to these phase errors is desirable. We use the same experiment setup as the imaging-with-multiples examples. We show the two erroneous initial guesses of the source wavelet in Fig. 13(a), and show their corresponding solution paths in terms of model errors in Fig. 13(b). We can see despite that both initial guesses of

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

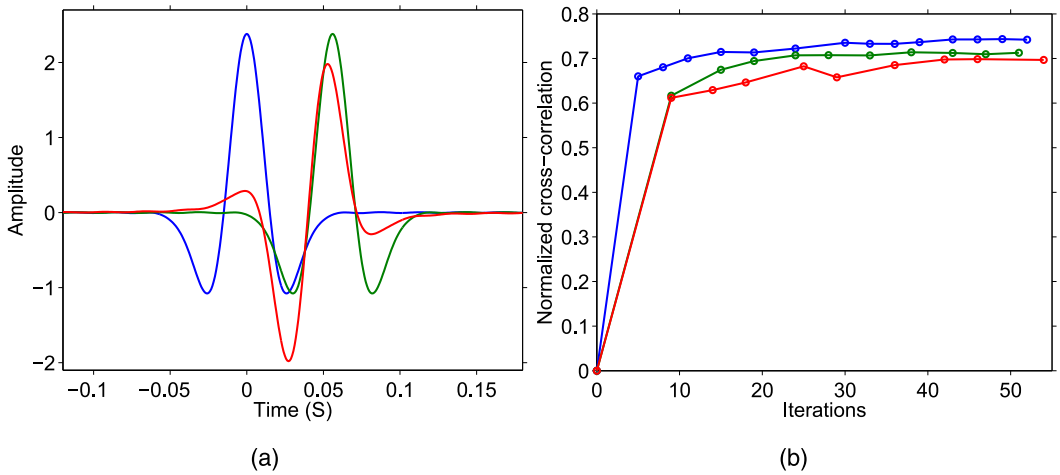

Figure 13. (a) The true source wavelet and the two erroneous initial guesses of the wavelet. The initial guess in green has a significant time shift, and the one in red has a 90° phase rotation plus a time shift. (b) The corresponding solution paths in terms of model errors measured using NCC.

the source wavelet can lead to local minima, the solution paths as well as the final model errors (and hence the final estimated source wavelets) are comparable to the case where we use an impulsive initial guess, demonstrating the robustness of the proposed method with respect to errors in the initial guess of the wavelet.

5 DISCUSSION

5.1 Variable projection and alternating optimization

As the optimization approach used to solve the reduced formulation by variable-projection, that is, eq. (7), resembles the optimization method that alternates between two variables (e.g. used in EPSI, van Groenestijn & Verschuur 2009; Lin & Herrmann 2013), one may wonder about the difference between the two terms. To the authors' knowledge, the term 'variable projection' says more about the problem formulation, that is, to eliminate one variable that the forward model linearly depends on (in our case the unknown source wavelet) using the pseudo inverse (i.e. the least-squares solution), which leads to a formulation that is in the form of a projection onto the orthogonal complement of the column space of the forward model (Golub & Pereyra 2003; Aravkin & van Leeuwen 2012); the term 'alternating' optimization says more about the algorithmic procedure used to solve a nonlinear least-squares formulation.

Actually, solving the reduced formulation by variable projection inevitably involves the alternation of the two variables. Golub & Pereyra (2003) stated that 'the variable projection algorithm consists of first minimizing for the nonlinear parameters (which also applies to the linear case—in our case it is the unknown model parameters $\mathbf{x}$) and then using the optimal value obtained for the nonlinear parameters to solve for the linear parameters'. The key of solving the reduced formulation is that, the pseudo-inverse (i.e. the least-squares solution) instead of an alternating gradient descent should be computed for the linear variable each time the nonlinear parameters are updated.

5.2 A note on the 'true-amplitude' seismic imaging

From the above discussion, we can see that unless the true source wavelet is known, obtaining the true-amplitude seismic image is out of the question when primary-only data are used due to the scaling ambiguity. Incorporating surface-related multiples in the inversion formulation indicates that we can resolve the ambiguity, that is, we get both the source estimates and the image with definite amplitudes. However, these definite amplitudes are the 'true' amplitudes only if the linearized modelling engine can accurately reproduce the input wavefield given the true source wavefield. In practice, many factors hinder this requirement from being satisfied, for example, scaling of the modelling engine, scaling of the SRME-type multiple-prediction to match the true amplitudes of multiples, and all sorts of coherent 'modelling errors' in the data. Nevertheless, the presented method provides a viable approach for the least-squares type of seismic imaging algorithm with an unknown source wavelet to be one step closer towards the retrieval of true amplitudes.

5.3 Modelling errors related to the free-surface

Data pre-processing, as we discussed in the paragraph following eq. (12), is of utmost importance for imaging with surface-related multiples based on the SRME relation. Because we formulate our imaging-with-multiple problem using a data-fitting objective, correct modelling of primaries and surface-related multiples in the presence of a free surface is the foundation for successful inversions.

The SRME relation requires that we only remove the receiver ghosts by separating the up- and downgoing wavefields (receiver ghosts are the downgoing wavefield), but retain the source ghosts in the data. However, as the Green's function in the SRME relation should be

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

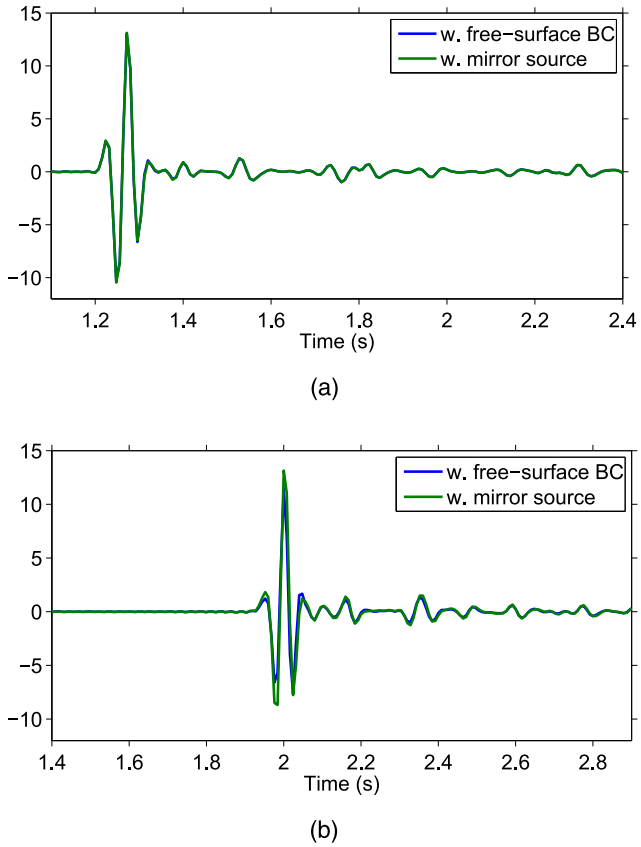

Figure 14. Comparison of two primary data sets that contain source ghosts modelled with a free surface boundary condition (in blue, no multiples arrive by the end of the time axes in the figures) and with the mirror source (in green). (a) Zero-offset traces. (b) Traces at about 2.3 km receiver offset.

surface-free, we can only approximate the source ghosts using a 'mirror' source, that is, as explained above, we place an extra virtual source with the opposite sign of the true source (because the free-surface has a reflectivity close to -1) at a position that is symmetric to the true source w.r.t. the surface level (again, no actual surface exists as the Green's function is surface-free). Our synthetic example shows that the accuracy of this approximation is fairly high for zero-offset traces but decreases with the receiver offset. Fig. 14 shows comparisons of two primary data sets with source ghosts modelled with a free-surface and with the mirror source (the finite-difference scheme, except for the boundary condition at the surface, remains the same for these two data sets). Two zero-offset traces and two far-offset (about 2.3 km receiver offset) traces are shown. The error at farther offsets may be one cause for the errors in the estimated source wavelet shown in Fig. 10.

Our synthetic example also shows that multiple-prediction using the SRME relation (implicitly embedded in eq. 13) can be fairly accurate. We again compare two multiple data sets, one modelled with a free-surface (predicted primaries using the mirror-source approach, as shown above, are subtracted from the modelled total upgoing data to obtain a rough estimate of the multiples) and another one modelled using the SRME relation. Note that as primaries, these multiples are upgoing wavefields with source ghosts. Two zero-offset traces and two far-offset traces, after proper scaling (this scalar is one reason that source estimation for imaging-with-multiple is necessary), are shown in Fig. 15. For field data, lots of factors can affect successful applications of the SRME relation: missing near-offset data, acquisition aperture for multiple-prediction, cable-feathering, out-of-plane propagation effect, etc. (Dragoset & Jeričević 1998; van Groenestijn & Verschuur 2009). As a result, inaccurate multiple prediction can be a major source of errors for imaging with multiples.

5.4 Source estimation by pre-processing

The proposed method aims to obtain a high-resolution seismic image via sparsity-constrained least-squares migration, without the prior knowledge of the source wavelet. As mentioned previously, because reliable source estimates only exist at frequencies sampled in the last LASSO(τ) subproblem, this method does not replace source estimation methods during pre-processing, which may be needed for certain data processing applications.

When a source wavelet estimated during pre-processing is accurate enough, one can use it to image primary-only data using the compressive imaging method proposed by Herrmann & Li (2012). As shown in Figs 2(b) and 4(b), this approach produces high-fidelity image in this case. However, to obtain an accurate estimation of a source wavelet in pre-processing, for example, as used in full waveform inversion (Pratt 1999; Virieux & Operto 2009), can be computationally expensive by solving extra wave equations. By comparison, the proposed

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

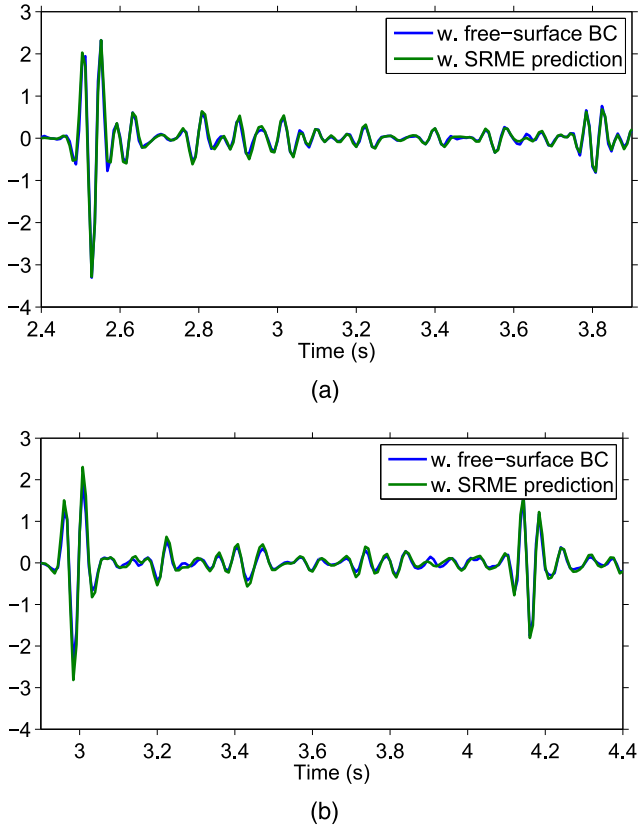

Figure 15. Comparison of two multiple data sets modelled with a free surface boundary condition (in blue) and using the SRME relation. (a) Zero-offset traces. (b) Traces at about 2.3 km receiver offset.

method has the advantage that the extra computation for the source estimates by eq. (6) is minimal compared to wave-equation solves, while the convergence of the model is still comparable to the case where we use the true source wavelet, as demonstrated in Fig. 12.

In the joint imaging of primaries and surface-related multiples, the proposed method leads to a 50 per cent computational overhead compared with the case where a properly scaled source wavelet is given—'properly scaled' in the sense that this wavelet leads to the separation of the primary and multiple wavefields. However, as in SRME or EPSI, this scaling of the source wavelet cannot be figured out in advance, but should be estimated as part of the inversion (Verschuur 2011).

6 CONCLUSIONS

We have proposed a fast least-squares imaging procedure that, unlike conventional RTM, does not require prior knowledge of the source wavelet which is difficult to accurately determine. We formulated this wavelet-free imaging procedure as a nonlinear sparse-inversion problem, by tightly integrating source estimation into the compressive imaging framework, and proposed to solve the problem using variable projection. Because of the blind-deconvolutional nature of this problem, applying variable projection alone could not address the scaling ambiguity in estimating the source wavelet and the seismic image. We therefore continued by resolving the issue through the incorporation of surface-related multiples in the inversion process (although at the expense of an increased computational cost). We subsequently provided a series of examples demonstrating that our method can produce highly accurate estimates of the source wavelet and high-resolution seismic images that are comparable to the ones obtained with the true source wavelet. We also provided examples where we overcome the scaling ambiguity by incorporating surface-related multiples, yielding definite amplitudes for the estimates of the source wavelet as well as for the seismic image. In all these examples, we controlled the simulation cost of the inversion to be more or less comparable to a single RTM with all the data, without compromising the quality of the images.

ACKNOWLEDGEMENTS

We would like to thank the authors of SPGℓ_1, Curvlab, and iWave. We also thank the authors of the SEG/EAGE salt model and the Sigsbee 2B model.

This work was in part financially supported by the Natural Sciences and Engineering Research Council of Canada via the Collaborative Research and Development Grant DNOISEII (375142-08). This research was carried out as part of the SINBAD II project which is

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

supported by the following organizations: BG Group, BGP, CGG, Chevron, ConocoPhillips, DownUnder GeoSolutions, Hess, Petrobras, PGS, Schlumberger, Sub Salt Solutions and 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 Probl.,* **28**(11), 115016, doi:10.1088/0266-5611/28/11/115016.

Aravkin, A.Y., van Leeuwen, T., Calandra, H. & Herrmann, F.J., 2012. Source estimation for frequency-domain FWI with robust penalties, in *74th EAGE Conference and Exhibition 2012,* p. P018, EAGE.

Aravkin, A.Y., Burke, J. & Friedlander, M., 2013. Variational properties of value functions, *SIAM J. Optim.,* **23**(3), 1689–1717.

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.

Candès, E.J. & Donoho, D.L., 2004. New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities, *Commun. Pure Appl. Math.,* **57**(2), 219–266.

Davydenko, M., Verschuur, D. & van Groenestijn, G., 2015. Full wavefield migration applied to field data, in *77th EAGE Conference and Exhibition 2015,* pp. We–N101–10, EAGE.

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

Esser, E., Lin, T.T., Wang, R. & Herrmann, F.J., 2015. A lifted ℓ_1/ℓ_2 constraint for sparse blind deconvolution, in *77th EAGE Conference and Exhibition 2015,* pp. Th–P6–07, EAGE.

Fomel, S., Backus, M., DeAngelo, M., Murray, P. & Hardage, B.A., 2003. Multicomponent seismic data registration for subsurface characterization in the shallow Gulf of Mexico, p. OTC 15117, Offshore Technology Conference, doi:10.4043/15117-MS.

Golub, G. & Pereyra, V., 1973. The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, *SIAM J. Numer. Anal.,* **10**(2), 413–432.

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

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

Herrmann, F.J., 2012. Pass on the message: recent insights in large-scale sparse recovery, in *74th EAGE Conference and Exhibition 2012,* p. B022, EAGE.

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

Herrmann, F.J., Moghaddam, P.P. & Stolk, C., 2008a. Sparsity- and continuity-promoting seismic image recovery with curvelet frames, *Appl. Comput. Harmon. Anal.,* **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.

Kaufman, L., 1975. A variable projection method for solving separable nonlinear least squares problems, *BIT Numer. Math.,* **15**(1), 49–57.

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering: Theory and Application,* pp. 206–220, eds Bednar, J.B., Redner, R., Robinson, E. & Weglein, A., SIAM.

Li, M., Rickett, J. & Abubakar, A., 2013. Application of the variable projection scheme for frequency-domain full-waveform inversion, *Geophysics,* **78**(6), R249–R257.

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

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 and Exhibition 2013,* pp. Th–01–05, EAGE.

Lu, S., Whitmore, D., Valenciano, A. & Chemingui, N., 2014. Enhanced subsurface illumination from separated wavefield imaging, *First Break,* **32,** 87–92.

Métivier, L., 2011. Interlocked optimization and fast gradient algorithm for a seismic inverse problem, *J. Comput. Phys.,* **230**(19), 7502–7518.

Métivier, L., Lailly, P., Delprat-Jannaud, F. & Halpern, L., 2011. A 2D nonlinear inversion of well-seismic data, *Inverse Probl.,* **27**(5), 055005, doi:10.1088/0266-5611/27/5/055005.

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.

Mulder, W.A. & Plessix, R.E., 2003. One-way and two-way wave-equation migration, in *SEG Technical Program Expanded Abstracts 2003,* pp. 881–884.

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.

Peters, B. & Herrmann, F.J., 2014. A sparse reduced hessian approximation for multi-parameter wavefield reconstruction inversion, in *SEG Technical Program Expanded Abstracts 2014,* pp. 1206–1210.

Peters, B., Herrmann, F.J. & van Leeuwen, T., 2014. Wave-equation based inversion with the penalty method – adjoint-state versus wavefield-reconstruction inversion, in *76th EAGE Conference and Exhibition 2014,* pp. Tu–E106–05, EAGE.

Pratt, R.G., 1999. Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model, *Geophysics,* **64**(3), 888–901.

Rickett, J., 2013. The variable projection method for waveform inversion with an unknown source function, *Geophys. Prospect.,* **61**(4), 874–881.

Robinson, E.A., 1967. Predictive decomposition of time series with application to seismic exploration, *Geophysics,* **32**(3), 418–484.

Stockham, T.G., Cannon, T.M. & Ingebretsen, R.B., 1975. Blind deconvolution through digital signal processing, *Proc. IEEE,* **63**(4), 678–692.

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

The SMAART JV, 2014. Sigsbee2B FS & NFS 2D synthetic datasets. Available at: http://www.delphi.tudelft.nl/SMAART/sigsbee2b.htm, last accessed February 2014.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso, *J. R. Stat. Soc. B,* **58**(1), 267–288.

Tu, N. & Herrmann, F.J., 2015a. Fast imaging with surface-related multiples by sparse inversion, *Geophys. J. Int.,* **201**(1), 304–317.

Tu, N. & Herrmann, F.J., 2015b. Fast least-squares imaging with surface-related multiples: application to a North-Sea data set, *Leading Edge,* **34**(07), 788–794.

Tu, N., Li, X. & Herrmann, F.J., 2013. Controlling linearization errors in ℓ_1 regularized inversion by rerandomization, in *SEG Technical Program Expanded Abstracts,* vol. 32, pp. 4640–4644.

Ulrych, T.J., Velis, D.R. & Sacchi, M.D., 1995. Wavelet estimation revisited, *Leading Edge,* **14**(11), 1139–1143.

van den Berg, E. & Friedlander, M.P., 2008. Probing the Pareto frontier for basis pursuit solutions, *SIAM J. Sci. Comput.,* **31**(2), 890–912.

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

van Leeuwen, T. & Mulder, W.A., 2009. A variable projection method for waveform inversion, in *71st EAGE Conference and Exhibition,* p. U024, EAGE.

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022

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.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics,* **74**(6), WCC1–WCC26.

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.

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

Zhang, D. & Schuster, G.T., 2014. Least-squares reverse time migration of multiples, *Geophysics,* **79**(1), S11–S21.

Downloaded from https://academic.oup.com/gji/article/205/3/1492/654727 by guest on 21 August 2022