Source-independent full wavefield converted-phase elastic migration velocity analysis

A. H. Shabelansky, A. E. Malcolm,[*] M. C. Fehler, X. Shang[†] and W. L. Rodi

Earth Resources Laboratory, Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
E-mail: andreys@mit.edu

SUMMARY

Converted phase (CP) elastic seismic signals are comparable in amplitude to the primary signals recorded at large offsets and have the potential to be used in seismic imaging and velocity analysis. We present an approach for CP elastic wave equation velocity analysis that does not use source information and is applicable to surface-seismic, microseismic, teleseismic and vertical seismic profile (VSP) studies. Our approach is based on the cross-correlation between reflected or transmitted PP and CP PS (and/or SS and CP SP) waves propagated backward in time, and is formulated as an optimization problem with a differential semblance criterion objective function for the simultaneous update of both *P*- and *S*-wave velocity models. The merit of this approach is that it is fully data-driven, uses full waveform information, and requires only one elastic backward propagation to form an image rather than the two (one forward and one backward) propagations needed for standard reverse-time migration. Moreover, as the method does not require forward propagation, it does not suffer from migration operator source aliasing when a small number of shots are used. We present a derivation of the method and test it with a synthetic model and field micro-seismic data.

Key words: Inverse theory; Body waves; Seismic tomography; Computational seismology; Wave propagation.

1 INTRODUCTION

In recent years, full waveform seismic imaging and velocity analysis methods have become standard and the use of elastic waves is now drawing more attention. Converted phase (CP) waves are an integrated part of the recorded elastic seismic signal and are investigated in numerous studies in the research areas of vertical seismic profile (VSP) data (e.g. Esmersoy 1990; Stewart 1991; Xiao & Leaney 2010), surface reflection (e.g. Purnell 1992; Stewart *et al.* 2003; Hardage *et al.* 2011) and transmission seismic data (e.g. Vinnik 1977; Vinnik *et al.* 1983; Bostock *et al.* 2001; Rondenay *et al.* 2001; Brytic *et al.* 2012; Shang *et al.* 2012; Shabelansky *et al.* 2013). In particular, for example Xiao & Leaney (2010) and Shang *et al.* (2012) showed that the CP seismic images can be calculated using one elastic propagation without using source information (i.e. location, mechanism and time-function). Source information is generally considered mandatory in standard seismic imaging and velocity analysis. However, in

passive monitoring source information is generally not available and in active source surveys seismic data require special treatment for frequency matching due to coupling differences between soil and vibro-seis or dynamite casing. These factors affect the accuracy of the imaging and velocity estimation and add computational and processing cost. Moreover, CP elastic seismic imaging is shown to have higher resolution in Xiao & Leaney (2010) and fewer artifacts than reflection type imaging in Shabelansky *et al.* (2012).

In this study, we present a source-independent CP (SICP) velocity analysis method that is formulated based on the framework of CP imaging and wave equation migration velocity analysis (WEMVA; e.g. Biondi & Sava 1999; Sava & Biondi 2004; Shen 2004, 2012; Albertin *et al.* 2006). We refer to this method as SICP-WEMVA. Like WEMVA, SICP-WEMVA depends strongly on starting (initial) velocity models (*P*- and *S*-wave speeds) and optimization algorithms with their parameters. SICP-WEMVA is typically less sensitive to the cycle skipping that is a problem for full waveform inversion, but it has lower resolution than full waveform inversion. The objective functional for converted *P* to *S* phases in source-dependent WEMVA (i.e. CP-WEMVA using forward/source propagation for *P* wave and backward propagation for *S* wave) appears to be convex (Yan 2010), and thus we expect SICP-WEMVA to also exhibit favourable properties for estimating large-scale velocity models. Unlike WEMVA, SICP-WEMVA (and CP-WEMVA) uses interference between

[*]Now at: Earth Sciences Department, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada.
[†]Now at: Shell Exploration and Production Int. Company Inc., 900 Louisiana Street, Houston, TX 70002, USA.

different wave types (i.e. *P* and *S* waves), and thus the resolution and stability of the two methods are not the same. Also, since SICP-WEMVA back-propagates data solely from receivers and does not depend on source information, it can be performed locally in the vicinity of the receivers only. This reduces the computational cost of iterative velocity analysis.

This paper is divided into three parts. In the first part, we present a SICP imaging condition in an extended domain. We use this to formulate an objective functional for gradient-based optimization. In the second part we present the derivation of the SICP-WEMVA velocity model optimization scheme. In the third part we show results of applying SICP-WEMVA to a synthetic model and its application to field data from a geothermal reservoir with abundant natural and induced seismicity. We also include three appendices addressing details of SICP-WEMVA.

2 THEORY OF SICP-WEMVA

SICP-WEMVA is a gradient-based iterative optimization scheme whose residuals are calculated from extended-domain migrated images. The gradients are formed from the backward and forward propagating in time seismic data and image residuals. We present below the derivation of the method, a summary of the algorithm and its practical implementations.

2.1 Extended SICP imaging condition (ESICP-IC)

We start our derivation using the isotropic elastic wave equation

$$
\ddot{\underline{u}} = \frac{\lambda + 2\mu}{\rho} \nabla \nabla \cdot \underline{u} - \frac{\mu}{\rho} \nabla \times \nabla \times \underline{u} + \frac{1}{\rho} \nabla \lambda (\nabla \cdot \underline{u})
$$
$$
+ \frac{1}{\rho} \nabla \mu \cdot \left[(\nabla \underline{u}) + (\nabla \underline{u})^T \right], \tag{1}
$$

where $\underline{u}(\underline{x}, t)$ and $\ddot{\underline{u}}(\underline{x}, t)$ are the displacement and acceleration vector wavefields, $\lambda(\underline{x})$, $\mu(\underline{x})$ and $\rho(\underline{x})$ are the two Lamé parameters and density, ∇, $\nabla\cdot$ and $\nabla\times$ are the gradient, divergence and curl and $\underline{x} = (x, y, z)$, and t are the spatial and time variables, respectively. The right hand side of eq. (1) consists of four terms: two with Lamé parameters and two with their gradients. The terms with the gradients are significant only at interfaces/discontinuities in subsurface medium and they are responsible for generation of reflected, transmitted and CP seismic data.

For the purpose of imaging, we assume smooth Lamé parameters (i.e. taking only the first two terms on the right-hand side of eq. 1), and obtain (Aki & Richards 2002, p. 64)

$$
\ddot{\underline{u}} = \hat{\alpha} \nabla \nabla \cdot \underline{u} - \hat{\beta} \nabla \times \nabla \times \underline{u}, \tag{2}
$$

where the parameters $\hat{\alpha}(\underline{x})$ and $\hat{\beta}(\underline{x})$ are defined through the *P*- and *S*-wave velocities, $\alpha(\underline{x})$ and $\beta(\underline{x})$, as

$$
\hat{\alpha} = \alpha^2 = \frac{\lambda + 2\mu}{\rho}, \qquad \hat{\beta} = \beta^2 = \frac{\mu}{\rho}. \tag{3}
$$

Since we use the isotropic elastic wave equation, the acceleration wavefield can be decomposed as $\ddot{\underline{u}} = \ddot{\underline{u}}_p + \ddot{\underline{u}}_s$, where

$$
\ddot{\underline{u}}_p(\underline{x}, t) = \hat{\alpha}(\underline{x}) \nabla \nabla \cdot \underline{u}(\underline{x}, t), \quad \ddot{\underline{u}}_s(\underline{x}, t) = -\hat{\beta}(\underline{x}) \nabla \times \nabla \times \underline{u}(\underline{x}, t). \tag{4}
$$

Then, the SICP-IC for N_e sources (i.e. explosions or earthquakes) is given as the zero lag in time cross-correlation between the back-propagated *P* and *S* acceleration vector-wavefields, $\ddot{\underline{u}}_p$ and $\ddot{\underline{u}}_s$,

$$
I(\underline{x}) = \sum_j^{N_e} \int_T^0 \ddot{\underline{u}}_p^j(\underline{x}, t) \cdot \ddot{\underline{u}}_s^j(\underline{x}, t) \, dt, \tag{5}
$$

where $\cdot$ is the dot product between vector components (e.g. vertical, radial and transverse), the superscript j refers to the source index, and T is the maximum recording time; it is at the lower limit of the integral (i.e. the data are propagated backward in time).

This imaging condition has three very important properties: first, no source information (i.e. location, mechanism and time-function) is required; second the image can be constructed only in the vicinity of the receivers (i.e. far from the sources); and third, the image can be constructed during the backward propagation and no wavefield storage is required. These properties reduce the computational cost and memory storage, and more importantly improve the image quality in comparison with standard reverse-time migration (RTM). Note also that we decompose the *P*- and *S*-wavefields using eq. (4) rather than using the more computationally efficient Helmholtz decomposition. The reason for this choice stems from the fact that the separated wavefields, using eq. (4), have consistent amplitude polarity for imaging, unlike those obtained by Helmholtz decomposition, which require additional treatment for signal amplitude (e.g. Du *et al.* 2012).

To provide intuition for the imaging condition in eq. (5), we show in Fig. 1(a) a snapshot of an elastic wave propagation from a
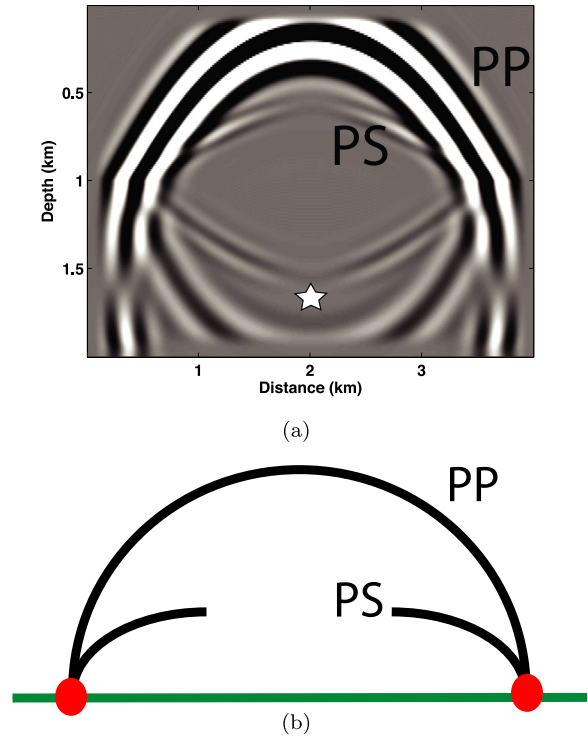

(a)

(b)

Figure 1. (a) A snapshot of the *Z*-component elastic wavefield generated from an isotropic point source at the position $(x, z) = (2.0, 1.7)$ km, marked with a star, and propagating through a horizontal interface at a depth of 1 km. (b) Schematic illustration of the wavefields shown in (a) above the horizontal interface. The black curves refer to the transmitted *P* and *S* wavefields (PP and PS) through the interface, marked by horizontal green line. The big red dots mark the points where the two wavefields interfere constructively and an image is formed.
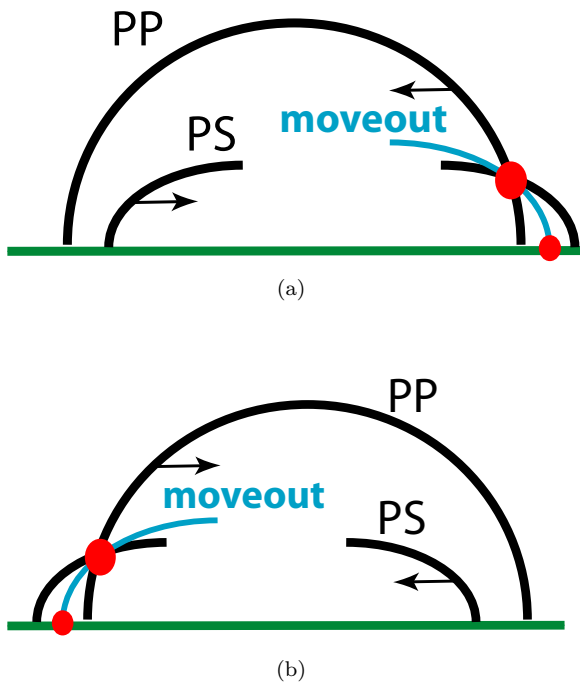
Figure 2. Schematic illustration of the construction of the space-lag CIG along x direction, h_x, using ESICP-IC (eq. 6). The black curves are as in Fig. 1(b) and the big red dots in (a) and (b) mark the current points of interference between the wavefields while the small red dots show the $h_x = 0$ position. The blue curve connecting the big and small red dots follows the points of interference and thus is called the moveout curve. This moveout is obtained by shifting the wavefields in opposite directions; these directions are marked with arrows in (a) and (b), and the length of the shift is called the space-lag.

single point source, through a horizontal interface. Fig. 1(b) illustrates the wavefields schematically above the interface. The red dots in Fig. 1(b) correspond to constructively interfering image points at which the energy of the zero-lag in time cross correlation between *P*- and *S*-wavefields is maximized. When we sum over multiple sources, N_e, in eq. (5), the signal-to-noise ratio (S/N) of the interfering image points increases.

This concept is valid when the *P*- and *S*-wave speed models are correct. However, when the velocities have error, the interference between *P* and *S* waves does not occur at the correct image points and energy leaks to adjacent points. To quantify the leak and interference mispositioning, we introduce an extended source-independent converted-phase imaging condition (ESICP-IC) in the so-called subsurface space lag, $\underline{h} = (h_x, h_y, h_z)$, as

$$I(\underline{x}, \underline{h}) = \sum_j^{N_e} \int_T^0 \ddot{u}_p^j(\underline{x} - \underline{h}, t) \cdot \ddot{u}_s^j(\underline{x} + \underline{h}, t)\, \mathrm{d}t. \qquad (6)$$

The concept of extension is adopted from acoustic reflection imaging where the extended image is typically called a common image gather (CIG), and the subsurface space lag, $\underline{h}$, is called the subsurface offset (e.g. Rickett & Sava 2002). However, as source information is not involved in SICP-IC and ESICP-IC, the notion of offset is not well defined.

In Fig. 2 we show schematically the construction of these image gathers, given by eq. (6), in 2-D for a single source and horizontal space lag, h_x [i.e. we show $I(x = x_f, z; h_x, h_z = 0)$ where x_f is a fixed horizontal image point] that are obtained by continuously

shifting the *P*-wavefield (marked PP) to the right and *S*-wavefield (marked PS) to the left, as marked with the arrows in Fig. 2(a). The waves interfere at new points during the shift (i.e. space-lag); these interference points are marked with a blue line and are called the moveout. The small red dot in Fig. 2(a) corresponds to the initial interference point, marked in Fig. 1(b) by the big red dot, and the big red dot in Fig. 2(a) marks the current point of interference along the moveout. Fig. 2(b) shows the same schematic for negative horizontal space lags, when the *P*-wavefield is shifted to the left and *S*-wave to the right. Note that the curved moveout can appear linear for small lags.

To illustrate numerically the construction of the moveout in the space lag gathers for the same horizontal interface model, discussed for Figs 1 and 2, we generate a set of isotropic *P* sources from below the horizontal layer at a depth of 1.7 km. We construct the gathers using forward propagated *P* and *S* wavefields. In Figs 3(a) and (b), we show moveouts for an image point at $x_f = 2$ km obtained from sources generated at $(x, z) = (1.5, 1.7)$ km and $(x, z) = (2.5, 1.7)$ km, below the interface, respectively. These moveouts verify the behaviour presented in Figs 2(a) and (b) for correct *P*- and *S*-wave velocity models. The summation of the two gathers (Figs 3a and b) is shown in Fig. 3(c), where we observe that only the energy around $h_x = 0$ at the depth of the layer interferes constructively. By adding more sources, generated from different horizontal positions below the interface, we obtain more focusing of the energy around $h_x = 0$ (see result in Fig. 3d obtained with 14 sources).

As mentioned above, this focusing is achieved only when the *P*- and *S*-wave velocities are correct. When the velocities have error, the energy is no longer focused. To demonstrate this, we add to the correct *S*-wave model an elliptical lens (i.e. anomaly) with a maximum perturbation of 15 per cent of the background model and perform the same numerical experiment. Fig. 4(a) shows a result obtained with the same 14 sources as that shown in Fig. 3(d) with the incorrect *S*-wave velocity. The idea behind the SICP-WEMVA is to design an optimization procedure that will minimize the energy outside of $\underline{h} = 0$ (see the residual gather for h_x in Fig. 4b) by updating the *P*- and *S*-velocity models. This optimization procedure is discussed in the next section.

2.2 Derivation of the SICP-WEMVA optimization

To relate the deviation of energy from $\underline{h} = 0$ in the extended subsurface space lag image gather (shown in Fig. 4) to the error in velocity, we formalize a gradient-based optimization problem by minimizing the energy at $\underline{h} \neq 0$. We use the differential semblance criterion (Symes & Carazzone 1991; Shen 2004, 2012) for the objective functional, *J*, as

$$J = \frac{1}{2} \int \int_{-H}^{H} \underline{h}^2 I^2(\underline{x}, \underline{h})\, \mathrm{d}\underline{h}\, \mathrm{d}\underline{x}, \qquad (7)$$

where $\underline{H} = (H_x, H_y, H_z)$ is the maximum subsurface space-lag, and $\underline{h} I(\underline{x}, \underline{h})$ is the residual subsurface space-lag CIG; Fig. 4(b) shows a 2-D CIG along the horizontal *x* direction, h_x, with $H_x = 9.5$ km. The choice of maximum space lag is addressed in Appendix A. In general, we add additional regularization terms to eq. (7) as discussed in for example, Shen & Symes (2008) and Shen (2012). These terms do not affect the understanding of the derivation and so will not be further discussed.
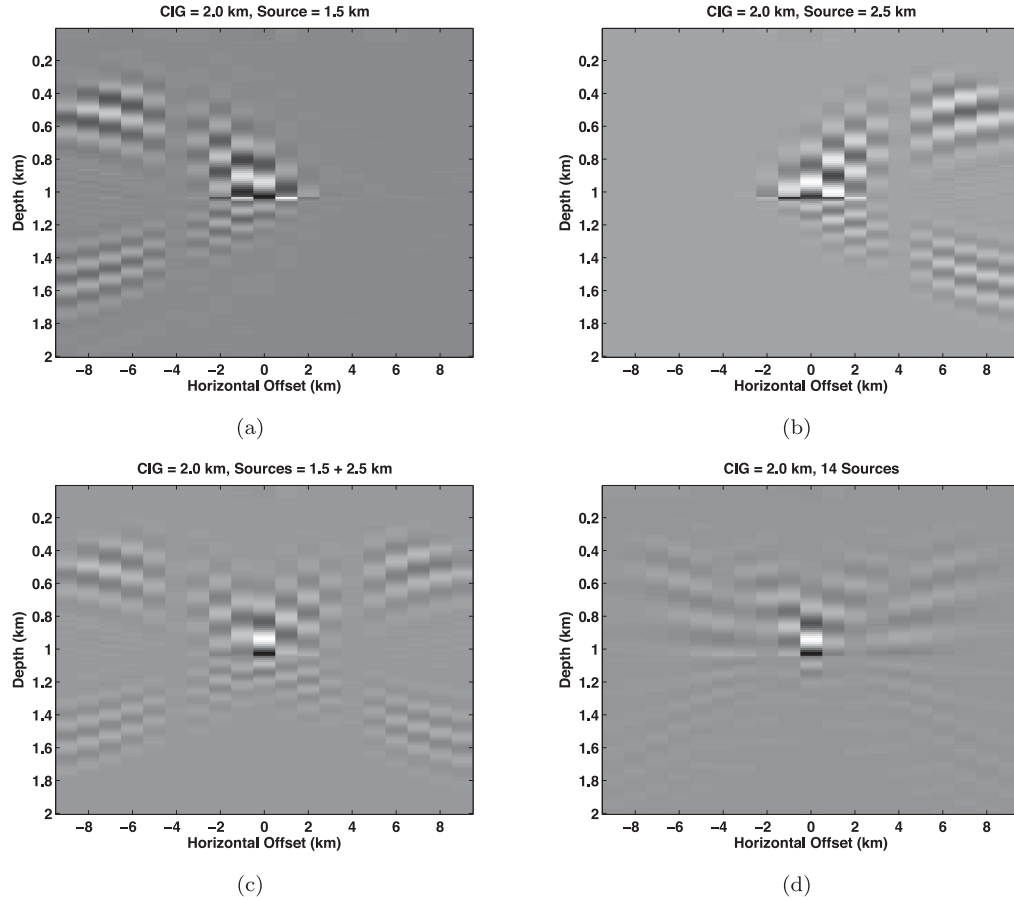
Figure 3. Converted phase subsurface space lag CIGs at $x_f = 2$ km, $I(x = x_f, z, h_x, h_z = 0)$, obtained with the correct velocity models from a source at position (a) $(x, z) = (1.5, 1.7)$ km, and (b) $(x, z) = (2.5, 1.7)$ km. (c) The summation of the two subsurface space lag CIGs from (a) and (b). (d) The summation over 14 sources generated at the depth of $z = 1.7$ km with a horizontal increment of 0.25 km.
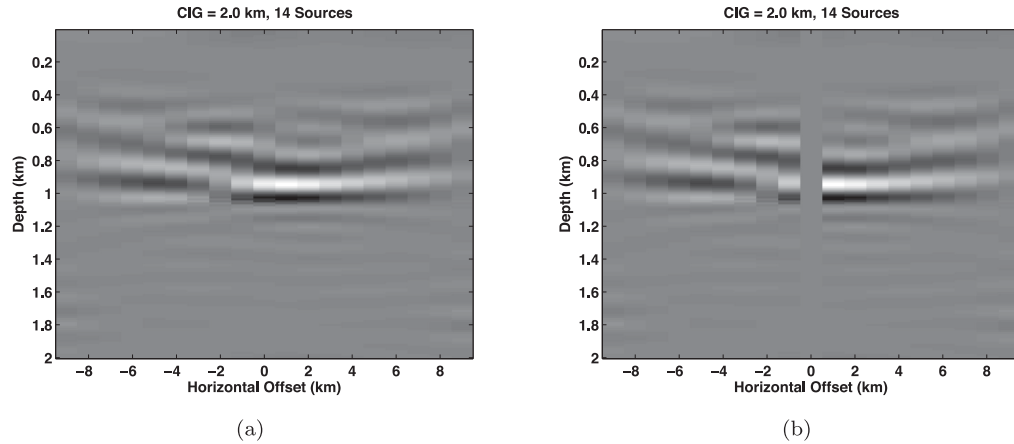

Figure 4. (a) Converted phase subsurface space lag CIGs at $x_f = 2$ km, $I(x = x_f, z, h_x, h_z = 0)$, obtained with the incorrect S-wave velocity with the same shots that are shown in Fig. 3(d). (b) The residual gather used for the SICP-WEMVA (i.e. signal at $h_x = 0$ was removed).

To formulate our problem as a gradient-based optimization, we use perturbation theory to calculate the gradients of the objective function with respect to the model parameters. We seek to obtain

$$\delta J = \int \left(K_{\hat{\alpha}}(\underline{x}) \delta \hat{\alpha}(\underline{x}) + K_{\hat{\beta}}(\underline{x}) \delta \hat{\beta}(\underline{x}) \right) d\underline{x}, \qquad (8)$$

where $K_{\hat{\alpha}}$ and $K_{\hat{\beta}}$ are the sensitivity kernels, associated with gradients of the objective function, J, that is perturbed with respect to

model parameters, $\hat{\alpha}$ and $\hat{\beta}$, respectively. The model parameters $\hat{\alpha}$ and $\hat{\beta}$ are defined in eq. (3).

The sensitivity kernels (gradients) of eq. (8) are derived in detail in Appendix B, and have the following convolutional forms

$$K_{\hat{\alpha}}(\underline{x}) = \frac{-1}{\hat{\alpha}(\underline{x})} \sum_j^{N_e} \int_T^0 \underline{\ddot{u}}_p^j(\underline{x}, t) \cdot \underline{v}_p^j(\underline{x}, T - t) \, dt \qquad (9)$$

and

$$K_{\hat{\beta}}(\underline{x}) = \frac{1}{\hat{\beta}(\underline{x})} \sum_j^{N_e} \int_T^0 \underline{v}_s^j(\underline{x}, t) \cdot \underline{\ddot{u}}_s^j(\underline{x}, T-t) \, dt, \qquad (10)$$

where the vector wavefields $\underline{v}_p$ and $\underline{v}_s$ are calculated by forward propagation using adjoint sources

$$\underline{v}_p^j(\underline{x}, T-t) = \left(L^{-1}\right)^* \nabla\nabla \cdot \int_{-H}^H \hat{\alpha}(\underline{x}+\underline{h})\underline{\ddot{u}}_s^j(\underline{x}+2\underline{h}, t)$$
$$\times R(\underline{x}+\underline{h}, \underline{h}) \, d\underline{h} \qquad (11)$$

and

$$\underline{v}_s^j(\underline{x}, T-t) = \left(L^{-1}\right)^* \nabla \times \nabla \times \int_{-H}^H \hat{\beta}(\underline{x}-\underline{h})\underline{\ddot{u}}_p^j(\underline{x}-2\underline{h}, t)$$
$$\times R(\underline{x}-\underline{h}, \underline{h}) \, d\underline{h}, \qquad (12)$$

where $(L^{-1})^*$ is the adjoint of the inverse of the isotropic elastic wave equation operator (here L^* is defined as forward propagation) and $R(\underline{x}, \underline{h}) = \underline{h}^2 I(\underline{x}, \underline{h})$ (see Appendix B for more details).

To reduce computational burden, we replace $\nabla\nabla \cdot$ in eq. (11) and $\nabla \times \nabla \times$ in eq. (12) by respective weights a and b (that are obtained from analytic derivation of spatial derivatives), and calculate the adjoint sources as

$$\underline{v}_p^j(\underline{x}, T-t) = \left(L^{-1}\right)^* a \int_{-H}^H \hat{\alpha}(\underline{x}+\underline{h})\underline{\ddot{u}}_s^j$$
$$\times (\underline{x}+2\underline{h}, t) R(\underline{x}+\underline{h}, \underline{h}) \, d\underline{h} \qquad (13)$$

and

$$\underline{v}_s^j(\underline{x}, T-t) = \left(L^{-1}\right)^* b \int_{-H}^H \hat{\beta}(\underline{x}-\underline{h})\underline{\ddot{u}}_p^j$$
$$\times (\underline{x}-2\underline{h}, t) R(\underline{x}-\underline{h}, \underline{h}) \, d\underline{h}. \qquad (14)$$

With these gradients of the objective function, we set up the standard gradient-based optimization scheme as

$$\underline{m}_{i+1} = \underline{m}_i + v_{i_k}\underline{p}_i, \qquad (15)$$

where the model parameter vector $\underline{m} = (\hat{\alpha} \ \hat{\beta})^T$ with $\hat{\alpha}$ and $\hat{\beta}$ sorted into row vectors, $\underline{p}$ is the line search direction column vector and v is the step length scalar. The superscript T denotes transpose and subscripts i and k refer to the indices of the iteration of the search direction and that of the step length, respectively. The efficient calculation of the step length, v, for SICP-WEMVA is given in Appendix C.

The search direction is given by the conjugate gradient method (Hestenes & Stiefel 1952) as

$$\underline{p}_0 = -C_0\underline{g}_0; \qquad (16)$$

$$\underline{p}_i = -C_i\underline{g}_i + \eta_i \underline{p}_{i-1}, \qquad (17)$$

where $\underline{g} = (K^{\hat{\alpha}} \ K^{\hat{\beta}})^T$, $K^{\hat{\alpha}}$ and $K^{\hat{\beta}}$ are the gradients found in eqs (9) and (10) and sorted into row vectors, C is a pre-conditioner (here we use it as a smoothing function), and scalar η_i is defined as in for example, Rodi & Mackie (2001) as

$$\eta_i = \frac{\underline{g}_i^T C_i \underline{g}_i}{\underline{g}_{i-1}^T C_{i-1}\underline{g}_{i-1}}. \qquad (18)$$

2.3 Algorithm

Each iteration of the proposed algorithm consists of the following steps:

(i) Propagate each elastic shot gather, j, backward in time, and store the acceleration wave fields $\underline{\ddot{u}}_p^j(\underline{x}, t)$ and $\underline{\ddot{u}}_s^j(\underline{x}, t)$.

(ii) Construct extended image gathers $I(\underline{x}, \underline{h})$ from all shots using eq. (6).

(iii) Construct residual extended image gathers: $R(\underline{x}, \underline{h}) = \underline{h}^2 I(\underline{x}, \underline{h})$.

(iv) Calculate $\underline{v}_p^j(\underline{x}, T-t)$ and $\underline{v}_s^j(\underline{x}, T-t)$ for each shot from the stored $\underline{\ddot{u}}_p^j(\underline{x}, t)$, $\underline{\ddot{u}}_s^j(\underline{x}, t)$ and constructed $R(\underline{x}, \underline{h})$.

(v) Construct the sensitivity kernels $K_{\hat{\alpha}}(\underline{x})$ and $K_{\hat{\beta}}(\underline{x})$ (gradients).

(vi) Update model parameters $\hat{\alpha}(\underline{x})$ and $\hat{\beta}(\underline{x})$ using the gradients (sensitivity kernels).

2.4 Practical considerations and implementation

There are three important practical considerations that need to be addressed during the SICP-WEMVA calculation. The first consideration is the smoothing and regularization of the gradients, $K_{\hat{\alpha}}$ and $K_{\hat{\beta}}$. Many studies address the problems associated with the construction and smoothing of the WEMVA gradient (e.g. Shen & Symes 2008; Fei & Williamson 2010; Vyas & Tang 2010; Shen & Simes 2013). In our study, we regularize the gradients in the so-called vertical time domain (e.g. Alkhalifah *et al.* 2001; Shabelansky 2007), which stabilizes and speeds up the convergence. The transformation for vertical time is given by

$$K_{\hat{\alpha}}(\underline{x}) \longrightarrow K_{\hat{\alpha}}(\underline{\tau}^\alpha)$$
$$K_{\hat{\beta}}(\underline{x}) \longrightarrow K_{\hat{\beta}}(\underline{\tau}^\beta), \qquad (19)$$

where the vertical times $\underline{\tau}^\alpha$ and $\underline{\tau}^\beta$ are defined as

$$\underline{\tau}^\alpha = \left(\tau_x^\alpha, \tau_y^\alpha, \tau_z^\alpha\right) = \int_0^{\underline{x}} \frac{d\underline{\xi}}{\alpha} \quad \text{and} \quad \underline{\tau}^\beta = \left(\tau_x^\beta, \tau_y^\beta, \tau_z^\beta\right) = \int_0^{\underline{x}} \frac{d\underline{\xi}}{\beta}.$$

The second consideration is the presence of the surface waves in the data and the excitation of false body-wave modes during the back propagation. The back-propagation of the surface waves may contaminate the image in the shallow part, as they cannot be separated into purely *P* and *S* waves. The false modes are generated due to the imperfect acquisition geometry, and become less severe with dense receiver coverage. Nevertheless, this effect is observable and will be addressed and illustrated below. To alleviate both of these effects and stabilize the optimization procedure, we apply muting in the space-lag domain. However, more sophisticated approaches may be beneficial.

The third practical consideration is related to the propagation of the adjoint *P*- and *S*-wave sources (eqs 13 and 14). In the 2-D examples below we propagate both sources simultaneously using only one propagation where the *P*-wavefield propagates in the *P*-SV plane and the *S*-wavefield propagates in the SH plane. This propagation is possible in 2-D as the wavefields are completely decoupled in 2-D. In 3-D, however, each adjoint source wavefield needs to be propagated separately.

3 NUMERICAL TESTS

To examine the proposed algorithm we test it with two data sets: a synthetic data set and a field data set from the passive source
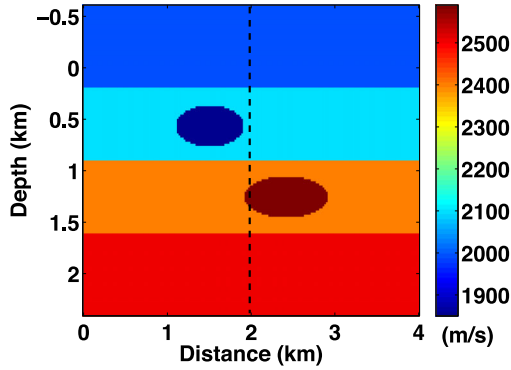
Figure 5. True *S*-wave speed used in the synthetic test. The dashed vertical line refers to the location of the CIG gathers, x_f, shown in Figs 6, 9 and 10.

microseismic monitoring of a geothermal area in Iceland. The acquisition geometry is transmission for both tests, that is the recording stations are placed on the surface and sources are located at depth (some earthquakes in the field data occurred outside of the imaging/computational region). All elastic wave solutions are modeled with a 2-D finite-difference solver, using a second order in time staggered-grid pseudo-spectral method with perfectly matched layer (PML) absorbing boundary conditions (Kosloff *et al.* 1984; Carcione 1999; Marcinkovich & Olsen 2003).

3.1 Synthetic test

We test SICP-WEMVA with a synthetic *S*-wave speed model, shown in Fig. 5, and constant *P*-wave speed and density of 4500 m s^{-1} and

2500 kg m^{-3}, respectively. The number of gridpoints in the model is $N_z = 150$ and $N_x = 200$, and the spatial increments are $\Delta x = \Delta z = 20$ m. We generate 35 isotropic (explosive) sources equally distributed at 1.8 km depth with horizontal increment of 100 m using a Ricker wavelet with a peak frequency of 45 Hz and Δt of 0.001 s. The data are recorded with two-component receivers that are equally distributed and span the same computational grid at a depth of 0 km (i.e. $\underline{x}_{obs} = (x, 0)$ km). Figs 6(a) and (b) show representative common shot gathers for *Z* and *X* components, $u_z(\underline{x}_{obs}, t)$ and $u_x(\underline{x}_{obs}, t)$, from a source at $(x, z) = (2.0, 1.8)$ km. In Fig. 6(c), we also present the *P* component of the seismic shot gather, calculated by $\nabla \cdot \underline{u}(\underline{x}_{obs}, t)$, to support the assumption made in eq. (B16) (Appendix B) that the *P* wavefield is barely affected by a contrast in the *S*-wave speed, evidenced by the lack of energy arriving after the direct *P* arrival.

Having calculated the elastic seismic data (Fig. 6), we test SICP-WEMVA with the smoothed background *S*-wave speed model, shown in Fig. 7(a) (i.e. elliptical velocity inclusions were omitted). In Fig. 7(b), we show the *S* wave model after five iterations. We observe that the smooth part of the true velocity model is reconstructed. To compare the effect of the inverted velocity model on imaging, we calculate the CP images using the true, initial and inverted *S*-wave speeds, along with the smoothed true *S*-wave speed. In Fig. 8(c), we observe that the migrated image obtained using the inverted model shows significant improvement and gradual convergence towards the true image compared to that obtained with the initial model (Fig. 8b). We show the space-lag image gathers for the image point at $x_f = 2$ km in Figs 9 and 10. The strong energy extending linearly in depth (marked with green arrows in Fig. 9a)
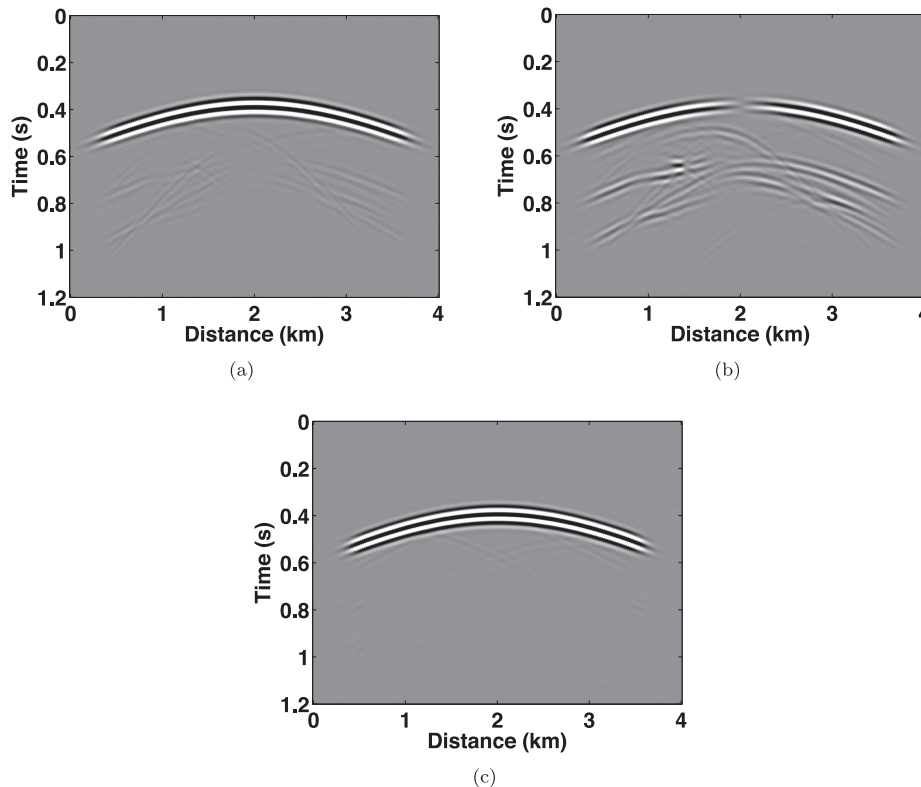

(a)

(b)

(c)

Figure 6. Common shot gathers calculated at the depth of 0 km due to an isotropic source from $(x, z) = (2.0, 1.8)$ km: (a) *Z*-component, $u_z(x, t)$; (b) *X*-component, $u_x(x, t)$; (c) *P*-component calculated during the propagation as $\nabla \cdot u$, shown to support the assumption made in eq. (B16) (Appendix B) that the *P*-wavefield is barely affected by a contrast in the *S*-wave speed, evidenced by the weak energy arriving at the times after the direct *P*-wave arrivals.
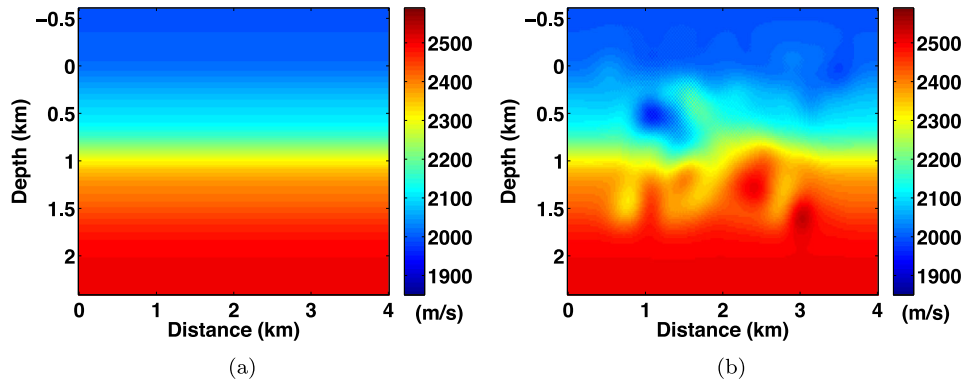
Figure 7. *S*-wave speed models: (a) initial and (b) after five iterations of SICP-WEMVA. The smooth part of the true velocity model (Fig. 5) is reconstructed in the inferred model.
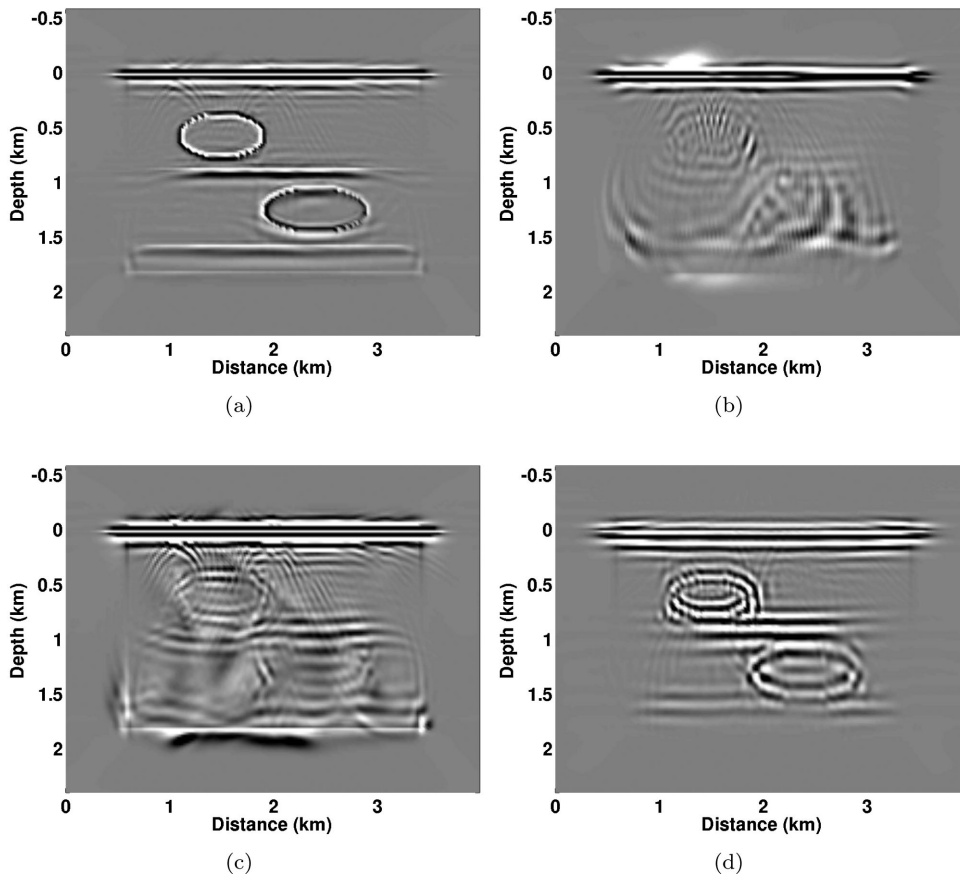

Figure 8. Results of the source-independent converted-phase elastic seismic imaging (SICP-IC) obtained with: (a) true, (b) initial and (c) inferred *S*-wave speeds that are shown in Figs 5 and 7. (d) SICP image obtained with the smoothed true *S*-wave velocity model to show the resolution of the expected image reconstruction with SICP-WEMVA. *P*-wave speed and density are constant, 4500 m s^{-1} and 2500 kg m^{-3}, respectively.

corresponds to the energy introduced by the false wave modes. In Fig. 10, we show the same gathers after zooming in and muting the false wave modes. Note that SICP-WEMVA was applied to the muted gathers. We observe that although some energy remains at the non zero-lags, the energy is more focused around zero lags in the gather calculated with the inverted model compared to that obtained with the initial *S*-wave speed.

The resolution of the velocity reconstruction depends on the sufficient sampling of the medium by both *P* and CP *S* (or/and *S* and CP *P*) back propagated wavefields (i.e. illumination) and is directly related to the imaging resolution through data frequency content

and receiver aperture. An analytic imaging resolution analysis for comparison between the PP and PS imaging was conducted in Xiao & Leaney (2010). This comparison highlighted the superiority of the spatial resolution for PS over the PP imaging.

3.2 Field data example

The second example uses passive source, microseismic field data from a geothermal area in a region of Iceland that has abundant natural and induced seismicity (see Fig. 11a). We test SICP-WEMVA
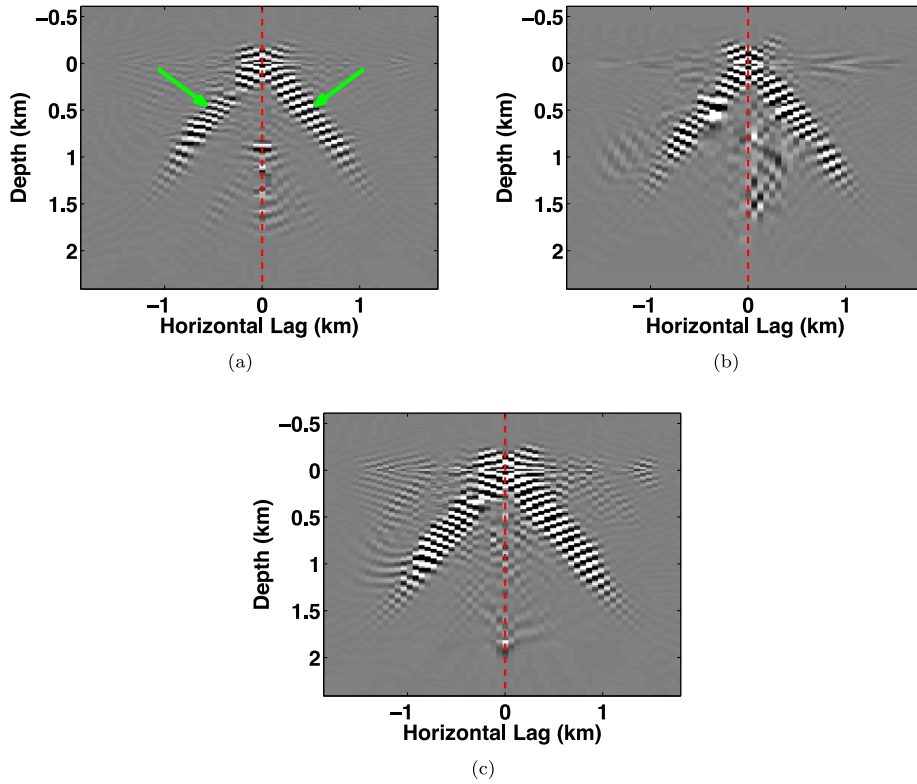
Figure 9. Converted-phase space-lag common image gathers with the (a) true, (b) initial and (c) inverted *S*-wave speed model. The strong energy linearly extended in depth, marked with green arrows in (a), corresponds to the false wave-mode oscillation.
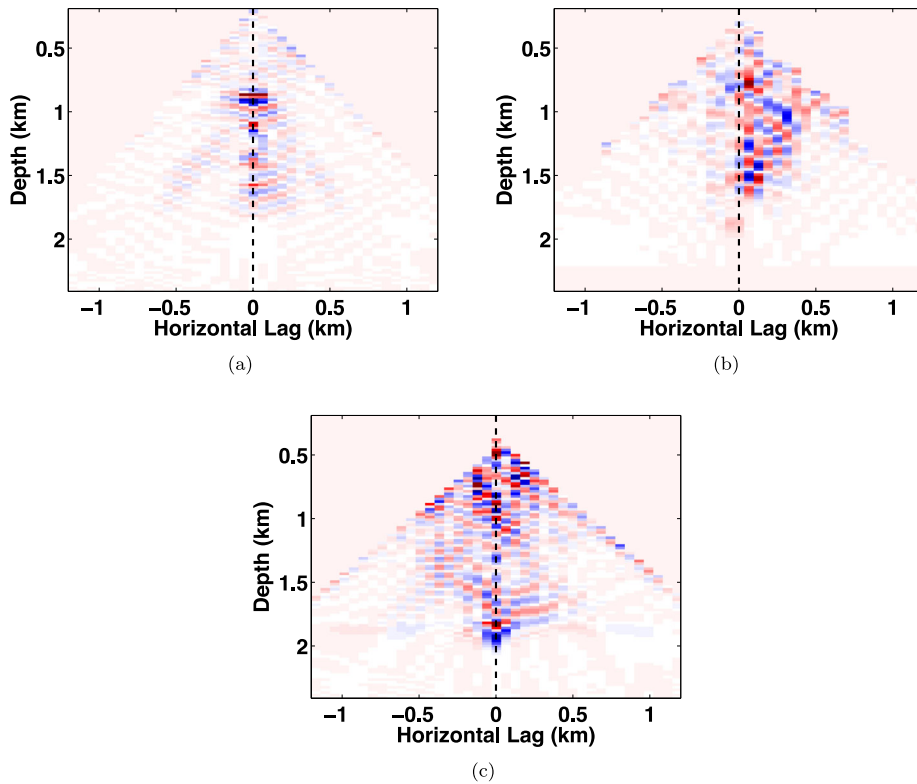

Figure 10. Converted-phase space-lag common image gathers (CIGs) from Fig. 9 shown in different colour after the false wave mode energy was muted out and the horizontal axis was zoomed in. The CIGs are obtained with (a) true, (b) initial and (c) inverted *S*-wave speed model after five iterations of SICP-WEMVA. The input for SICP-WEMVA is the residual image gathers after the false wave-mode energy has been muted out. Note that although some energy remains at the non-zero lags, the energy is more focused around the zero lags (black vertical dashed line) in the gather after the five iterations of SICP-WEMVA.
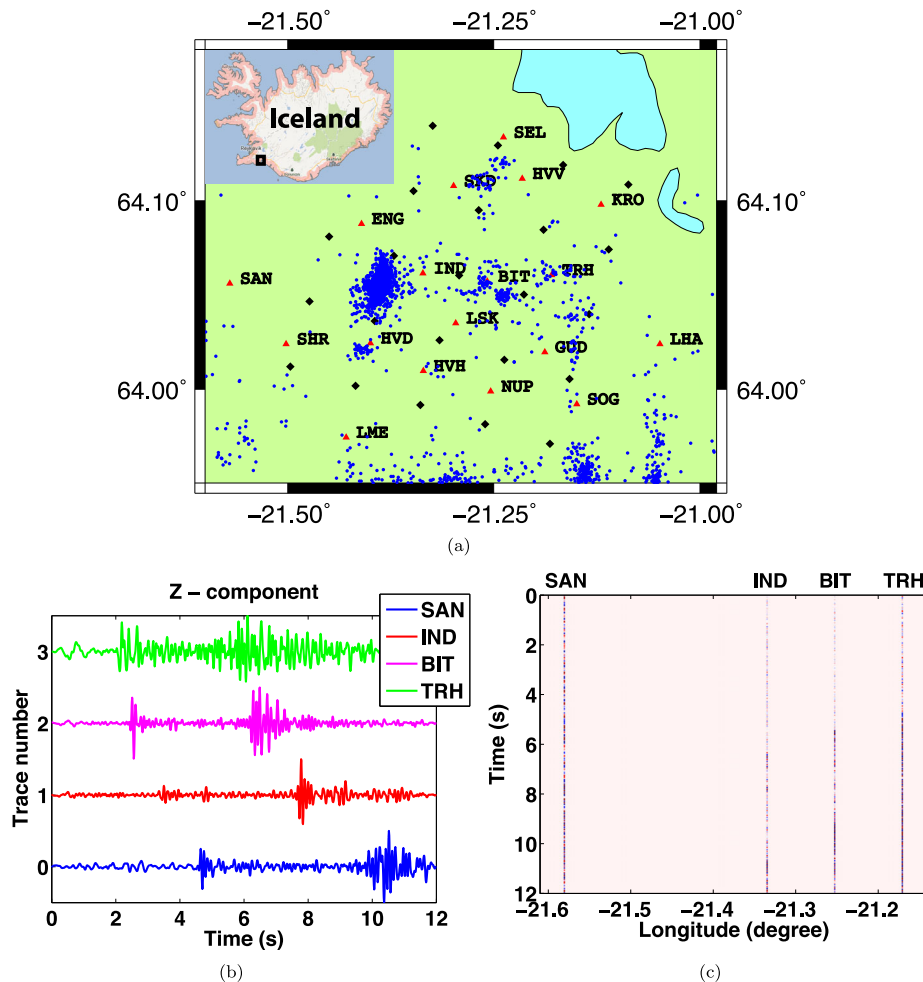
Figure 11. (a) Map of the investigated area in Iceland with microseismic field data. The blue dots mark the earthquake locations projected onto the surface, the red triangles refer to locations of the recording stations named with three letters, and the black diamonds denote the computational grid points of the regional traveltime tomography done by Tryggvason *et al.* (2002). (b) Representative normalized and filtered *Z*-component traces that are recorded at the surface along a 2-D line at the latitude of 64.06° and sorted into (c) a common shot gather. The relative position of the traces in the common shot gather corresponds to the relative position of their stations on the surface; the trace #0 from station SAN in (b) is the leftmost trace in (c).

with four stations (SAN, IND, BIT and TRH) along a 2-D line at the latitude of 64.06° and longitudes between −21.6° and −21.10°. In Figs 11(b) and (c), we show four representative traces sorted into a common shot gather (*Z*-component) from a single earthquake. Each trace is of a time record of $T = 12$ s and $\Delta t = 0.005$ s that was band-pass filtered between 2 and 12 Hz. The relative position of the traces in the common shot gather corresponds to the relative locations of the stations at the surface, $\underline{x}_{\mathrm{obs}} = (x, z_{\mathrm{surface}} = 0 \text{ km})$. The computational grid is $N_z = 90$ and $N_x = 300$ and the spatial increments are $\Delta z = 0.2$ km and $\Delta x = 0.15$ km. To construct an image (using eq. 5) of the geothermal area that is located between longitudes −21.4° and −21.2°, we use 32 events (i.e. earthquakes) of moment magnitude between 0.9 and 1.2. Initial *P*- and *S*-wave speeds were taken from the model of Tryggvason *et al.* (2002) obtained using regional-scale traveltime tomography. These velocity models are shown in Figs 12(a) and (b), along with the SICP image in Fig. 12(c). Note that the shallow part of the SICP image is completely contaminated by aliasing caused by the sparse station distribution on the surface and thus is muted out. By applying the SICP-WEMVA optimization for the deep part of the Earth (between 2.5 and 10 km), we update the *P*- and *S*-velocity models (Figs 12d and e) and the SICP image (Fig. 12f). We use 27 iterations of conjugate gradient for

updating the velocity models (see the convergence curve in Fig. 13a). We observe that in the area of the reservoir (longitudes between −21.38° and −21.22°) the *P*-wave velocity has slightly increased while the *S*-wave velocity decreased along vertical trends. Although this result indicates a high Vp/Vs and may suggest magma conduits, more study is needed to understand the geologic implications of this result. The image produced with the updated velocities (in Fig. 12f) shows more focused structure and reveals previously obscured structure (e.g. the region within the marked red ellipse). We also show in Fig. 13(b) the subsurface space-lag gather along the *x*-direction, h_x, at longitude of −21.34°, where we observe that the energy in the updated gather is more focused around zero space-lag than that obtained with the initial models. These gathers, along with the convergence curve, suggest that the SICP-WEMVA optimization converges towards a reasonable solution.

4 DISCUSSION

In this section, we address the most pressing assumptions underlying SICP-WEMVA and discuss their implications for the accuracy, stability, robustness and applicability of the method.
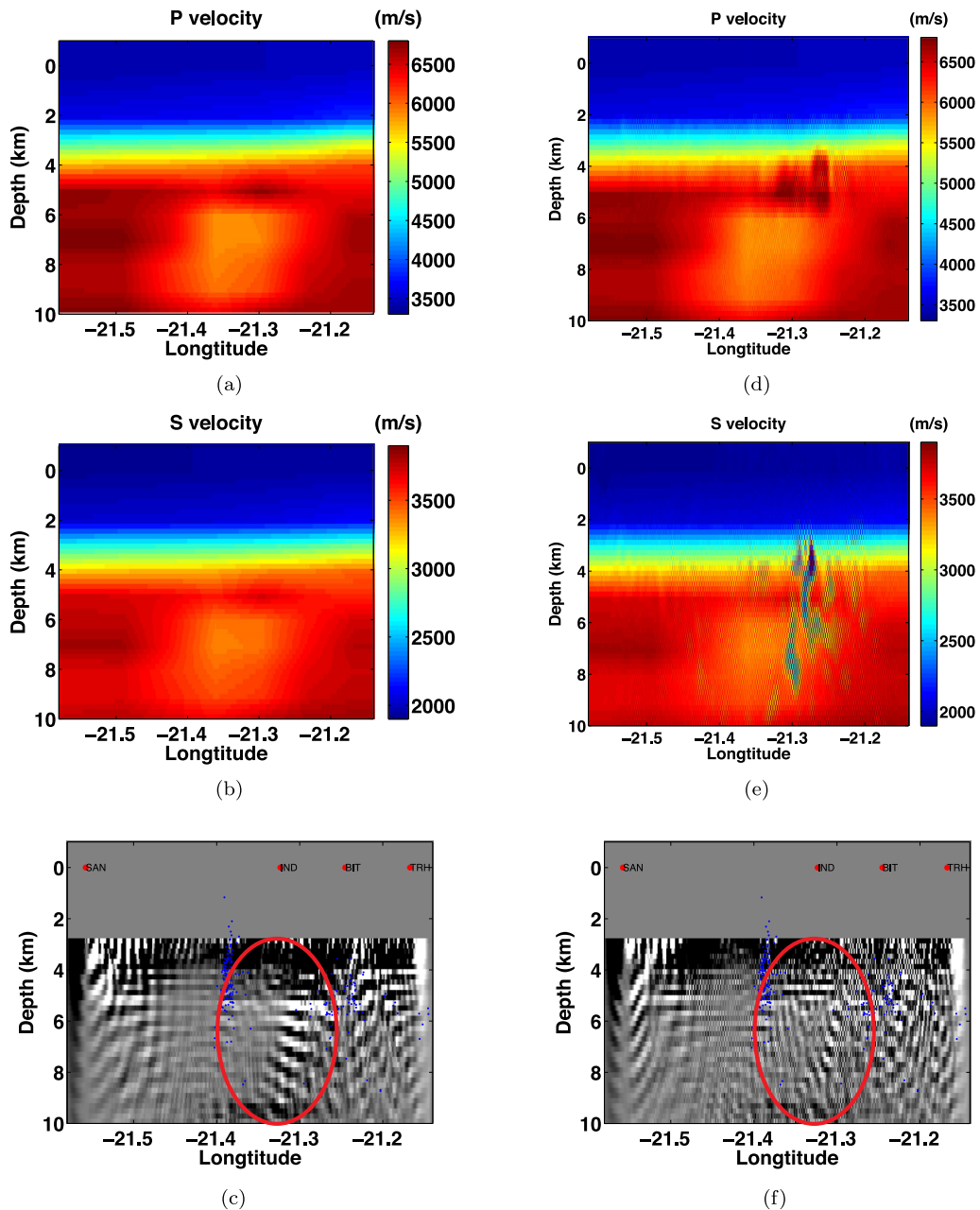
Figure 12. Left-hand side: the initial (a) *P*-wave speed, (b) *S*-wave speed and (c) SICP image obtained with the *P*- and *S*-wave speeds from (a) and (b). Right-hand side: the results of SICP-WEMVA after 27 iterations (d) *P*-wave speed, (e) *S*-wave speed and (f) SICP image obtained with the *P*- and *S*-wave speeds from (d) and (e). The blue dots in (c) and (f) mark the location of natural and induced microseismic events and the red ellipse highlights the region where the structure has improved. The shallow part of the images are contaminated due to sparse station distribution and thus were muted. The horizontal axis corresponds to distance of 22 km (i.e. exaggerated by 2.2 times in comparison to the vertical depth axis).

First, SICP-WEMVA is derived based on the assumption of an isotropic medium in which the total acceleration wavefield is decomposed into the *P*- and *S*-wavefield components. Although the wavefield separation can be achieved in certain anisotropic cases (e.g. Yan 2010; Zhang & McMechan 2010; Yan & Sava 2011) for imaging purposes, additional mathematic effort is needed to derive the sensitivity kernel to update the velocity model for the anisotropic medium. Note however, that there are studies for reflection type source-dependent anisotropic WEMVA (e.g. Li *et al.* 2012; Li 2013; Weibull & Arntsen 2013) that could potentially be applied to the source-independent converted-wave case constructed here.

The second assumption is related to the differential semblance operator (eq. 7) for the residual image, which is defined as a multiplication of the extended image by *h*. Although this operator is very intuitive and was derived in Shen & Symes (2008) from the derivative operator that measures the flatness of the seismic energy in the angle domain CIG, it has been recognized that a more sophisticated operator is required for better focusing of the energy at zero subsurface space-lag and reducing noise at large lags (e.g. Shan & Wang 2013; Weibull & Arntsen 2013).

The third consideration addresses the source illumination or number of sources required to focus the energy around zero space-lag, even with the true velocity model. It is known that for reflection
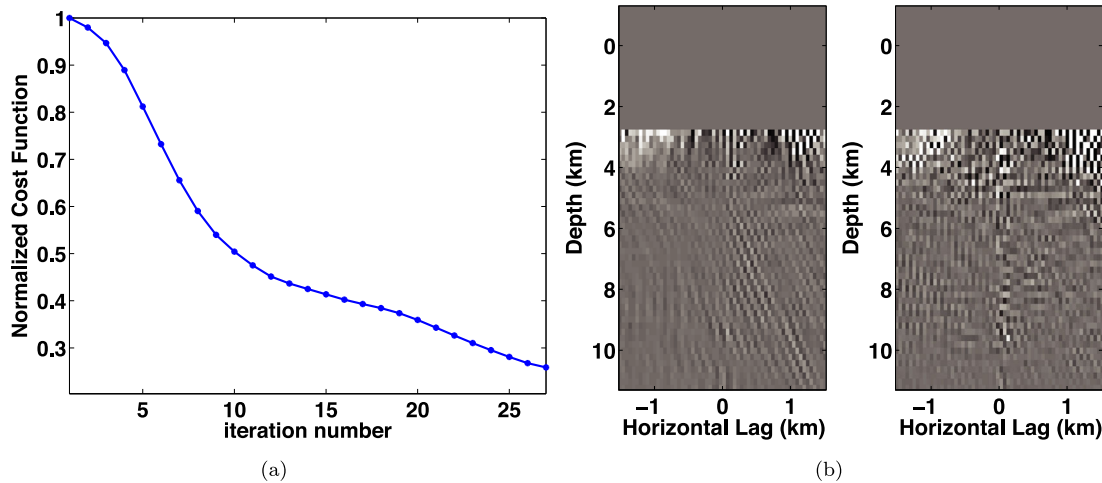
Figure 13. (a) Convergence curve for SICP-WEMVA. (b) Converted phase subsurface space-lag CIG with initial and updated velocity models shown in Fig. 12.

type WEMVA (and CP-WEMVA), many sources are required to obtain focusing. For transmission SICP-WEMVA, in contrast, the number of sources could be smaller (see Fig. 3d) because of the low frequency content in the data due to the presence of the direct waves. The comparison between WEMVA/CP-WEMVA and SICP-WEMVA will be discussed elsewhere.

The fourth and last issue is the applicability of the SICP-WEMVA to different types of seismic data: active source surface-seismic and VSP, and passive tele-seismic data. For surface-seismic data, where the CP signal in the data is generally of a reflection type for both *P* and (converted) *S* waves, reflection SICP-WEMVA may have similar limitations with source illumination and frequency content as the standard reflection WEMVA/CP-WEMVA, as discussed above. For VSP studies, SICP-WEMVA may have limitations with the illumination due to the geometry of the monitoring wells, and it depends on the wave types that are recorded: reflection, transmission or both types simultaneously. The use of both wave types simultaneously might produce images and velocity models with improved resolution from SICP-WEMVA but may also introduce additional complexities. For tele-seismic data, the separation of different body-wave phases should be clearer than with microseismic data, shown in this study, and should produce clear large-scale images and velocity models, but may be restricted by the high amplitude surface waves in the record and thus would require procedures for windowing out the surface-waves before using body-wave information in the SICP-WEMVA. However, SICP-WEMVA does not require source information, thus it does not distinguish between different types of data and acquisitions, and can be applied similarly to each data set as long as CP waves are present in the signal. Nevertheless, more research needs to be conducted to better understand the method's performance on each individual data type.

5 CONCLUSION

We have developed a fully automatic, data-driven optimization method for velocity update using CP waves. We presented a CP imaging condition in the extended space-lag domain for the objective functional. We derived a gradient-based iterative optimization method based on this objective functional and tested the method using a synthetic model and field microseismic data. The results show the applicability of the method in particular to microseismic

applications where source information is not directly available. This opens up the possibility for source-independent full-wavefield velocity analysis, a methodology that previously was not available, thus considerably improving the quality of velocity analysis and reducing the computational and processing cost.

ACKNOWLEDGEMENTS

REFERENCES

Aki, K. & Richards, P.G., 2002. *Quantitative Seismology, Theory and Methods,* 2nd edn, University Science Books.

Albertin, U., Sava, P., Etgen, J. & Maharramov, M., 2006. Adjoint wave-equation velocity analysis, in *Proceedings of the 74th Annual International Meeting, SEG,* New Orleans, Expanded Abstracts, pp. 3009–3013.

Alkhalifah, T., Fomel, S. & Biondi, B., 2001. The space–time domain: theory and modelling for anisotropic media, *Geophys. J. Int.,* **144**(1), 105–113.

Biondi, B. & Sava, P., 1999. Wave-equation migration velocity analysis, in *Proceedings of the EAGE/SEG Summer Research Workshop,* Trieste, Italy, August/September 2003, Expanded Abstracts, pp. 1723–1726.

Bostock, M., Rondenay, S. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves – 1. Theory for oblique incidence, *J. geophys. Res.: Solid Earth (1978–2012),* **106**(B12), 30 771–30 782.

Brytic, V., de Hoop, M.V. & van der Hilst, R.D., 2012. Elastic-wave inverse scattering based on reverse time migration with active and passive source reflection data, *Inverse Problems Applicat: Inside Out II,* **60,** 411–453.

Carcione, J.M., 1999. Staggered mesh for the anisotropic and viscoelastic wave equation, *Geophysics,* **64,** 1863–1866.

Du, Q., Zhu, Y. & Ba, J., 2012. Polarity reversal correction for elastic reverse time migration, *Geophysics,* **77**(2), S31–S41.

Esmersoy, C., 1990. Inversion of P and SV waves from multicomponent offset vertical seismic profiles, *Geophysics,* **55**(1), 39–50.

Fei, W. & Williamson, P., 2010. On the gradient artifacts in migration velocity analysis based on differential semblance optimization, in *SEG Technical Program Expanded Abstracts 2010,* Vol. 29, pp. 4071–4076.

Hardage, B.A., DeAngelo, M.V., Murray, P.E. & Sava, D., 2011. *Multicomponent Seismic Technology,* Society of Exploration Geophysicists.

Hestenes, M.R. & Stiefel, E., 1952. Methods of conjugate gradients for solving linear systems, *J. Res. Natl. Bureau Stand.,* **49**(6), 409–436.

Kosloff, D., Reshef, M. & Loewenthal, D., 1984. Elastic wave calculations by the fourier method, *Bull. seism. Soc. Am.,* **74**(3), 875–891.

Li, Y., 2013. Image-guided WEMVA for azimuthal anisotropy, *SEG Technical Program,* Expanded Abstracts, pp. 4709–4714.

Li, Y., Shen, P. & Perkins, C., 2012. VTI migration velocity analysis using RTM, in *SEG Technical Program,* Expanded Abstracts, pp. 1–5.

Marcinkovich, C. & Olsen, K., 2003. On the implementation of perfectly matched layers in a three-dimensional fourth-order velocity-stress finite difference scheme, *J. geophys. Res.,* **108**(B5), 18:1–16.

Purnell, G.W., 1992. Imaging beneath a high-velocity layer using converted waves, *Geophysics,* **57**(11), 1444–1452.

Rickett, J.E. & Sava, P.C., 2002. Offset and angle-domain common image-point gathers for shot-profile migration, *Geophysics,* **67**(3), 883–889.

Rodi, W. & Mackie, R.L., 2001. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics,* **66**(1), 174–187.

Rondenay, S., Bostock, M. & Shragge, J., 2001. Multiparameter two-dimensional inversion of scattered teleseismic body waves 3. application to the cascadia 1993 data set, *J. geophys. Res.: Solid Earth (1978–2012),* **106**(B12), 30 795–30 807.

Sava, P. & Biondi, B., 2004. Wave-equation migration velocity analysis. I. Theory, *Geophys. Prospect.,* **52**(6), 593–606.

Shabelansky, A., 2007. Full Wave Inversion, *Master's thesis,* Tel Aviv University.

Shabelansky, A., Malcolm, A., Fehler, M. & Bakku, S., 2012. Seismic imaging of hydraulically-stimulated fractures: a numerical study of the effect of the source mechanism, in *Proceedings of the SEG Meeting,* Expanded Abstracts, pp. 1–5.

Shabelansky, A., Malcolm, A., Fehler, M., Shang, X. & Rodi, W., 2013. Converted phase elastic migration velocity analysis, in *Proceedings of the SEG Meeting,* Expanded Abstracts, pp. 1–5.

Shan, G. & Wang, Y., 2013. RTM based wave equation migration velocity analysis, in *Proceedings of the SEG Meeting,* Expanded Abstracts, pp. 1–5.

Shang, X., de Hoop, M. & van der Hilst, R., 2012. Beyond receiver functions: passive source reverse time migration and inverse scattering of converted waves, *Geophys. Res. Lett.,* **39**, 1–7.

Shen, P., 2004. Wave equation migration velocity analysis by differential semblance optimization, *PhD thesis,* Rice University.

Shen, P., 2012. An RTM based automatic migration velocity analysis in image domain, in *Proceedings of the SEG Technical Program* Expanded Abstracts, pp. 1–5.

Shen, P. & Simes, W., 2013. Subsurface domain image warping by horizontal contraction and its application to wave-equation migration velocity analysis, in *Proceedings of the SEG Technical Program,* Expanded Abstracts, pp. 4715–4719.

Shen, P. & Symes, W.W., 2008. Automatic velocity analysis via shot profile migration, *Geophysics,* **73**(5), VE49–VE59.

Stewart, R.R., 1991. Rapid map and inversion of P-SV waves, *Geophysics,* **56**(6), 859–862.

Stewart, R.R., Gaiser, J.E., Brown, R.J. & Lawton, D.C., 2003. Converted-wave seismic exploration: applications, *Geophysics,* **68**(1), 40–57.

Stoer, J. & Bulirsch, R., 2002. *Introduction to Numerical Analysis,* Vol. 12, Springer.

Symes, W. & Carazzone, J.J., 1991. Velocity inversion by differential semblance optimization, *Geophysics,* **56**(5), 654–663.

Tryggvason, A., Rögnvaldsson, S. & Flóvenz, Ó.G., 2002. Three-dimensional imaging of the P- and S-wave velocity structure and earthquake locations beneath Southwest Iceland, *Geophys. J. Int.,* **151**(3), 848–866.

Vinnik, L., 1977. Detection of waves converted from P to SV in the mantle, *Phys. Earth planet. Inter.,* **15**(1), 39–45.

Vinnik, L., Avetisjan, R. & Mikhailova, N., 1983. Heterogeneities in the mantle transition zone from observations of P-to-SV converted waves, *Phys. Earth planet. Inter.,* **33**(3), 149–163.

Vyas, M. & Tang, Y., 2010. Gradients for wave-equation migration velocity analysis, in *Proceedings of the the 72nd EAGE Conference and Exhibition Incorporating SPE EUROPEC 2010,* Barcelona, Spain, 2010 June 14–17.

Weibull, W.W. & Arntsen, B., 2013. Anisotropic migration velocity analysis using reverse-time migration, *Geophysics,* **79**(1), R13–R25.

Xiao, X. & Leaney, W., 2010. Local vertical seismic profiling (VSP) elastic reverse-time migration and migration resolution: salt-flank imaging with transmitted P-to-S waves, *Geophysics,* **75**(2), S35–S49.

Yan, J., 2010. Wave-mode separation for elastic imaging in transversely isotropic media, *PhD thesis,* Colorado School of Mines.

Yan, J. & Sava, P., 2011. Improving the efficiency of elastic wave-mode separation for heterogeneous tilted transverse isotropic media, *Geophysics,* **76**(4), T65–T78.

Zhang, Q. & McMechan, G.A., 2010. 2D and 3D elastic wavefield vector decomposition in the wavenumber domain for VTI media, *Geophysics,* **75**(3), D13–D26.

APPENDIX A: MAXIMUM SPACE-LAG SELECTION PROCEDURE

For choosing the optimal maximum space-lags, $\underline{H}$, we need to derive and solve another optimization problem. Although we do not solve the optimization problem explicitly, here we give a recipe for the simple selection of a nearly optimal $\underline{H}$. We reduce the space-lag 3-D domain to 1-D along the x direction and take the derivative of eq. (7), the objective functional of SICP-WEMVA, with respect to maximum space-lags, H_x, using the finite difference approximation, and set it to zero as

$$\frac{\partial J}{\partial H_x} \approx \frac{1}{2\Delta H_x} \left(\int \int_{-H_{x_2}}^{H_{x_2}} h_x^2 I^2(\underline{x}, h_x)\, dh_x\, d\underline{x} - \int \int_{-H_{x_1}}^{H_{x_1}} h_x^2 I^2(\underline{x}, h_x)\, dh_x\, d\underline{x} \right) = 0, \tag{A1}$$

where $\Delta H_x = H_{x_2} - H_{x_1}$. For the sake of simplicity we omit the dependence on $\underline{x}$, and by manipulating the limits of integration, eq. (A1) becomes

$$\frac{\partial J}{\partial H_x} \approx \frac{1}{2\Delta H_x} \left(\int_{H_{x_1}}^{H_{x_2}} h_x^2 I^2(h_x)\, dh_x - \int_{-H_{x_1}}^{-H_{x_2}} h_x^2 I^2(h_x)\, dh_x \right) = 0 \tag{A2}$$

or

$$\frac{\partial J}{\partial H_x} \approx \frac{1}{2\Delta H_x} \left(\int_{H_{x_1}}^{H_{x_1}+\Delta H_x} h_x^2 I^2(h_x)\, dh_x - \int_{-H_{x_1}}^{-H_{x_1}-\Delta H_x} h_x^2 I^2(h_x)\, dh_x \right) = 0. \tag{A3}$$

Thus, to obtain the nearly optimal solution for H_x we calculate the integrals in eq. (A3) solely between $-H_{x_1}$ and $-H_{x_1} - \Delta H_x$ and between H_{x_1} and $H_{x_1} + \Delta H_x$, which in general are small. We use the regula falsi algorithm (e.g. Stoer & Bulirsch 2002, p. 339) to iteratively solve eq. (A3) to select the values of H_{x_1}.

APPENDIX B: DERIVATION OF THE SENSITIVITY KERNEL FOR SICP-WEMVA

To derive the sensitivity kernels, $K_{\hat{\alpha}}$ and $K_{\hat{\beta}}$, we perturb the objective function (eq. 7) and obtain

$$\delta J = \int \int \delta I(\underline{x}, \underline{h}) R(\underline{x}, \underline{h}) \, d\underline{x} \, d\underline{h}, \tag{B1}$$

where $R(\underline{x}, \underline{h}) = \underline{h}^2 I(\underline{x}, \underline{h})$ and

$$\delta I(\underline{x}, \underline{h}) = \sum_j^{N_e} \int \left(\delta \ddot{\underline{u}}_p^j(\underline{x} - \underline{h}, t) \cdot \ddot{\underline{u}}_s^j(\underline{x} + \underline{h}, t) + \ddot{\underline{u}}_p^j(\underline{x} - \underline{h}, t) \cdot \delta \ddot{\underline{u}}_s^j(\underline{x} + \underline{h}, t) \right) dt. \tag{B2}$$

Substituting eq. (B2) into eq. (B1) yields

$$\delta J := \delta J_p + \delta J_s, \tag{B3}$$

where

$$\delta J_p := \sum_j^{N_e} \int \int \int \left(\delta \ddot{\underline{u}}_p^j(\underline{x} - \underline{h}, t) \cdot \ddot{\underline{u}}_s^j(\underline{x} + \underline{h}, t) \right) R(\underline{x}, \underline{h}) \, dt \, d\underline{x} \, d\underline{h} \tag{B4}$$

and

$$\delta J_s := \sum_j^{N_e} \int \int \int \left(\ddot{\underline{u}}_p^j(\underline{x} - \underline{h}, t) \cdot \delta \ddot{\underline{u}}_s^j(\underline{x} + \underline{h}, t) \right) R(\underline{x}, \underline{h}) \, dt \, d\underline{x} \, d\underline{h}. \tag{B5}$$

To find $\delta \ddot{\underline{u}}_p^j$ and $\delta \ddot{\underline{u}}_s^j$ in eqs (B4) and (B5), we perturb eq. (4) and obtain:

$$\delta \ddot{\underline{u}}_p^j = \delta \hat{\alpha} \nabla \nabla \cdot \underline{u}^j + \hat{\alpha} \nabla \nabla \cdot \delta \underline{u}^j \tag{B6}$$

and

$$\delta \ddot{\underline{u}}_s^j = -\delta \hat{\beta} \nabla \times \nabla \times \underline{u}^j - \hat{\beta} \nabla \times \nabla \times \delta \underline{u}^j, \tag{B7}$$

both of which depend on $\delta \underline{u}^j$. To evaluate $\delta \underline{u}^j$, we consider eq. (2) with any input source function $\underline{f}$ as a linear system given as

$$L\underline{u}^j = \underline{f}, \tag{B8}$$

where L is the isotropic elastic wave operator defined as

$$L = \hat{\alpha} \nabla \nabla \cdot - \hat{\beta} \nabla \times \nabla \times - \partial_t^2. \tag{B9}$$

By perturbing eq. (B8) we obtain

$$\delta L \underline{u}^j + L \delta \underline{u}^j = 0 \tag{B10}$$

or

$$\delta \underline{u}^j = -L^{-1} \delta L \underline{u}^j. \tag{B11}$$

The perturbation of the modelling operator, L, is

$$\delta L = \delta \hat{\alpha} (\nabla \nabla \cdot) - \delta \hat{\beta} (\nabla \times \nabla \times), \tag{B12}$$

which turns eq. (B11) into

$$\delta \underline{u}^j = -L^{-1} \left(\delta \hat{\alpha} \frac{\ddot{\underline{u}}_p^j}{\hat{\alpha}} + \delta \hat{\beta} \frac{\ddot{\underline{u}}_s^j}{\hat{\beta}} \right), \tag{B13}$$

where $\frac{\ddot{\underline{u}}_p^j}{\hat{\alpha}} = \nabla \nabla \cdot \underline{u}^j$ and $\frac{\ddot{\underline{u}}_s^j}{\hat{\beta}} = -\nabla \times \nabla \times \underline{u}^j$ (see eq. 4).

Substituting eq. (B13) into eqs (B6) and (B7) yields

$$\delta \ddot{\underline{u}}_p^j = \frac{\delta \hat{\alpha}}{\hat{\alpha}} \ddot{\underline{u}}_p^j - \hat{\alpha} \nabla \nabla \cdot L^{-1} \left(\frac{\delta \hat{\alpha}}{\hat{\alpha}} \ddot{\underline{u}}_p^j + \frac{\delta \hat{\beta}}{\hat{\beta}} \ddot{\underline{u}}_s^j \right) \tag{B14}$$

and

$$\delta \ddot{\underline{u}}_s^j = \delta \hat{\beta} \frac{\ddot{\underline{u}}_s^j}{\hat{\beta}} + \hat{\beta} \nabla \times \nabla \times L^{-1} \left(\delta \hat{\alpha} \frac{\ddot{\underline{u}}_p^j}{\hat{\alpha}} + \delta \hat{\beta} \frac{\ddot{\underline{u}}_s^j}{\hat{\beta}} \right). \tag{B15}$$

Since model parameters $\hat{\alpha}$ and $\hat{\beta}$ are smooth, we assume that the perturbation in P- and S-wavefields are not affected by perturbations in S- and P-wave speeds, respectively. Thus, we can drop the $\delta\hat{\beta}$ term in $\delta\ddot{\underline{u}}_p^j$ and the $\delta\hat{\alpha}$ term in $\delta\ddot{\underline{u}}_s^j$ to obtain

$$\delta\ddot{\underline{u}}_p^j = \frac{\delta\hat{\alpha}}{\hat{\alpha}}\ddot{\underline{u}}_p^j - \hat{\alpha}\nabla\nabla\cdot L^{-1}\frac{\delta\hat{\alpha}}{\hat{\alpha}}\ddot{\underline{u}}_p^j \tag{B16}$$

and

$$\delta\ddot{\underline{u}}_s^j = \frac{\delta\hat{\beta}}{\hat{\beta}}\ddot{\underline{u}}_s^j + \hat{\beta}\nabla\times\nabla\times L^{-1}\frac{\delta\hat{\beta}}{\hat{\beta}}\ddot{\underline{u}}_s^j. \tag{B17}$$

Next, we substitute eqs (B16) into (B4) and obtain

$$\delta J_p = \sum_j^{N_e}\int\int\int\left(\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\ddot{\underline{u}}_p^j(\underline{x}-\underline{h},t) - \hat{\alpha}(\underline{x})\nabla\nabla\cdot L^{-1}\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\ddot{\underline{u}}_p^j(\underline{x}-\underline{h},t)\right)\cdot\left(\ddot{\underline{u}}_s^j(\underline{x}+\underline{h},t)R(\underline{x},\underline{h})\right)\,\mathrm{d}t\,\mathrm{d}\underline{x}\,\mathrm{d}\underline{h}$$

$$= \int\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\left(\sum_j^{N_e}\int\int(\ddot{\underline{u}}_p^j(\underline{x}-\underline{h},t))\cdot(\ddot{\underline{u}}_s^j(\underline{x}+\underline{h},t)R(\underline{x},\underline{h}))\,\mathrm{d}t\,\mathrm{d}\underline{h}\right)\,\mathrm{d}\underline{x}$$

$$- \sum_j^{N_e}\int\int\int\left(\hat{\alpha}(\underline{x})\nabla\nabla\cdot L^{-1}\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\ddot{\underline{u}}_p^j(\underline{x}-\underline{h},t)\right)\cdot\left(\ddot{\underline{u}}_s^j(\underline{x}+\underline{h},t)R(\underline{x},\underline{h})\right)\,\mathrm{d}t\,\mathrm{d}\underline{x}\,\mathrm{d}\underline{h}. \tag{B18}$$

In the first integral we substitute the imaging condition, eq. (5), and in the second integral we recognize an inner product and so we can replace it with an adjoint operator [i.e. $\left(\hat{\alpha}\nabla\nabla\cdot L^{-1}\right)^* = (L^{-1})^*\nabla\nabla\cdot\hat{\alpha}$, where superscript $*$ refers to adjoint] in order to isolate $(\delta\hat{\alpha})/\hat{\alpha}$:

$$\delta J_p = \int\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\left(\int I(\underline{x},\underline{h})R(\underline{x},\underline{h})\,\mathrm{d}\underline{h}\right)\,\mathrm{d}\underline{x}$$

$$- \int\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\left(\sum_j^{N_e}\int\int\ddot{\underline{u}}_p^j(\underline{x}-\underline{h},t)\cdot\left((L^{-1})^*\nabla\nabla\cdot\hat{\alpha}(\underline{x})\ddot{\underline{u}}_s^j(\underline{x}+\underline{h},t)R(\underline{x},\underline{h})\right)\,\mathrm{d}t\,\mathrm{d}\underline{h}\right)\,\mathrm{d}\underline{x}. \tag{B19}$$

To make the second integral computationally efficient and integrate over $\underline{h}$ only once before applying the adjoint operator, we apply a shift in spatial coordinates (e.g. Shen 2004), and obtain

$$\delta J_p = \int\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\left(\int\underline{h}^2 I^2(\underline{x},\underline{h})\,\mathrm{d}\underline{h}\right)\,\mathrm{d}\underline{x}$$

$$- \int\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\left(\sum_j^{N_e}\int\ddot{\underline{u}}_p^j(\underline{x},t)\cdot\left((L^{-1})^*\nabla\nabla\cdot\left\{\int\hat{\alpha}(\underline{x}+\underline{h})\ddot{\underline{u}}_s^j(\underline{x}+2\underline{h},t)R(\underline{x}+\underline{h},\underline{h})\,\mathrm{d}\underline{h}\right\}\right)\,\mathrm{d}t\right)\,\mathrm{d}\underline{x}$$

$$= \int\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\left(\int\underline{h}^2 I^2(\underline{x},\underline{h})\,\mathrm{d}\underline{h}\right)\,\mathrm{d}\underline{x} - \int\frac{\delta\hat{\alpha}(\underline{x})}{\hat{\alpha}(\underline{x})}\left(\sum_j^{N_e}\int_T^0\ddot{\underline{u}}_p^j(\underline{x},t)\cdot\underline{v}_p^j(\underline{x},T-t)\,\mathrm{d}t\right)\,\mathrm{d}\underline{x}$$

$$:= \int\delta\hat{\alpha}(\underline{x})\left(K_{\hat{\alpha}}^D(\underline{x}) + K_{\hat{\alpha}}^A(\underline{x})\right)\,\mathrm{d}\underline{x}. \tag{B20}$$

The sensitivity kernels in the first and second integrals are denoted by $K_{\hat{\alpha}}^D$ and $K_{\hat{\alpha}}^A$, where the superscripts D and A refer to *direct* and *adjoint*, respectively, based on their computation procedure (i.e. for the *direct* term, adjoint wavefield is not calculated). Note that in our experience with the examples above $K_{\hat{\alpha}}^D$ is significantly smaller than $K_{\hat{\alpha}}^A$ and thus are dropped. We therefore use $K_{\hat{\alpha}}^A$ for the sensitivity kernel used in eq. (9). However, if $K_{\hat{\alpha}}^A$ is not small it might underestimate the gradient calculation as $K_{\hat{\alpha}}^D$ is always positive.

Similarly we substitute eqs (B17) into (B5) to derive δJ_s:

$$\delta J_s = \int\int\sum_j^{N_e}\left(\int\ddot{\underline{u}}_p^j(\underline{x}-\underline{h},t)\cdot\frac{\delta\hat{\beta}(\underline{x})}{\hat{\beta}(\underline{x})}\ddot{\underline{u}}_s^j(\underline{x}+\underline{h},t)\,\mathrm{d}t\right)R(\underline{x},\underline{h})\,\mathrm{d}\underline{x}\,\mathrm{d}\underline{h}$$

$$+ \int\sum_j^{N_e}\int\int\left(\ddot{\underline{u}}_p^j(\underline{x}-\underline{h},t)\cdot\hat{\beta}(\underline{x})\nabla\times\nabla\times L^{-1}\frac{\delta\hat{\beta}(\underline{x})}{\hat{\beta}(\underline{x})}\ddot{\underline{u}}_s^j(\underline{x}+\underline{h},t)\right)R(\underline{x},\underline{h})\,\mathrm{d}t\,\mathrm{d}\underline{x}\,\mathrm{d}\underline{h}$$

$$= \int\frac{\delta\hat{\beta}(\underline{x})}{\hat{\beta}(\underline{x})}\int\int\underline{h}^2 I^2(\underline{x},\underline{h})\,\mathrm{d}\underline{x}\,\mathrm{d}\underline{h}$$

$$+ \int\frac{\delta\hat{\beta}(\underline{x})}{\hat{\beta}(\underline{x})}\sum_j^{N_e}\int\int\left((L^{-1})^*\nabla\times\nabla\times\hat{\beta}(\underline{x})\ddot{\underline{u}}_p^j(\underline{x}-\underline{h},t)R(\underline{x},\underline{h})\right)\cdot\ddot{\underline{u}}_s^j(\underline{x}+\underline{h},t)\,\mathrm{d}t\,\mathrm{d}\underline{x}\,\mathrm{d}\underline{h}$$

$$= \int \frac{\delta \hat{\beta}(\underline{x})}{\hat{\beta}(\underline{x})} \int \int \underline{h}^2 I^2(\underline{x}, \underline{h}) \, d\underline{x} \, d\underline{h}$$

$$+ \int \frac{\delta \hat{\beta}(\underline{x})}{\hat{\beta}(\underline{x})} \sum_j^{N_e} \int \left((L^{-1})^* \nabla \times \nabla \times \left\{ \int \hat{\beta}(\underline{x} - \underline{h}) \underline{\ddot{u}}_p^j(\underline{x} - 2\underline{h}, t) R(\underline{x} - \underline{h}, \underline{h}) \, d\underline{h} \right\} \right) \cdot \underline{\ddot{u}}_s^j(\underline{x}, t) \, dt \, d\underline{x}$$

$$= \int \frac{\delta \hat{\beta}(\underline{x})}{\hat{\beta}(\underline{x})} \int \int \underline{h}^2 I^2(\underline{x}, \underline{h}) \, d\underline{x} \, d\underline{h} + \int \frac{\delta \hat{\beta}(\underline{x})}{\hat{\beta}(\underline{x})} \sum_j^{N_e} \int_T^0 \underline{v}_s^j(\underline{x}, T - t) \cdot \underline{\ddot{u}}_s^j(\underline{x}, t) \, dt \, d\underline{x}$$

$$:= \int \delta \hat{\beta}(\underline{x}) \left(K_{\hat{\beta}}^D(\underline{x}) + K_{\hat{\beta}}^A(\underline{x}) \right) \, d\underline{x}. \tag{B21}$$

As for $K_{\hat{\alpha}}$, the sensitivity kernel, $K_{\hat{\beta}}(\underline{x})$, given in eq. (10), is simply $K_{\hat{\beta}}^A(\underline{x})$ and $K_{\hat{\beta}}^D(\underline{x})$ is dropped. Note that the *direct* terms in eqs (B20) and (B21) are the same, $K_{\hat{\alpha}}^D(\underline{x}) = K_{\hat{\beta}}^D(\underline{x})$, which suggests that $K^D(\underline{x})$ should have a larger bias on the *S*-wave speed estimation rather that of the *P*-wave. The effect of the *direct* terms on the full sensitivity kernels requires further investigation.

APPENDIX C: AN EFFICIENT CALCULATION OF THE STEP LENGTH FOR SICP-WEMVA OPTIMIZATION

To find an optimal step length, ν, for the gradient-based optimization in eq. (15), we need to choose a step length that will set to zero the derivative of the objective functional with respect to the step length. In other words, if $J(\nu_{i_{k+1}}) = J(\underline{m}_i + \nu_{i_k} \underline{p}_i)$, then

$$\frac{\partial J}{\partial \nu_{i_k}} \left(\nu_{\text{optimal}} \right) = \frac{\partial J}{\partial \underline{m}_i} \cdot \frac{\partial \underline{m}_i}{\partial \nu_{i_k}} = \underline{g}_{i_k} \cdot \underline{p}_i = 0. \tag{C1}$$

One way to find the optimal ν's is to use the line search method: start with an initial guess and use the bisection method (e.g. Rodi & Mackie 2001) or the regula falsi algorithm for iterative update of ν, such that eq. (C1) is satisfied. To satisfy this equation we need to calculate gradient, $\underline{g}_{i_k}$, at each iteration of the line search along a given search direction. Hence the calculation of each gradient (in SICP-WEMVA) is accomplished with two propagations in eqs (9) and (10). We propose a faster way to calculate the ν's without calculating the gradient at each iteration; we calculate the objective functional for two consecutive ν's (from either initial guesses or bisection method) with only one propagation and then calculate the derivative of eq. (C1) using the finite difference approximation as

$$\frac{\partial J}{\partial \nu_{i_k}} = \frac{J(\nu_{i_k}) - J(\nu_{i_{k-1}})}{\nu_{i_k} - \nu_{i_{k-1}}} = 0. \tag{C2}$$

Note that for eq. (C2), we calculate the gradient solely when we define a new search direction.