Geophysical Prospecting

EAGE EUROPEAN ASSOCIATION OF GEOSCIENTISTS & ENGINEERS

Geophysical Prospecting, 2015, **63**, 1050–1069 doi: 10.1111/1365-2478.12225

Image-warping waveform tomography

Francesco Perrone*,† and Paul Sava

Center for Wave Phenomena, Colorado School of Mines, Golden, CO 80401, USA

Received December 2013, revision accepted November 2014

ABSTRACT

Imaging the change in physical parameters in the subsurface requires an estimate of the long wavelength components of the same parameters in order to reconstruct the kinematics of the waves propagating in the subsurface. One can reconstruct the model by matching the recorded data with modeled waveforms extrapolated in a trial model of the medium. Alternatively, assuming a trial model, one can obtain a set of images of the reflectors from a number of seismic experiments and match the locations of the imaged interfaces. Apparent displacements between migrated images contain information about the velocity model and can be used for velocity analysis. A number of methods are available to characterize the displacement between images; in this paper, we compare shot-domain differential semblance (image difference), penalized local correlations, and image-warping. We show that the image-warping vector field is a more reliable tool for estimating displacements between migrated images and leads to a more robust velocity analysis procedure. By using image-warping, we can redefine the differential semblance optimization problem with an objective function that is more robust against cycle-skipping than the direct image difference. We propose an approach that has straightforward implementation and reduced computational cost compared with the conventional adjoint-state method calculations. We also discuss the weakness of migration velocity analysis in the migrated-shot domain in the case of highly refractive media, when the Born modelling operator is far from being unitary and thus its adjoint (migration) operator poorly approximates the inverse.

INTRODUCTION

Determining accurate propagation velocities in the subsurface is crucial in seismic imaging and exploration geophysics. A kinematically accurate velocity model allows us to correctly reconstruct Green's functions and thus to precisely image the subsurface discontinuities responsible for the data acquired at the surface or in boreholes. An accurate image of the geologic structures is key for planning drilling operations and for reducing hazards associated with drilling through faults. Seismic propagation velocities are sensitive to the stress conditions in the subsurface and carry information about the orientation of the principal stresses and overpressures (Carcione 2007), which are both important for mitigating hazards during drilling operations and reducing risk in production.

In recent years, a main distinction has been drawn between methods that recover the velocity model by matching the recorded data with simulated synthetic wavefields modeled in a trial velocity model (Tarantola 1984; Pratt 1999; Sirgue and Pratt 2004), and methods that focus on optimizing properties of the migrated images (Faye and Jeannot 1986; Al-Yahya 1989; Chavent and Jacewitz 1995; Biondi and Sava 1999; Sava, Biondi, and Etgen 2005; Albertin *et al.* 2006). Full-waveform inversion (FWI) is the name used to indicate techniques that operate in the data domain and focus on the minimization of some functional of the data residual. Migration velocity analysis (MVA), or image-domain tomography, identifies all methods that reconstruct the velocity model by analysing the similarity between migrated images.

*E-mail: fperrone@mines.edu
†Now at: CGG, Crompton Way, Crawley, West Sussex RH10 9QN, UK.

© 2015 European Association of Geoscientists & Engineers

FWI is simple to implement but very sensitive to every inconsistency between the recorded and modeled data. It requires an accurate parameterization of the physics of wave propagation, knowledge of the source signature, and a kinematically reliable starting model in order to converge to the global minimum of the objective function (Santosa and Symes 1989; Kelly *et al.* 2010). MVA is not as sensitive to errors in the source signature, and it is more robust against errors in the initial velocity model; it is characterized by an objective function that is convex over a wide range of model perturbations and thus intrinsically more suitable for gradient-based optimization procedures. However, as it ignores the amplitude information in the data for inversion purposes, MVA cannot achieve the high-resolution results that FWI promises in the ideal scenario (Warner *et al.* 2012). The main goal of MVA is to produce a kinematically accurate velocity model that will eventually be used as a starting model by FWI to reconstruct small-scale high-resolution details. In this work, we introduce an MVA methodology based on the apparent shift between single-shot (or single experiment) migrated images.

MVA is based on the idea that the subsurface is time-invariant on the timescale of the seismic experiment. The images obtained by seismic migration from a series of experiments must agree on the position of the reflectors in the subsurface since the structures that reflect the energy injected by our sources do not change over time. This is the idea behind the semblance principle (Sattlegger 1975; Al-Yahya 1989).

Differential semblance optimization (DSO) is a popular solution for measuring the consistency of a set of images and correcting the errors in the migration velocity model (Symes 1993). Applications of DSO include both regularization for data domain inversion (Symes and Carazzone 1991; Plessix 2006) and stand-alone MVA techniques in an extended space (Symes 2008; Shen and Symes 2008).

Shot-domain migrated images are obtained by applying the adjoint single-scattering Born operator to the acquired data. Because the migration operator is an adjoint and not an inverse operator, in extremely refractive media, the individual migrated images can be contaminated by kinematic artefacts (Stolk and Symes 2004). This problem is addressed by stacking, which makes migration an approximation of the inverse operator up to a dip-dependent scaling factor. For velocity analysis, we want to preserve the dimensionality or the data and avoid stacking. Data can be mapped in the image domain as a function of a surface or subsurface offset, a reflection angle, or even a shot coordinate. These mappings allow us to maintain the dimensionality of the data and use the semblance principle to evaluate the correctness of the velocity model. However, some domains are more suitable than others for MVA; Zhang *et al.* (2010) give a quick yet clear explanation of the strong and weak points of different mappings. Angle-domain common-image gathers (ADCIGs) are currently the tool of choice to assess the quality of the migration velocity model. In 2D, Sava and Fomel (2003) proposed a numerically efficient algorithm for computing ADCIGs; however, the computation of angle gathers in 3D is still a topic of active research.

DSO in the shot-migrated domain (Symes 1993; Plessix 2006) is very easy to implement but suffers from the non-optimality of the single-shot migrated images for velocity analysis. The shot domain is likely the least suitable domain for constructing gathers because of the aforementioned migration artefacts that contaminate the single-experiment migrated images (Stolk and Symes 2004; Zhang *et al.* 2010). On the other hand, shot-domain common-image gathers are an automatic output of migration schemes, such as reverse-time migration, and do not require any additional processing. Moreover, in the particular case of reverse-time migration, single-shot migrated images have the interesting property of being the partial gradients of FWI; thus, shot-domain MVA can, in principle, be used to integrate and to regularize the FWI objective function (Symes 1993; Fleury and Perrone 2012; Biondi and Almomin 2013)

This paper deals with the estimation of the velocity model using image-warping. Image-warping refers to a family of digital signal processing techniques to either estimate the apparent shift between two input signals or to apply a user specified distortion to an input signal (Hale 2009). In this work, we use the estimate of the apparent shift between single-experiment migrated images to reconstruct a kinematically accurate velocity model. Image-warping allows us to counterbalance the noise in the shot-domain image gathers with the lateral coherence of the imaged structures. Perrone *et al.* (2014) derived an approximation for the image perturbation based on the estimated displacement field and design a linearised wave-equation MVA algorithm based on the method developed by Sava and Biondi (2004). The method presented by Perrone *et al.* (2014) relies on several approximations and linearisations: the wave-equation MVA operator is obtained from the phase-shift continuation operator. The entire inversion procedure stems from the particular form of the extrapolation operator that allows one to derive a linearized operator that links perturbation in the image space to perturbation in the model parameters. Because of the strong assumption of a phase-shift continuation wave extrapolator, this method cannot be directly applied to images obtained using different engines (e.g., reverse-time

migration). Moreover, events that are not correctly handled by the extrapolation operator (e.g., diving waves and overturning reflections) become noise for the inversion scheme. In order to generalize the wave-equation MVA method for nonlinear inversion, Perrone and Sava (2012) proposed a second image-domain wavefield tomography approach where the residual is computed from estimates of local shifts between migrated images. The shifts are obtained from penalized local correlations of images from nearby experiments, and the adjoint-state method allows us to compute the gradient of the objective function and to implement a local gradient-based nonlinear inversion. The method presented by Perrone and Sava (2012) involves the solution of the full two-way wave equation and thus integrates reverse-time migration into the inversion procedure. The method shows weak points when reflectors are located near the acquisition surface and when several reflectors fall inside the same local window, thus biasing the computation of the local correlations through the crosstalk between the closely spaced events.

We extend the linearized wave-equation MVA algorithm of Perrone *et al.* (2014) to incorporate a two-way extrapolation engine and to directly exploit the displacement field. The displacement vector field is used to estimate an image perturbation; then, through single-scattering Born modelling, the wavefield perturbation is computed. The correlation between the scattered and background wavefields represents the kernel for the model update. We compare the objective functions based on DSO, penalized local image correlations, and image-warping, and show that the displacement field obtained from image-warping leads to a robust objective function and model update. From the MVA perspective, our approach can be related to the work of Xu *et al.* (2012), which follows a two-step, migration/demigration to compute the Fréchet derivative of the objective function with respect to the smooth part of the model and to recover the long wavenumber components of the velocity field. Another work on a similar path is the differential waveform inversion presented by Chauris and Plessix (2012) again in the framework of FWI. The use of an image-warping algorithm for model parameter estimation has been recently investigated in the context of FWI and data-domain techniques; Ma and Hale (2013) used dynamic image-warping (Hale 2013) to estimate shifts between simulated and recorded data and implement a travel-time inversion algorithm using connective functions (Luo and Schuster 1991) to link shifts and velocity errors. Dynamic image-warping has also been used in the context of image-domain tomography for recovering time-lapse model perturbation from 1D apparent shifts between baseline and monitor images (Yang *et al.* 2014). Baek *et al.* (2014) has implemented a more sophisticated warping algorithm to automatically register 1D traces in the data domain. They construct fractionally warped data that are then used to implement a least-squares inversion algorithm that is robust against cycle-skipping.

THEORY

DSO in the shot-image domain minimizes the energy of the difference of neighbouring images Symes (1993):

$$\mathcal{J}_{DSO}(m) = \frac{1}{2}\sum_i \left\| R_{i+1} - R_i \right\|^2, \tag{1}$$

where $R_i(\mathbf{x})$ is the migrated image obtained from the ith shot, $m(\mathbf{x})$ represents the model parameter, and $\mathbf{x} = (x, y, z)$ is the position vector. Perrone *et al.* (2014) showed that the objective function in equation (1) can be approximated using the concept of image-warping as

$$\mathcal{J}_{WARP}(m) = \frac{1}{2}\sum_i \left\| \nabla R_i \cdot \mathbf{u}_i \right\|^2, \tag{2}$$

where $\mathbf{u}_i(\mathbf{x})$ is the displacement vector field that maps $R_i(\mathbf{x})$ into $R_{i+1}(\mathbf{x})$, and $\nabla R_i(\mathbf{x})$ is the gradient of the migrated image. Displacement vectors $\mathbf{u}(\mathbf{x})$ are obtained by an iterative search of the maximum of local correlations and warping of the input images (Hale 2009); the algorithm is the same used for measuring the apparent shift between migrated images in 4D seismic processing (Hale D., Cox B. and Hatchell P.J. 2008). Assuming that the two shots illuminate the same portion of the subsurface and that the migrated images are thus similar, the displacement vector field is such that the difference between the first image after warping and the second image is "small" or, equivalently,

$$R_i(\mathbf{x} + \mathbf{u}(\mathbf{x})) \approx R_{i+1}(\mathbf{x}), \tag{3}$$

and is obtained by computing the (possibly) fractional correlation lag that maximizes the local correlation panels:

$$\mathbf{u}(\mathbf{x}) = \arg\max_{\lambda} c(\mathbf{x}, \boldsymbol{\lambda}), \tag{4}$$

where

$$c(\mathbf{x}, \boldsymbol{\lambda}) = \int_{w(\mathbf{x})} R_i\left(\boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}\right) R_{i+1}\left(\boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) d\boldsymbol{\xi} \tag{5}$$

is the local correlation of two images $R_i(\mathbf{x})$ and $R_{i+1}(\mathbf{x})$, and $w(\mathbf{x})$ are seamlessly overlapping windows. The method developed by Hale (2009) does not impose any smoothness constraint to the estimated shifts; instead, it applies signal whitening using local prediction-error filters and a cyclic search and shift along every axis to obtain a smooth estimate of the displacement field.

Perrone and Sava (2012) used penalized local correlations between migrated images to measure the shift normal to the structural dip. Their objective function is the energy of the estimated normal shift. Penalized local correlations can be differentiated with respect to the state variables and allow one to use the adjoint-state method to compute the gradient of the objective function (Lions and Magenes 1972; Fichtner, Bunge and Igel *et al.* 2006).

A rigorous adjoint-state calculation for the objective function in equation (2) can be worked out using the idea of connective functions introduced by Luo and Schuster (1991) for travel-time tomography. In appendix, we show the calculations of the sources of the adjoint wavefield necessary to compute the gradient of the objective function in equation (2) using connective functions and give an estimate of the increase computational cost with respect to our proposed approach.

The relation between the image difference and the dot product of image gradient and displacement field can be used to set up a wavefield tomography procedure that is based on the image difference objective function but that is more robust.

The adjoint sources $g_{s_i}(\mathbf{x}, t)$ and $g_{r_i}(\mathbf{x}, t)$ for the image-difference DSO objective function depend on the second derivative of the migrated images along the shot axis of the pre-stack image cube and the source and receiver wavefields $p_{s_i}(\mathbf{x}, t)$ and $p_{r_i}(\mathbf{x}, t)$ computed in the background model (Plessix 2006):

$$g_{s_i} = \left(R_{i-1} - 2R_i + R_{i+1}\right) p_{r_i}, \tag{6}$$

$$g_{r_i} = \left(R_{i-1} - 2R_i + R_{i+1}\right) p_{s_i}. \tag{7}$$

The functions in equations (6) and (7) are the force terms for the adjoint wave equations

$$\mathcal{L}(m)\, a_{s_i} = g_{s_i}, \tag{8}$$

$$\mathcal{L}(m)\, a_{r_i} = g_{r_i}, \tag{9}$$

where $a_{s_i}(\mathbf{x}, t)$ and $a_{r_i}(\mathbf{x}, t)$ are the adjoint wavefields and $\mathcal{L}(m)$ is the d'Alembert operator. The adjoint sources generate the scattered fields due to the reflectivity perturbation $\left(R_{i-1}(\mathbf{x}) - 2R_i(\mathbf{x}) + R_{i+1}(\mathbf{x})\right)$ and can be also interpreted as the *demigrated* reflectivity perturbation. The gradient of the objective function is then the zero-lag time correlation of the background and adjoint wavefields (Fichtner *et al.* 2006; Plessix 2006).

By inspecting equations (6) and (7), we can immediately see the practical issues of implementing waveform tomography by minimizing the energy of the image difference. Cycle-skipping may arise when the velocity model is severely inaccurate because different images map the same structures at inconsistent positions in the subsurface.

We propose to replace $\left(R_{i-1}(\mathbf{x}) - 2R_i(\mathbf{x}) + R_{i+1}(\mathbf{x})\right)$ in equations (6) and (7) with an approximation derived from image-warping. The apparent displacement field removes the cycle-skipping problem by warping one image into its neighbour. Let us indicate with $\mathbf{u}_i^-(\mathbf{x})$ and $\mathbf{u}_i^+(\mathbf{x})$ the apparent displacement fields that warp $R_i(\mathbf{x})$ into $R_{i-1}(\mathbf{x})$ and $R_i(\mathbf{x})$ into $R_{i+1}(\mathbf{x})$, respectively. Notice that the reference image $R_i(\mathbf{x})$ is the same for the two displacement vector fields: this is important to obtain smooth sensitivity kernels and avoid cycle skipping. By expanding in a Taylor series the warped $R_i(\mathbf{x} + \mathbf{u}(\mathbf{x}))$ image with respect to the displacement $\mathbf{u}_i(\mathbf{x})$, we can construct approximations of the image differences by computing the dot product of the image gradient with the displacement fields (Perrone *et al.* 2014):

$$\Delta R_i^- = R_{i-1} - R_i \approx \nabla R_i \cdot \mathbf{u}_i^-, \tag{10}$$

$$\Delta R_i^+ = R_{i+1} - R_i \approx \nabla R_i \cdot \mathbf{u}_i^+. \tag{11}$$

Using the expressions in equations (10) and (11), we can approximate the second derivative along the shot axis as follows:

$$R_{i-1} - 2R_i + R_{i+1} \approx \Delta R_i^- + \Delta R_i^+, \tag{12}$$

which is robust against cycle-skipping because only one image $R_i(\mathbf{x})$ enters the expression of the image perturbation.

NUMERICAL EXAMPLES

We test image-warping wavefield tomography for different depths of the reflecting interface and compare it with the direct image difference tomography approach (Symes 1993; Plessix 2006) and penalized local image correlation wavefield tomography (Perrone and Sava 2012). We show that image-warping wavefield tomography is more robust with respect to the depth of the reflector (i.e., the ratio between the source receiver offset and the depth of the reflector) and leads to more stable gradients for a model update. We also show how the method fails to produce sensible gradients in highly refractive media. In this case, the displacement estimation is stable, but the artefacts in the single-experiment migrated image lead to spurious scattered wavefields and artefacts in the sensitivity kernel. Finally, we test our method on the Marmousi model.

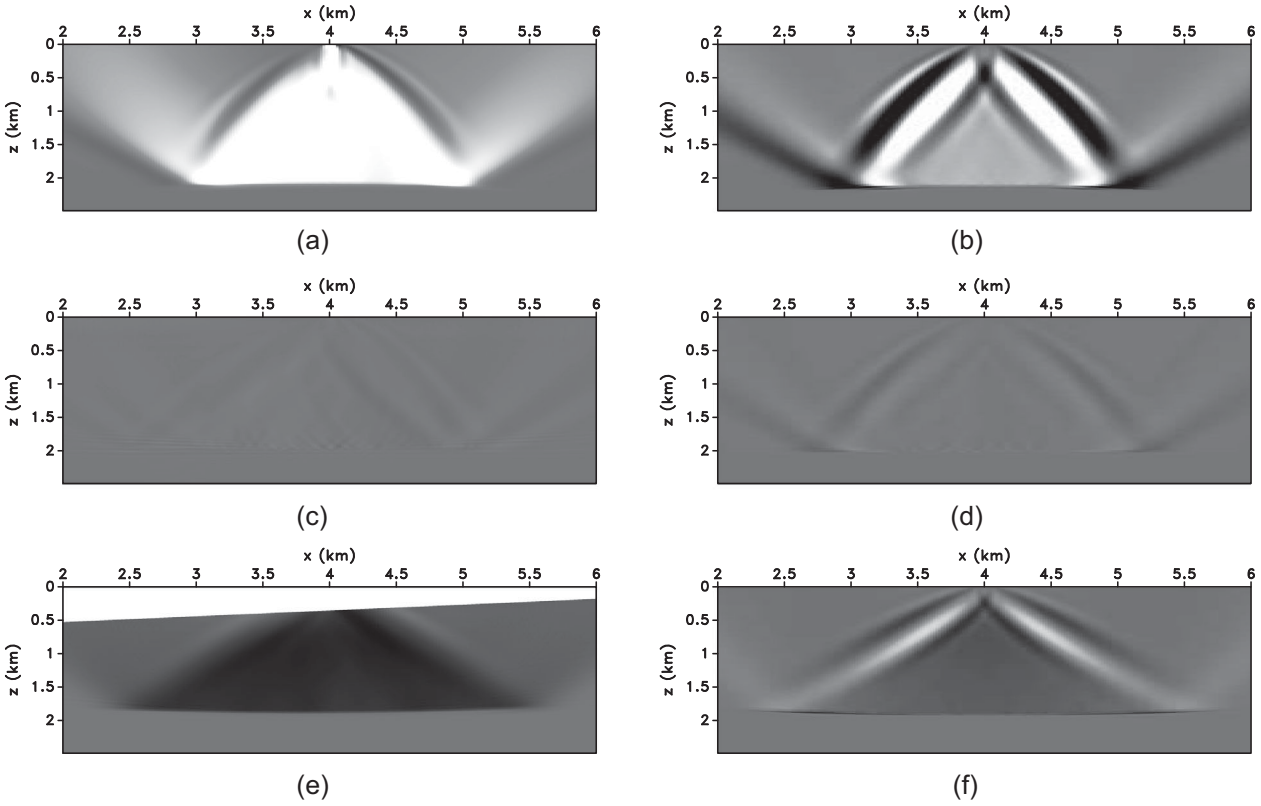


Figure 1 Sensitivity kernels obtained from a *deep interface* using *image-warping* wavefield tomography for (a) high, (c) correct, and (e) low velocities, and using *differential semblance* wavefield tomography for (b) high, (d) correct, and (f) low velocities.

Horizontal reflector

Let us consider a flat horizontal reflector in a homogeneous 2-km/s-velocity model. In the first case, the interface is at 2 km depth, and it is due to a density contrast. The data are generated using time-domain finite differences to solve the second-order partial differential equation for pressure. The source function is a Ricker wavelet with peak frequency of 15 Hz. In order to measure errors in the velocity model, we analyse two nearby shots: the shot position of the first shot is $x = 3.8$ km, and the distance between the shots is 200 m. We apply different homogeneous perturbations to the model, and we compute sensitivity kernels using using image-warping wavefield tomography and differential semblance. Figure 1 shows the comparison of the sensitivity kernels for three migration models. The rows of Fig. 1 are, from top to bottom, kernels for a high, the correct, and a low migration velocity; the columns, from left to right, show the results using image warping and standard differential semblance based on image difference. For the particular parameters in this example, the bandwidth of the data extends up to about 30 Hz, and the wavelength at 25 Hz for a velocity of 2 km/s is 80 m. The vector shift is due to both the distance between the shots and the velocity error and is not a simple vertical shift. At the large offsets, the displacement is over 50 m, which is already larger than half a wavelength. Although not all the frequencies are cycle-skipped, part of the frequency components of the signal are, and this introduces the strong sidelobes in the sensitivity kernels for standard shot-domain DSO based on image differences. Image-warping resolves the cycle-skipping between the migrated images that causes the strong sidelobes in differential semblance. Using image-warping to modify the expression of the adjoint source of differential semblance, we obtain very smooth sensitivity kernels that correctly represent the perturbation in the velocity model. Figure 2 compares the sensitivity analysis performed using the image-warping approximation of the image difference (the objective function in equation (2)) and standard differential semblance. Notice that, in Fig. 2a, we observe a more symmetrical behaviour with respect to the perturbation compared with Fig. 2b. However, the effect on the objective function behaviour is minimal, and the major difference is the quality of the model update; image-warping

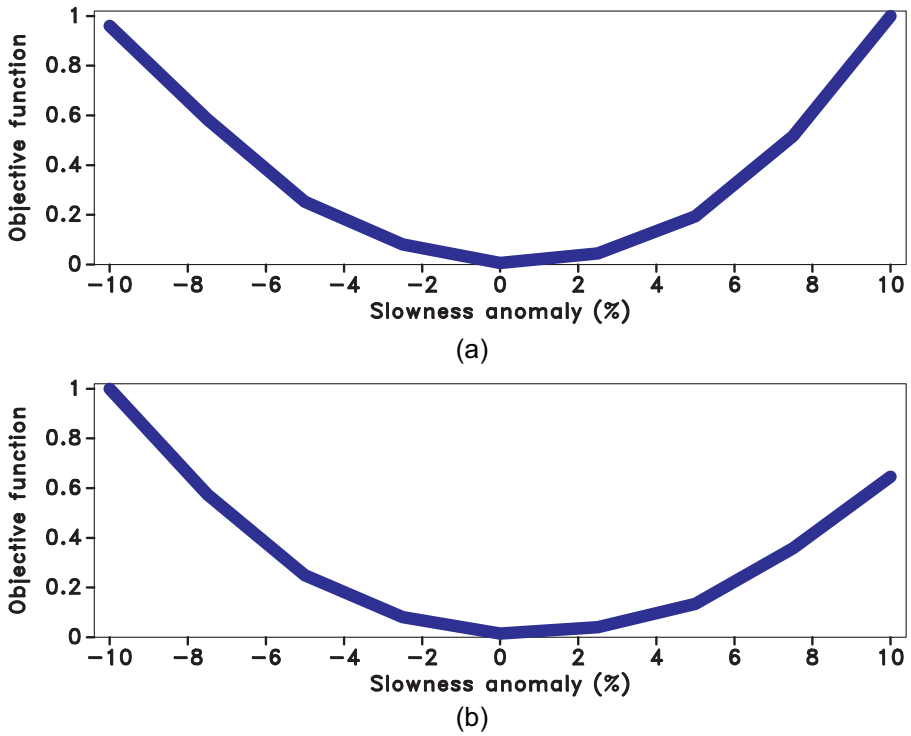


Figure 2 Objective functions for (a) image-warping wavefield tomography (equation (2)) and (b) *differential semblance* (equation (1)) evaluated for a set of model perturbations and the *deep interface*. The minimum of the objective function indicates the correct model. Differential semblance shows a slightly less resolute objective function for positive slowness anomalies.

gives a more regular update that can be related only to the robustness against cycle-skipping at longer offsets, i.e., where the differences between the images are more pronounced.

We also consider a shallower interface and compare the gradients obtained using image-warping with the ones obtained using penalized local correlations. We consider a model where the density interface is at 1 km depth. The model perturbations applied to the velocity field are the same used as in the previous case. The position of the first shot is at $x =$ 3.8 km, and the shot separation is 200 m. The comparison of the sensitivity kernels for image-warping and penalized local correlation wavefield tomography are shown in Fig. 3. In the left column, image-warping shows the ability to return kernels with the correct sign even when the interface is close to the surface, and cycle-skipping problems are enhanced because the difference between the apertures of the two shots increases. Penalized local correlations suffer from the strong bias at the edges of the aperture, where the two images do not overlap and the values of the correlation coefficients drop. The comparison of the objective functions for different values of the model perturbation in Fig. 4 highlights a strong bias for the penalized local correlations when the reflecting interface is shallow and the ratio between the depth of the interface and the shot separation is small. Nonetheless, the gradients obtained using image-warping for a deeper interface are comparable with the results reported by Perrone and Sava (2012) for the objective function employing local image correlations. Thus, image-warping and penalized local correlations are expected to perform similarly when the ratio between the depth of the reflector and the shot-point separation is large. When compared with Fig. 2a, image-warping appears to be insensitive to the depth of the reflector.

Highly refractive model and migration failure

When the model is highly refractive, kinematic artefacts contaminate the single-experiment migrated images, although the exact model is used to reconstruct the source and receiver wavefields (Stolk and Symes 2004). This fact is due to the very nature of the migration operator, which is the adjoint of the Born modelling operator and not its inverse.

Figure 5 shows the velocity model and density model we used to study the behaviour of image warping wavefield tomography in highly refractive media. The model contains

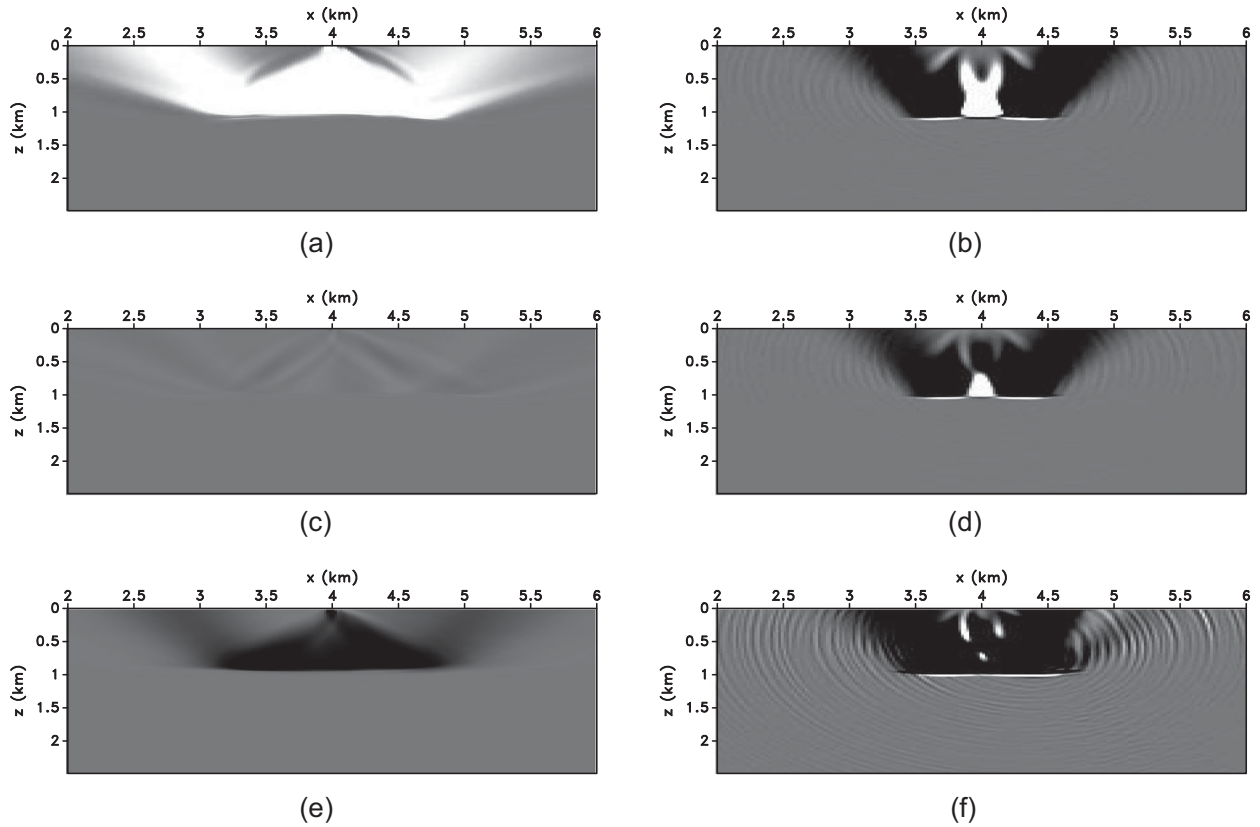


Figure 3 Sensitivity kernels obtained from a *shallow interface* using *image warping* wavefield tomography for (a) high, (c) correct, and (e) low velocities, and using penalized *local correlations* wavefield tomography for (b) high, (d) correct, and (f) low velocities. Penalized local correlation are strongly biased at the edges of the subsurface aperture.

a strong low-velocity Gaussian anomaly, whose minimum value is 40% lower than the background 2 km/s velocity. The reflector is simulated by a density contrast. The data are generated using a 15-Hz Ricker wavelet and a staggered-grid finite-difference algorithm with absorbing boundary conditions on every side. The horizontal sampling is 20 m, and the vertical sampling is 10 m.

The low-velocity area in the model creates triplications of the wavefields. The shot gather in Fig. 6 shows the complexity of the wavefield recorded at the surface. The source is located at $x = 3$ km on the surface, and receivers are at every grid point at the surface. Figure 7a shows the result of migrating the shot gather in Fig. 6 with the *exact* model. Notice that, at about $x = 5$ km, the image is distorted by the kinematic artefacts described in Stolk and Symes (2004). These artefacts do not satisfy the relationship between the dip field and displacement vector assumed in our inversion procedure. Moreover, since the migration operator is simply the adjoint, and not the inverse, of the Born modelling operator, the amplitudes of the image are corrupted as well. All these distortions affect the calculation of the sensitivity kernel. In Fig. 7b, we show the sensitivity kernel computed for the exact velocity model. Instead of having negligible amplitude everywhere, the sensitivity kernel has a strong and totally spurious contribution for the area of the image that is distorted by the migration artefacts. Notice that, since the velocity model is in fact correct, the part of the reflector that is imaged by waves that do not sense the low-velocity anomaly does not contribute to the total sensitivity kernel. The spurious kernel is entirely due to the triplication of the wavefront in the low-velocity anomaly and the associated artefacts in the single-experiment image.

Marmousi model

We also test our approach on the Marmousi model (Fig. 8a). We model 100 shots separated by 80 m. The spatial sampling is 8 m in both the x and z directions. Receivers are located at every grid point on the surface. The pressure data are modelled with a staggered-grid acoustic finite-difference

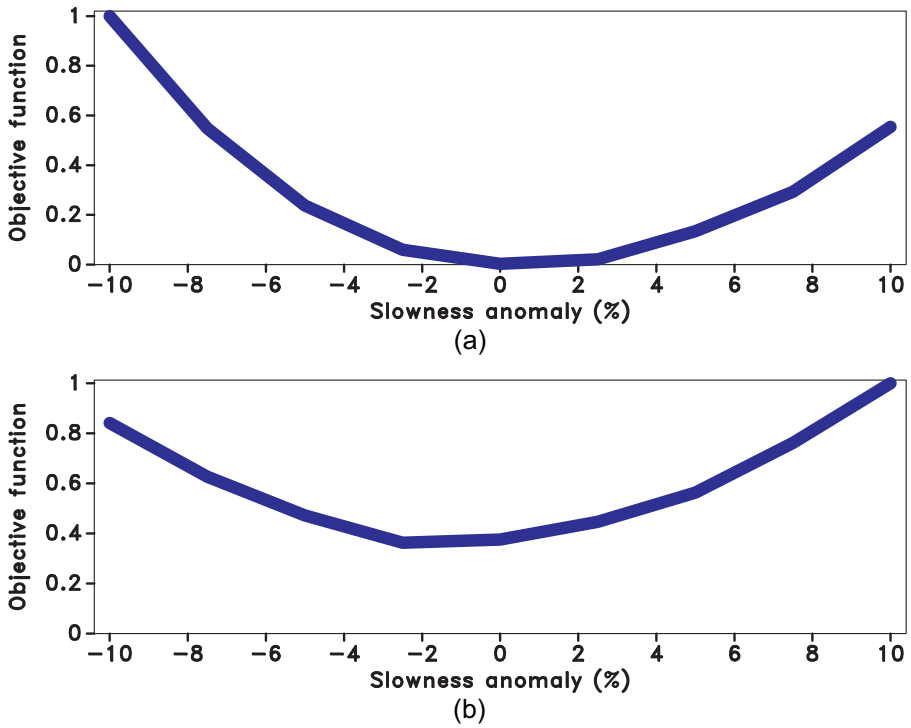


Figure 4 Objective functions for (a) *image-warping* wavefield tomography and (b) *penalized local correlations* evaluated for a set of model perturbations and the *shallow interface*. Penalized local correlation fail to identify the correct model.

code that solves the system of first-order partial differential equations for pressure and particle velocities. The source function is a Ricker wavelet with peak frequency of 15 Hz. Absorbing boundary conditions are used to remove free-surface multiples, but internal multiples are present in the data. We assume that free-surface multiples can be effectively removed from the data using, for example, SRME (Verschuur 1991). The Marmousi model represents a complex land-like scenario and is characterized by multi-pathing and strong refractions. This complexity is exemplified by the raypaths calculated in a smoothed version of the correct model (Fig. 8b) for a shot at $x = 4.8$ km.

Figure 9 shows the preprocessing applied to the shot gathers before imaging and inversion. In Fig. 9a, we display the shot gather for a source located at $x = 1.6$ km after removal of the direct arrival. In order to balance the amplitude of the reflection events and of the reflectors in the image, we equalized each trace by scaling them by a smooth local estimate of the rms value along the trace itself. Figure 9b shows the result of applying the amplitude equalization to the shot gather in Fig. 9a and then smoothly muting the refracted energy. Notice that the later arrival are much more visible after the amplitude equalization.

The initial model is the vertical gradient of velocity in Fig. 10a. The velocity increases 0.5 km/s with depth and is generally slower than in the correct model. The migrated image obtained using this model is shown in Fig. 10b. The central part of the image is non-interpretable, and all reflectors are mispositioned and defocused; the deep and strong reflectors are shifted about 500 m upward. The noise in the image is due to the strong errors in the migration model and the inconsistent imaging of the reflectors from different shots. Also, the internal multiples in the data are not imaged correctly since migration is based on single-scattering (Born) assumptions. The complex wave propagation, the strong refractions, and the the internal multiples contribute noise and artefacts in the migrated image. For example, at $x = 1$ km and $z = 2.7$ km, ghost reflectors due to strong internal multiples can be observed, and in the central part of the model, at $x = 4.5$ and $z = 2$ km, the unfocused events do not combine constructively and lead to random noise rather than an image of the geologic structures. In the latter case, the complexity of the faulted portion of the model creates scattering that is not correctly remapped using the initial model in Fig. 10a. These artefacts are the main source of noise for our wavefield tomography approach since they do not carry information

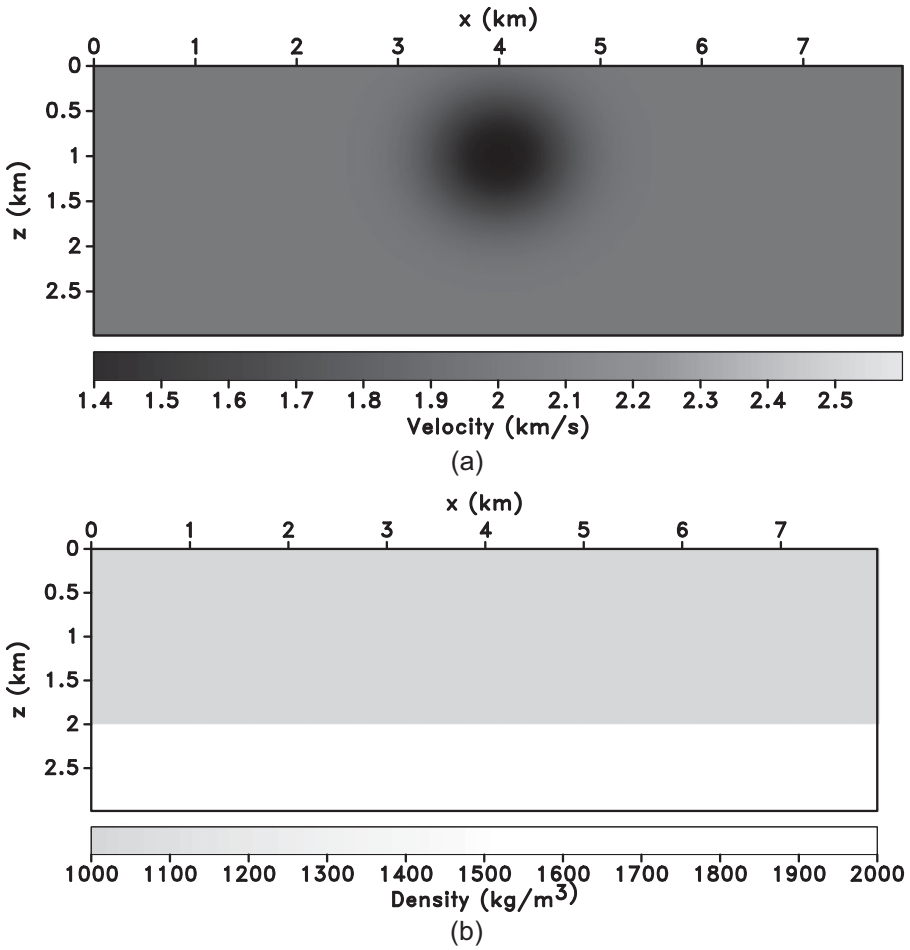


Figure 5 (a) Velocity model used to study the behaviour of the inversion in highly refracting media. The low-velocity anomaly is Gaussian shaped, and its minimum value is 40% lower than the 2 km/s background. (b) Density model used to generate reflections.

about specular reflections and since they decrease the reliability of the estimated displacement vector field, and thus of the adjoint sources. By analysing the partial images, we observe that the unfocused reverberations contribute steeply dipping and short wavelength artefacts, which can be attenuated by low passing the image in the horizontal direction. More sophisticated filtering techniques, e.g., in the curvelet domain, may help in preserving the amplitudes of the signal and thus further increasing the quality of the computed gradient.

For comparison, Figure 11 shows the gradient of the objective function computed using DSO and image-warping. The gradient of the objective function is computed with respect to slowness squared. Blue indicates that the velocity in the model should be increased, and red indicates that the velocity should be decreased. In Fig. 11a, the gradient computed from standard image differences (i.e., DSO) shows random values and does not represent a reliable direction for updating the model. In this case, the error in the migration model leads to severe cycle-skipping between adjacent images, especially at mid and large offsets. Image-warping correctly captures the shifts between the migrated images and, as a consequence, removes the cycle-skipping problem; the gradient in Fig. 11b shows a higher degree of regularity and supplies a more reliable update for the model. Notice that the regularity of the gradient in Fig. 11b decreases in the central part of the model. The faulted structure in the mid part of the model (at about $x = 4.5$ km) contains strong contrasts in propagation velocity and lateral discontinuities that produce strong reverberations and complex wavepaths. As a consequence, the single-shot migrated images are contaminated with strong artefacts compared with images of the more regular sedimentary portions of the model on the sides. Moreover, the gradient seems relatively insensitive to the different values of the velocity anomaly and does

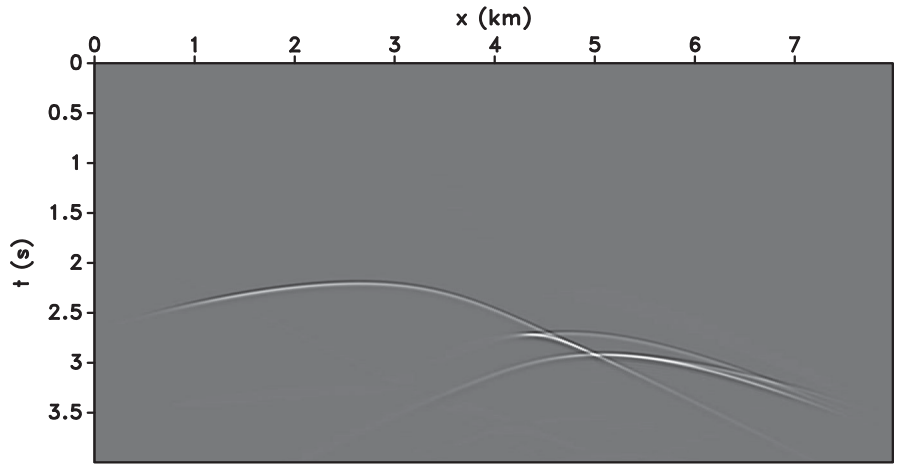


Figure 6 Shot gather for a source at $x = 3$ km. Notice the complicated response of the medium due to the triplications of the wavefield passing through the low-velocity anomaly.

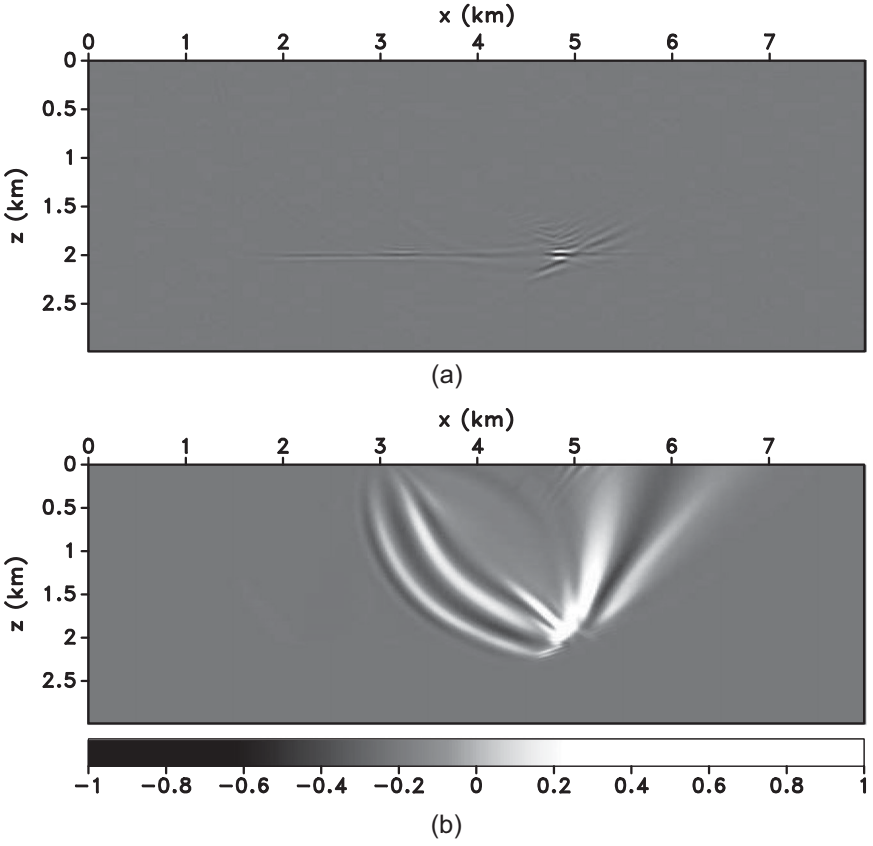


Figure 7 Migrated image obtained from the shot gather in Fig. 6 with the correct velocity model. Observe the distorted image of the reflectors and the artefacts introduced by the migration operator. (b) Sensitivity kernel obtained using image-warping wavefield tomography for the correct velocity model. The kernel should be very small and incoherent because the model is actually correct; the spurious kernel is thus completely due to the artefacts in the image.

not highlight the high-velocity layer and blocks on either the sides and the faulted area of the model.

We implement a simple steepest descent scheme to reconstruct the velocity model. The gradient of the objective function is smoothed along the structures in the image and then equalized using the source-side illumination map. In this inversion exercise, we focus on recovering the trend of the model. We further smooth the computed gradient using a triangular smoother 400 samples lateral extent and 80 samples vertical extent. Effectively, we aim to reconstruct the very long

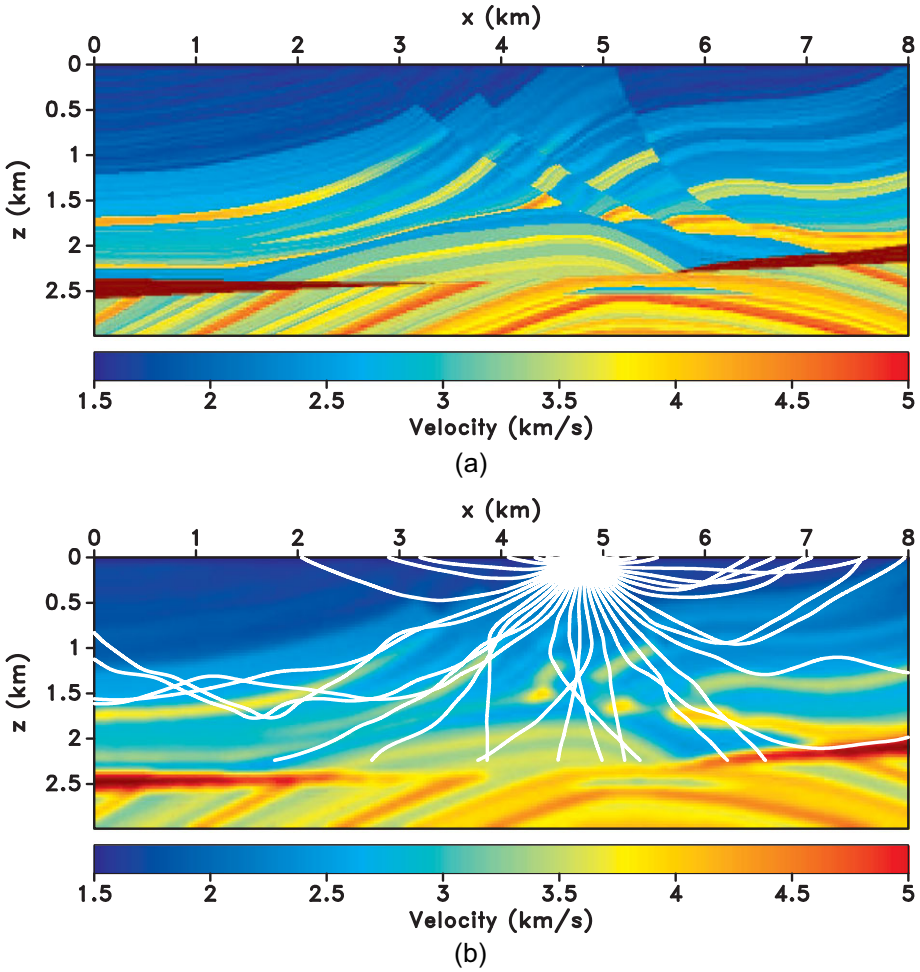


Figure 8 (a) Exact Marmousi velocity model. (b) Raytracing in a smooth version of the exact model for a shot at $x = 4.8$ km. The complex raypaths highlight the refractive behaviour of the Marmousi model.

wavelength components of the velocity field. The model update is obtained by scaling the gradient with a scalar step-length. The step-length is computed by parabolic fitting of three values of the objective function in equation (2). Figure 12 shows the final model recovered by image-domain wavefield tomography and the associated migrated image. The inversion reconstructs the very smooth background velocity model in Fig. 12a, which is characterized by a steeper velocity gradient than the initial model in Fig. 10a. By comparing the exact model in Fig. 8a and the final image in Fig. 12b, we can see that image-warping wavefield tomography is able to better position reflectors where the geology is relatively simple and the geometrical features of the image can be measured more precisely. In the shallower part of the model, the truncation of the sediments against the faults is now better interpretable (e.g., $x = 3.5$ km and $z = 1.0$ km and along the fault that reaches the surface at about $x = 5.0$ km). However, the quality of the image quickly decreases in the central, faulted portion of the model because of the high level of coherent noise in the partial images.

Figure 13 shows that the objective function decreases smoothly and monotonically, but the convergence slows down after about 50 iterations. The non-zero value of the objective function is mainly due to migration artefacts, whose contribution becomes significant with respect to the shifts measured on actual reflectors.

The initial model for image-domain tomography does not have the correct velocity gradient and leads to evident cycle-skipping in the data-domain inversion, as shown in Figure 14a. We used the model reconstructed by image-warping wavefield tomography as the initial guess for data-domain FWI. Figure 14b shows the final reconstructed model

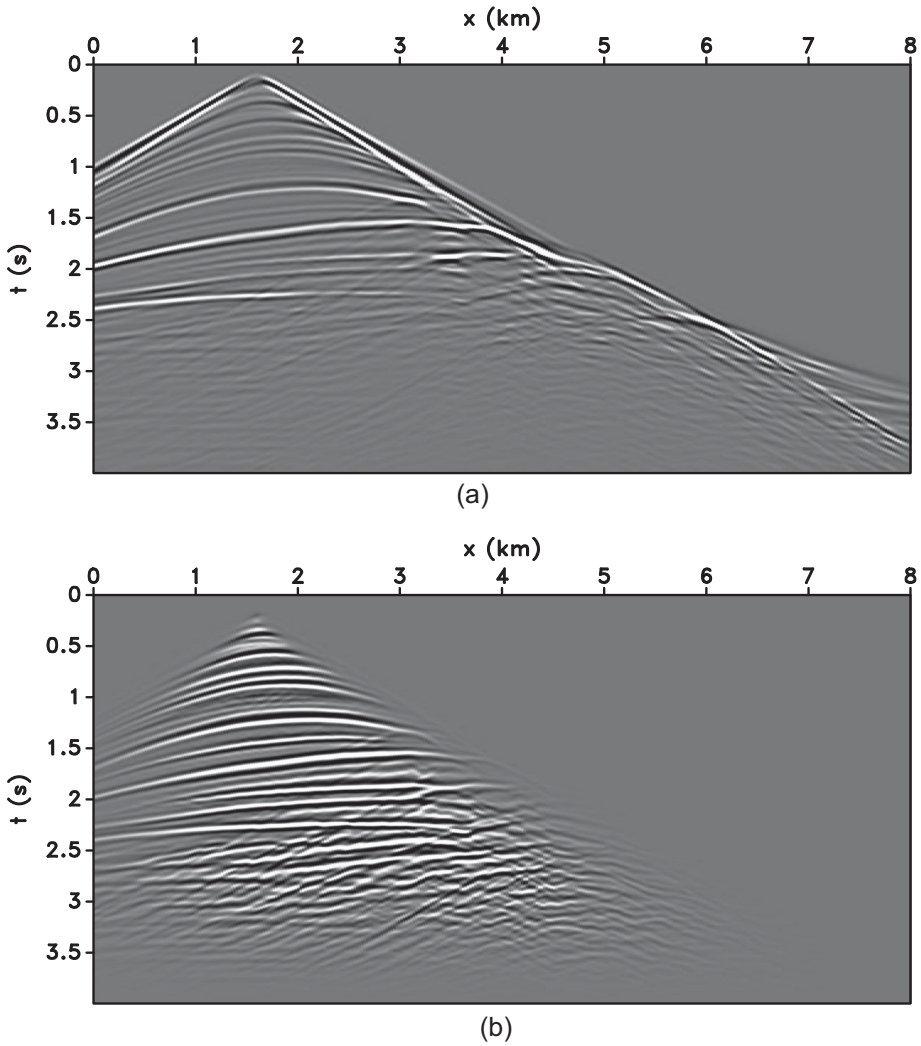


Figure 9 Shot gather with source at $x = 1.6$ km. (a) The reflection data after the removal of the direct arrival and (b) after applying amplitude equalisation and smooth tapering of the refracted energy. The amplitude equalization is implemented by scaling each trace by a smooth local estimate of the rms of the trace itself.

after 60 iterations of full-waveform inversion. The data used for the data-domain inversion test are the low-passed filtered version of the data used for image-domain tomography, the high cutoff frequency is 3 Hz, and the bandwidth of the low-passed data extends to about 5 Hz.

FWI converges to a high-resolution result and reconstructs high-wavenumber features of the model. Compared with the actual Marmousi model in Fig. 8a, we were able to accurately reconstruct the geology in the central part of the model and up to a depth of about 2.5 km. The quality of the result comes from the correct kinematics of the wavefields in the sedimentary part of the model, which was reconstructed by image-warping wavefield tomography, and allows one to correctly position the reflectors. Image-warping tomography recovers the very low wavenumber components of the velocity model, which are crucial to avoid cycle-skipping in data-domain inversion. The deeper part of the model, below 2.5 km, is poorly constrained by the data. Image-warping wavefield tomography is not able to correct the velocity gradient, and consequently, the quality of the data-domain inversion is lower, as well.

DISCUSSION

The image difference objective function proposed by Symes (1993) and Plessix (2006) as a regularization term for FWI

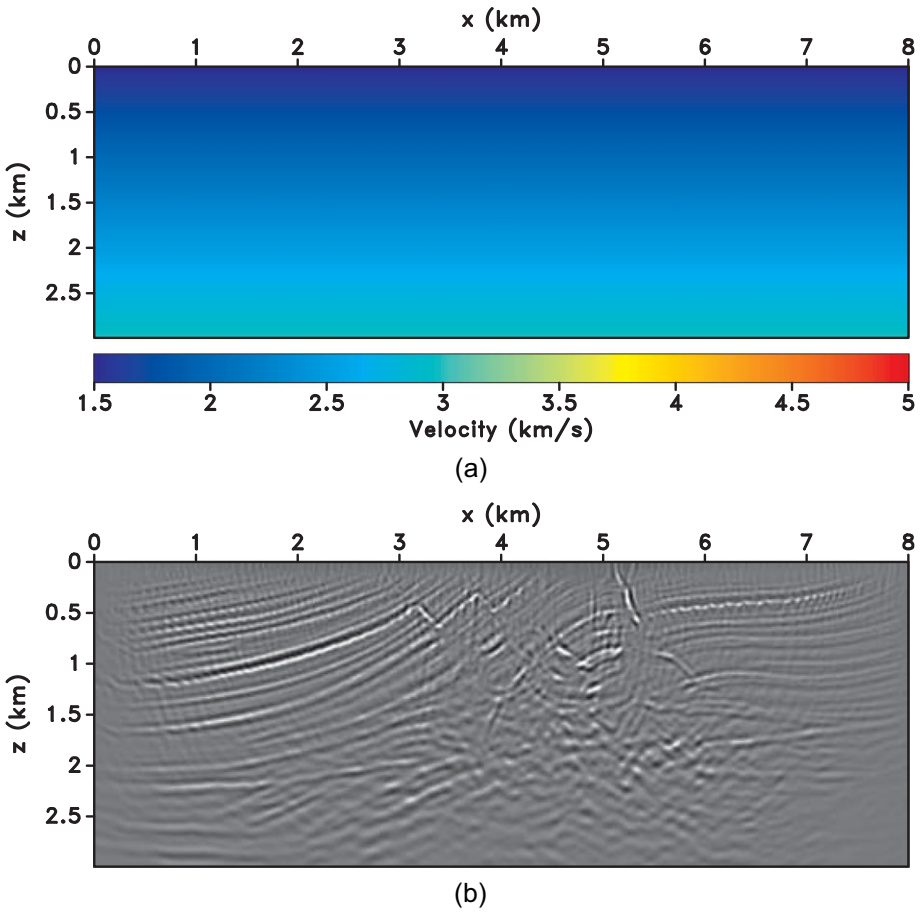


Figure 10 (a) The initial model represented by an incorrect $v(z)$ velocity. (b) The migrated image is distorted and not interpretable in the central part of the model. Observe the general upward shift of the image due to the bias in the velocity model. Ghost reflectors due to the strong internal multiples from the high-velocity inclusions can be observed at $x = 1$ km and $z = 2.5$ km.

is simple to implement and carries straightforward physical intuition about the action of the inversion. In media that are not strongly refractive, cycle-skipping can be effectively and efficiently avoided using the apparent displacement field between migrated images and constructing robust approximations of image differences. Estimating the displacement field and computing the image gradient have a negligible computational cost compared with migration. Moreover, we preserve the physical interpretation of the image perturbation that enters the expression of the adjoint sources in equations (6) and (7), which is useful for quality control and for quickly assessing the evolution of the inversion result. The residual evaluation is immediate, and the calculation of the adjoint sources is straightforward compared with the local correlation waveform tomography, where the computation of the adjoint sources requires nonstationary convolutions between the migrated images and the penalty operators followed by local scaling by the background wavefields. When compared with methods that require image gathers or extended images, image-warping wavefield tomography appears easier to implement and less computationally demanding. On the other hand, by stacking contributions from different experiments, methods based on image gathers are intrinsically less sensitive to random noise and can also attenuate the migration operator noise.

Cycle-skipping between migrated images depends on several parameters: not only the error in the velocity model but also the distance between the shots affect the similarity between the images. Intuitively, very distant shots illuminate different areas of the model, which may be quite different from the point of view of the geometry and orientation of the reflectors; in this case, the computation of either image differences or displacement vectors is meaningless because the two images do not have to be similar in the first place. If

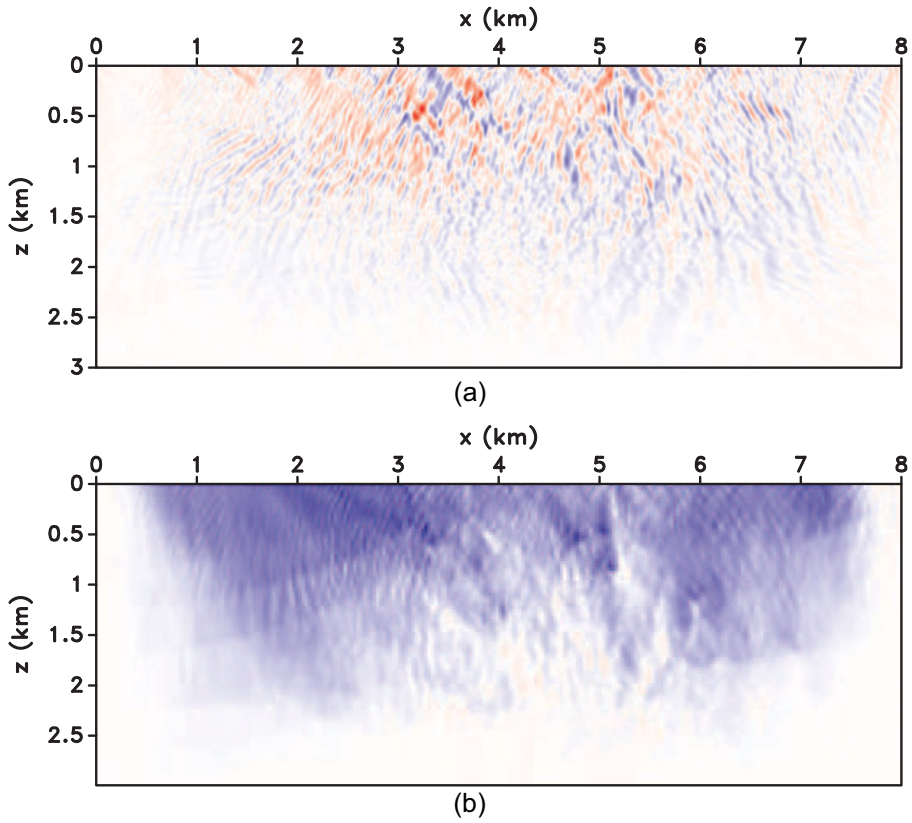


Figure 11 (a) The gradient of the standard DSO objective function and (b) the gradient obtained by modifying the adjoint sources using image-warping. In both cases, the gradient is with respect to slowness squared; blue indicates that slowness shall be decreased, and red indicates that slowness shall be increased. For this particular initial model, DSO does not provide a useful gradient for wavefield tomography.

the error in the model tends to zero, the image-warping objective function better approximates the image difference because the Taylor series expansion becomes more accurate. For larger velocity errors, cycle-skipping between seismic images may arise only in portions of the image space where the travel path of the waves is the longest. In this case, reducing the distance between the shots helps the image-difference objective function but does not remove cycle-skipping, which arises where the waves have accumulated a path difference larger than half a wavelength. Image-warping defines a one-to-one correspondence between the points in the two images, and it is thus intrinsically more robust against cycle-skipping than image-difference DSO.

The adjoint-state calculations for the objective function in equation (2) are computationally expensive and involved (see Appendix). Connective functions are necessary to establish a differentiable relationship between the warping vectors and the velocity model. From an inversion perspective, this step increases the dimensionality of the problem; Luo and Schuster (1991) operate on scalar shifts, whereas in the image domain, one manipulates vectors defined at every point in the discretised space. A further difference between the image and data domains is the presence (or lack thereof) of artefacts. The data domain is obviously free from migration artefacts but is more prone to cycle-skipping and requires long offset to constrain the long wavelength components of the model; on the contrary, the image domain can operate with relatively short offset data and is more robust against cycle-skipping when the error in the model is large.

Computing the relative shift between two images in local seamlessly overlapping windows requires some degree of overlapping of the two migrated images. The quantity of interest to qualitatively characterise the distance between the two images in simple subsurface geometries is the ratio between the shot-point separation and the depth of the reflectors. The smaller this ratio, the greater the overlap between migrated images. When we estimate the shift between the migrated images using penalized local correlations, the extent of the local window constrains the domain in which the maximum

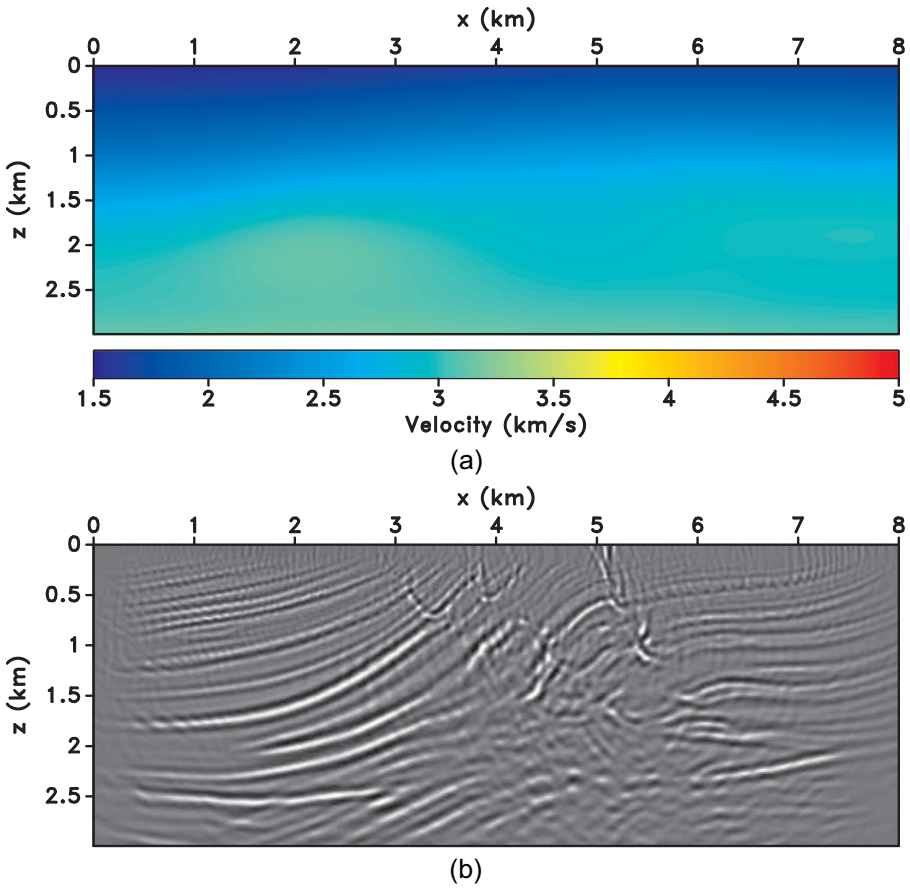


Figure 12 (a) Recovered velocity model and (b) migrated image using the result of the inversion. To reduce the influence of the migration artefacts, we focus on the long wavelength of the velocity field by smoothing the computed gradients and correct the bulk shift of the main reflectors.

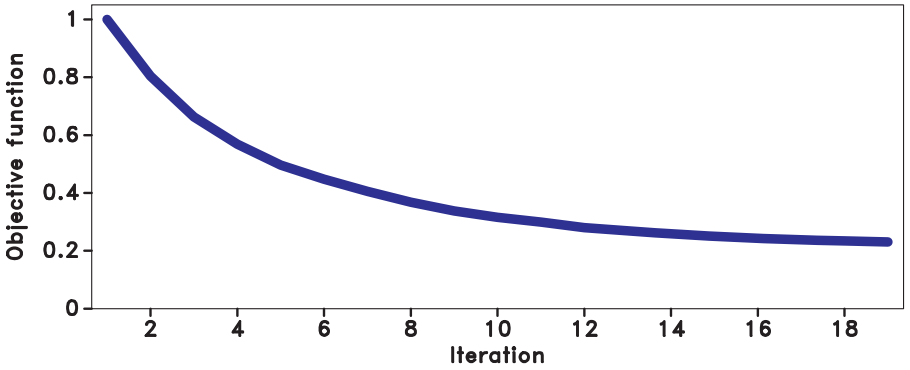


Figure 13 The value of the objective function decreases smoothly with iterations. Convergence slows down after 20 iterations because the residuals measured on migration artefacts becomes significant with respect to the residual measured on actual reflectors.

of the correlation must be. When the separation between the shot is large and the local window is not wide enough to capture the position of the maximum of the local correlation, we introduce an error in the evaluation of the residuals. Notice that increasing the size of the local correlation window implies a decrease in resolution. Moreover, if other events (more reflectors) enter the local window and/or the reflector changes orientation in space, increasing the extent of the local

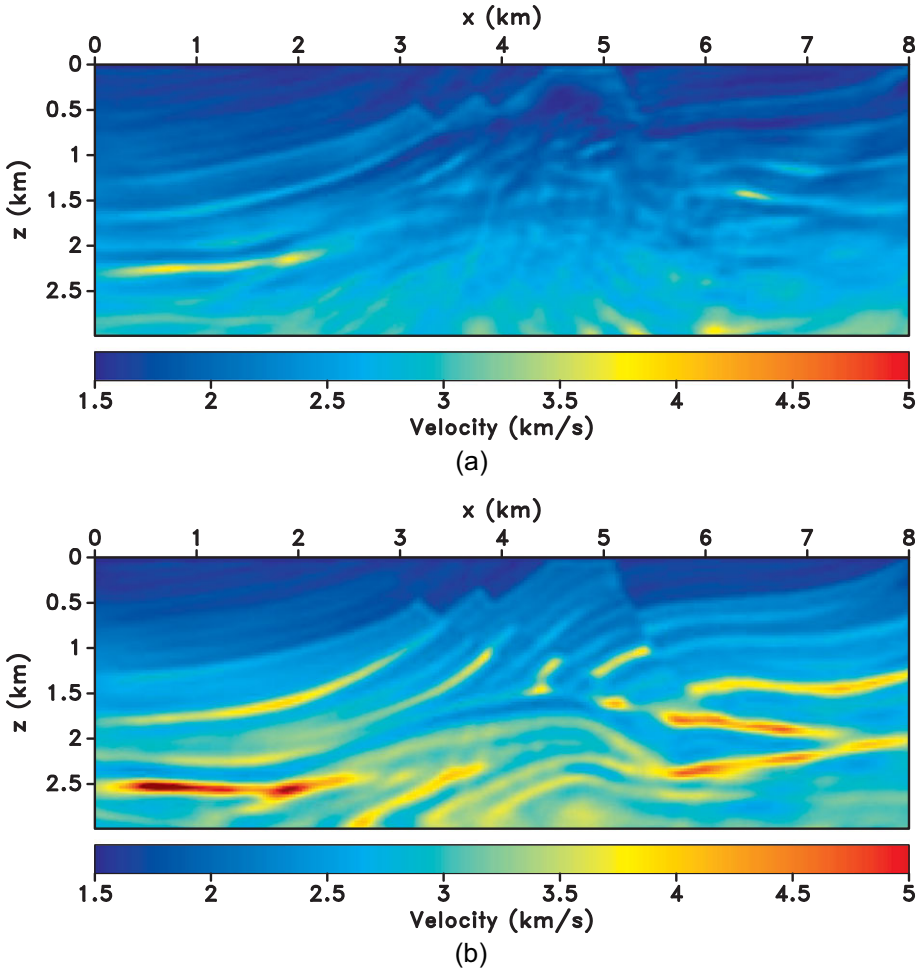


Figure 14 FWI result obtained using (a) the initial model for image-domain tomography and (b) the image-warping wavefield tomography inverted model in Fig. 12a as initial guess. The data are low-pass filtered with high cutoff frequency of 3 Hz. Observe that the initial model for image-domain wavefield tomography leads to cycle-skipping in the data domain. The improved kinematics of the model obtained by image-warping wavefield tomography removes the cycle-skipping problem in the data domain and allows one to reconstruct a high-resolution model.

window implies a drop in the values of the estimated correlation coefficients and thus a decrease in the quality of the estimated shifts. In penalized local correlation wavefield tomography, the construction of the penalty operators is based on geometrical assumptions on the shape of the correlation panels and their relationship with the local structural dip (Perrone and Sava 2012); when the two images do not have energy in the same local window (shallow areas) or a unique structural dip cannot be defined (discontinuities or unconformities), this geometrical relationship breaks, and additional artefacts are introduced. The displacement vector field is more robust than the penalized local correlations because of the iterative search and warping procedure used (Hale *et al.* 2008); thus, it allows us to obtain higher-quality velocity updates.

The main drawback of performing inversion in the shot-migrated image domain is related to the nature of the migration operator and to the associated migration artefacts. In highly refractive media, triplications and multi-pathing between the source and the receiver produce artefacts in the migrated image (Stolk and Symes 2004). The artefacts are due to the migration operator being the adjoint and not the inverse of the Born modelling operator; they have the same frequency content of the signal and, although stacking over experiments removes them from the final image, they severely contaminate the migrated images for individual shots. These kinematic artefacts represent the principal source of noise for our method: they do not satisfy the assumed relation between the directions of the image gradient and displacement

vector field; moreover, they introduce events with conflicting dips in the image that make the estimation of the displacement field noisier. Least squares migration (Nemeth, Wu, and Schuster 1999) may be a possible solution in the case of white Gaussian noise but would increase the computational cost and time. Unfortunately, migration artefacts do not follow a Gaussian distribution; thus, least squares migration is ineffective in removing migration artefacts from single-shot images.

A substantial difference of image-warping wavefield tomography with respect to methods based on extended images (Shen and Symes 2008; Yang and Sava 2011), which involve all shots in the construction of the adjoint sources, is the fundamentally different tomographic behaviour. As described by Vasconcelos, Sava, and Douma (2010), extended images represent the reflection response for virtual sources constructed in the subsurface at specific locations through interferometry. These sources illuminate a wide range of directions; thus, the sensitivity kernels obtained from each point have a wide spatial spectral content. On the contrary, our method (and conventional differential semblance in the shot domain) constructs the adjoint sources from single images, each of which is illuminated from a very limited range of direction. The poor illumination of the model that individual images supply reflects in extremely smooth sensitivity kernels that do not capture high wavenumber features in the model.

In complicated models, such as the Marmousi model, regularization plays an important role in balancing the strength of the update and in removing the natural bias that comes from the uneven illumination and the complexity of the subsurface structures. The smoothness and regularity of the estimated displacement vector field is also very important since it controls the quality of the adjoint sources and the smoothness of the sensitivity kernels. Nonetheless, image-warping wavefield tomography appears to be characterized by smooth gradients that can be used to easily correct large errors in the velocity models and then to bootstrap or regularize velocity analysis schemes that supply higher resolution.

CONCLUSIONS

The image-difference objective function for wavefield tomography in the migrated-shot domain can be made more robust using image-warping. The displacement vector field that warps a migrated image into the image obtained from a neighboring experiment removes cycle-skipping because every reflector in one image is mapped into the corresponding event in the second image.

Our method is more effective than local image correlation wavefield tomography for shallow reflectors, i.e., when the ratio between the source separation and the depth of the reflector is high. However, kinematic artefacts contaminate the image in highly refractive media and destabilize the computation of the sensitivity kernel. Nonetheless, in non-pathological cases, the image-warping objective function can measure the distance between different models and identify the most coherent images, thus enabling local gradient-based optimization.

An inversion test on the Marmousi model shows that image-warping wavefield tomography is able to correct strong errors in the velocity model where the geometry of the reflectors is relatively simple and regular. Although kinematic artefacts in the single-shot migrated images prevent a successful inversion in complex areas, the correct reconstruction of the very low wavenumber components of the model makes it possible to use the result of image-warping wavefield tomography as an initial model for higher resolution techniques and to avoid cycle-skipping in the data domain.

ACKNOWLEDGMENTS

This work was supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures. The reproducible numerical examples in this paper use the Madagascar open-source software package freely available from http://www.ahay.org and the Mines JTK freely available at https://github.com/dhale/jtk. We also thank the two anonymous reviewers and the Associate Editor, whose constructive comments helped improve the paper.

REFERENCES

Al-Yahya, K. 1989. Velocity analysis by iterative profile migration. *Geophysics* **54**, 718–729.

Albertin, U., Sava, P., Etgen, J. and Maharramov, M. 2006. Adjoint wave-equation velocity analysis. SEG Techincal program, Expanded Abstract, 3345–3349.

Baek, H., Calandra, H. and Demanet, L. 2014. Velocity estimation via registration-guided least-squares waveform inversion. *Geophysics* **79**, R79–R89.

Biondi, B. and Almomin, A. 2013. Tomographic full-waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis. *The Leading Edge* **32**, 1074–1080.

Biondi, B. and Sava, P. 1999. Wave-equation migration velocity analysis. SEG Techincal program, Expanded Abstract, 1723–1726.

Carcione, J. 2007. *Wave Fields in Real Media Wave Propagation in Anisotropic, Anelastic, Porous and Electromagnetic Media*, Vol. **38** Handbook of Geophysical Exploration. Elsevier. ISBN 9780080464084.

Chauris, H. and Plessix, R.-E. 2012. Investigating the differential waveform inversion. 74th EAGE Conference and Exhibition, Copenhagen, Denmark, WS06.

Chavent, G. and Jacewitz, C.A. 1995. Determination of background velocities by multiple migration fitting. *Geophysics* **60**, 476–490.

Claerbout, J.F. 1985. *Imaging the Earth's Interior*. Blackwell Publishing. ISBN 0865423040.

Faye, J.-P. and Jeannot, J.-P. 1986. Prestack migration velocity analysis from focusing depth imaging. SEG Techincal program, Expanded Abstract, 438–440.

Fichtner, A., Bunge, H.-P. and Igel, H. 2006. The adjoint method in seismology I. theory. *Physics of the Earth and Planetary Interiors* **157**, 86–104.

Fleury, C. and Perrone, F. 2012. Bi-objective optimization for the inversion of seismic reflection data: Combined FWI and MVA. SEG Techincal program, Expanded Abstract, 1–6.

Hale, D. 2009. A method for estimating apparent displacement vectors from time-lapse seismic images. *Geophysics* **74**, V99–V107.

Hale, D. 2013. Dynamic image warping of seismic images. *Geophysics* **78**, S105–S115.

Hale, D., Cox, B. and Hatchell, P.J. 2008. Apparent horizontal displacements in time-lapse seismic images. SEG Techincal program, Expanded Abstract, 3169–3173.

Kelly, S., Martinez, J.R., Tsimelzon, B. and Crawley, S. 2010. A comparison of inversion results for two full-waveform methods that utilize the lowest frequencies in dual-sensor recordings. SEG Techincal program, Expanded Abstract, 940–944.

Lions, J. and Magenes, E. 1972. *Non-Homogeneous Boundary Value Problems and Applications vol. 2*. Springer Verlag, Berlin Heidelberg. ISBN 3642652190.

Luo, Y. and Schuster, G.T. 1991. Wave-equation traveltime inversion. *Geophysics* **56**, 645–653.

Ma, Y. and Hale, D. 2013. Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion. *Geophysics* **78**, R223–R233.

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

Perrone, F. and Sava, P. 2012. Wavefield tomography based on local image correlation. 74th EAGE Conference and Exhibition, Copenhagen, Denmark, Expanded Abstracts, P233.

Perrone, F. Sava, P., Andreoletti, C. and Bienati, N. 2014. Linearized wave-equation migration velocity analysis by image warping. *Geophysics* **79**, S35–S46.

Plessix, R.-E. 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *International Journal of Geophysics* **167**, 495–503.

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

Santosa, F. and Symes, W.W. 1989. An analysis of least-squares velocity inversion. Society of Exploration Geophysicists. ISBN 9780931830785.

Sattlegger, J.W. 1975. Migration velocity determination: Part I. Philosophy. *Geophysics* **40**, 1–5.

Sava, P. and Biondi, B. 2004. Wave-equation migration velocity analysis. I. Theory. *Geophysical Prospecting* **52**, 593–606.

Sava, P., Biondi, B. and Etgen, J. 2005. Wave-equation migration velocity analysis by focusing diffractions and reflections. *Geophysics* **70**, U19–U27.

Sava, P. and Fomel, S. 2003. Angle-domain common-image gathers by wavefield continuation methods. *Geophysics* **68**, 1065–1074.

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

Sirgue, L. and Pratt, R. 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies. *Geophysics* **69**, 231–248.

Stolk, C.C. and Symes, W.W. 2004. Kinematic artifacts in prestack depth migration. *Geophysics* **69**, 562–575.

Symes, W.W. 1993. A differential semblance criterion for the inversion of multioffset seismic reflection data. *Journal of Geophysical Research* **98**, 2061–2073.

Symes, W.W. 2008. Migration velocity analysis and waveform inversion. *Geophysical Prospecting* **56**, 765–790.

Symes, W.W. and Carazzone, J.J. 1991. Velocity inversion by differential semblance optimization. *Geophysics* **56**, 654–663.

Tarantola, A. 1984. Inversion of seismic reflection data in the acoustic approximation. *Geophysics* **49**, 71–92.

Vasconcelos, I. Sava, P. and Douma, H. 2010. Nonlinear extended images via image-domain interferometry. *Geophysics* **75**, SA105–SA115.

Verschuur, D.J. 1991. Surface-related multiple elimination, an inversion approach. PhD thesis, Delft University of Technology, Netherlands.

Warner, M., Morgan, J., Umpleby, A., Stekl, I. and Guasch, L. 2012. Which physics for full-wavefield seismic inversion? 74th EAGE Conference and Exhibition, Copenhagen, Denmark, Expanded Abstracts, W029.

Xu, S., Wang, D., Chen, F. and Zhang, Y. 2012. Full waveform inversion for reflected seismic data. 74th EAGE Conference and Exhibition, Copenhagen, Denmark Expanded Abstracts, W024.

Yang, D., Malcolm, A. and Fehler, M. 2014. Using image warping for time-lapse image domain wavefield tomography. *Geophysics* **79**, WA141–WA151.

Yang, T. and Sava, P. 2011. Image-domain waveform tomography with two-way wave-equation. SEG Technical Program Expanded Abstracts, 2591–2596.

Zhang, Y., Xu, S., Tang, B., Bai, B., Huang, Y. and Huang, T. 2010. Angle gathers from reverse time migration. *The Leading Edge* **29**, 1364–1371.

APPENDIX

ADJOINT-STATE CALCULATION FOR IMAGE-WARPING OBJECTIVE FUNCTION

In this appendix, we show the adjoint-state calculations for the objective function in equation (2). For the sake of simplicity, we assume there is a pair of shots shots R_i (**x**) and R_{i+1} (**x**). To facilitate the understanding and help the reader,

we use a notation borrowed from Fichtner *et al.* (2006), who show the rationale at the base of the adjoint-state method and how the adjoint wavefields are actually derived.

Let us rewrite the objective function as an inner product:

$$\begin{aligned}\mathcal{J}_{WARP}(m) &= \tfrac{1}{2}\int (\nabla R_i(\mathbf{x})\cdot\mathbf{u}(\mathbf{x}))^2\, d\mathbf{x} \\ &= \tfrac{1}{2}\langle 1,\, f(\mathbf{x})\rangle,\end{aligned} \tag{A1}$$

where $f(\mathbf{x}) = (\nabla R_i(\mathbf{x})\cdot\mathbf{u}(\mathbf{x}))^2$, the angle brackets represent integration over the space variable as bilinear form with arguments 1 and $f(\mathbf{x})$, $m(\mathbf{x})$ represents the model parameter function, and $\mathbf{x} = (x, y, z)$ is the position vector. In the following, for the sake of simplicity, we will derive expressions assuming a 2D geometry and $\mathbf{x} = (x, z)$. We also drop the dependence on the space variable, unless it is necessary for the clarity of the expression. The dependence of the objective function from the source and receiver wavefields, $p_{s\,i}(\mathbf{x}, t)$ and $p_{r\,i}(\mathbf{x}, t)$, is implicit in the migrated image $R_i(\mathbf{x})$ Claerbout (1985) and in the displacement field $\mathbf{u}(\mathbf{x})$ that warps $R_i(\mathbf{x})$ into $R_{i+1}(\mathbf{x})$ (see equation (4)).

We can write the gradient of the objective function in equation (A1) as follows:

$$\begin{aligned}D_m\mathcal{J}_{WARP} &= \partial_{p_{s\,i}}\mathcal{J}_{WARP}D_m p_{s\,i} + \partial_{p_{r\,i}}\mathcal{J}_{WARP}D_m p_{r\,i} \\ &= \langle 1, \partial_{p_{s\,i}} f D_m p_{s\,i}\rangle + \langle 1, \partial_{p_{r\,i}} f D_m p_{r\,i}\rangle,\end{aligned} \tag{A2}$$

where $\partial_{p_{s\,i}}$ and $\partial_{p_{r\,i}}$ are the Fréchet derivatives with respect to the state variables, i.e., the source and receiver wavefields, and D_m represents the total derivative with respect to the model parameters. The derivatives $D_m p_{s\,i}$ and $D_m p_{r\,i}$ are the Fréchet derivatives of the state variables with respect to the model parameters; because of the dimensionality of the model, these terms are impossible to compute and store. The definition of the adjoint problem allows one to bypass the computation of these Fréchet derivatives, and for acoustic wave propagation, the adjoint problem is conveniently the solution of the adjoint wave equation with a modified source term (Fichtner *et al.* 2006). The terms $\partial_{p_{s\,i}} f$ and $\partial_{p_{r\,i}} f$ represent the so-called adjoint sources, i.e., the sources of the adjoint problem. In the following, we focus on the calculation of these adjoint sources since they are the key element in the adjoint-state method implementation.

The displacement vector $\mathbf{u}(\mathbf{x})$ is computed by extracting the argument (correlation lag) that maximizes the local correlation. We use an implicit function (Luo and Schuster 1991) to link the vector $\mathbf{u}(\mathbf{x})$ to a differentiable function of the background wavefields. Once we have defined the implicit function, we can use implicit differentiation to obtain the derivative of the displacement with respect to the state variables.

Let us consider the first term in equation (A2), which represents the source-side contribution to the gradient of the objective function. Using the chain rule of differentiation, we have

$$\partial_{p_{s\,i}} f = \partial_{\nabla R_i} f \partial_{p_{s\,i}} \nabla R_i + \partial_{\mathbf{u}} f \partial_{p_{s\,i}} \mathbf{u}. \tag{A3}$$

The first term is easy to compute since the function $f(\nabla R_i, \mathbf{u}) = (\nabla R_i(\mathbf{x})\cdot\mathbf{u}(\mathbf{x}))^2$ is a quadratic function of the image gradient, the ∇ operator is linear, and the image $R_i(\mathbf{x})$ is a linear functional of the source wavefield $p_{s\,i}(\mathbf{x}, t)$. Straightforward calculations give

$$\partial_{\nabla R_i} f \partial_{p_{s\,i}} \nabla R_i = (\nabla R_i\cdot\mathbf{u})\,\mathbf{u}\cdot\nabla p_{r\,i}. \tag{A4}$$

In the second term in equation (A2), $\partial_{\mathbf{u}} f \partial_{p_{s\,i}} \mathbf{u}$, the derivative of $f(\nabla R_i, \mathbf{u})$ with respect to the displacement vector $\mathbf{u}(\mathbf{x})$ is easy to compute because of the quadratic dependence of the objective function on $\mathbf{u}(\mathbf{x})$. Simple calculations give

$$\partial_{\mathbf{u}} f \partial_{p_{s\,i}} \mathbf{u} = (\nabla R_i\cdot\mathbf{u})\,\nabla R_i\cdot\partial_{p_{s\,i}}\mathbf{u}. \tag{A5}$$

On the other hand, the derivative of the displacement vector $\mathbf{u}(\mathbf{x})$ with respect to the state variable requires to define the displacement vectors in a different but equivalent fashion that would allow us to perform differentiation. Since the displacement vector corresponds to the stationary point of the local correlation, in order to compute the derivative with respect to the state variables, we use the following implicit definition:

$$\begin{aligned}&\mathbf{u} \text{ such that } F(\mathbf{x}, \boldsymbol{\lambda} = \mathbf{u}) \\ &= \nabla_{\boldsymbol{\lambda}} \int w(\mathbf{x}-\boldsymbol{\xi})\, R_i\left(\boldsymbol{\xi} - \frac{\boldsymbol{\lambda}}{2}\right) R_{i+1}\left(\boldsymbol{\xi} + \frac{\boldsymbol{\lambda}}{2}\right) d\boldsymbol{\xi}\Big|_{\boldsymbol{\lambda}=\mathbf{u}} \\ &= 0.\end{aligned} \tag{A6}$$

We can expand the expression for $F(\mathbf{x}, \boldsymbol{\lambda} = \mathbf{u})$ using the chain rule:

$$\begin{aligned}F\ &(\mathbf{x}, \boldsymbol{\lambda} = \mathbf{u}) \\ &= -\tfrac{1}{2}\int w(\mathbf{x}-\boldsymbol{\xi})\,\nabla R_i\left(\boldsymbol{\xi} - \tfrac{\mathbf{u}}{2}\right) R_{i+1}\left(\boldsymbol{\xi} + \tfrac{\mathbf{u}}{2}\right) d\boldsymbol{\xi} \\ &\quad + \tfrac{1}{2}\int w(\mathbf{x}-\boldsymbol{\xi})\, R_i\left(\boldsymbol{\xi} - \tfrac{\mathbf{u}}{2}\right)\nabla R_{i+1}\left(\boldsymbol{\xi} + \tfrac{\mathbf{u}}{2}\right) d\boldsymbol{\xi} = 0.\end{aligned} \tag{A7}$$

We can implicitly differentiate the function $F(\mathbf{x}, \boldsymbol{\lambda} = \mathbf{u})$ with respect to $\mathbf{u}(\mathbf{x})$ and the state variable $p_{s\,i}(\mathbf{x}, t)$ to obtain an expression for the derivative $\partial_{p_{s\,i}}\mathbf{u}$ in equation (A3).

The implicit function $F(\mathbf{x}, \mathbf{u}) = 0$ depends on the background wavefield $p_{s\,i}(\mathbf{x}, t)$ via the migrated image $R_i(\mathbf{x})$; by writing the differential with respect to $p_{s\,i}(\mathbf{x}, t)$ and $\mathbf{u}(\mathbf{x})$

$$\partial_{p_{s\,i}} F\, dp_{s\,i} + \partial_{\mathbf{u}} F\, d\mathbf{u} = 0, \tag{A8}$$

we can derive an expression for the variation of the displacement vector $\mathbf{u}(\mathbf{x})$ with respect to the wavefields $p_{s_i}(\mathbf{x}, t)$

$$\partial_{p_{s_i}} \mathbf{u} = -(\partial_{\mathbf{u}} F)^{-1} \partial_{p_{s_i}} F, \tag{A9}$$

given that $\partial_{\mathbf{u}} F$ is not singular.

The computation of $\partial_{p_{s_i}} F$ is relatively straightforward since F depends only on linear functions of the migrated image $R_i(\mathbf{x})$. The derivative $\partial_{\mathbf{u}} F$ is the multi-dimensional generalization of the denominator in in (Luo and Schuster 1991, equation (8(a)):

$$\begin{aligned}\partial_{\mathbf{u}} F &= \nabla_{\mathbf{u}} F \\ &= \frac{1}{4}\int w(\mathbf{x}-\boldsymbol{\xi})\,\nabla\nabla R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right) R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right) d\boldsymbol{\xi} \\ &\quad -\frac{1}{4}\int w(\mathbf{x}-\boldsymbol{\xi})\,\nabla R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right)\nabla R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right) d\boldsymbol{\xi} \\ &\quad -\frac{1}{4}\int w(\mathbf{x}-\boldsymbol{\xi})\,\nabla R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right)\nabla R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right) d\boldsymbol{\xi} \\ &\quad +\frac{1}{4}\int w(\mathbf{x}-\boldsymbol{\xi})\, R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right)\nabla\nabla R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right) d\boldsymbol{\xi},\end{aligned} \tag{A10}$$

where, in $2D$,

$$\nabla\nabla R_i\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right) = \begin{bmatrix} \dfrac{\partial^2 R_i\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right)}{\partial x^2} & \dfrac{\partial^2 R_i\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right)}{\partial x \partial z} \\ \dfrac{\partial^2 R_i\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right)}{\partial z \partial x} & \dfrac{\partial^2 R_i\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right)}{\partial z^2} \end{bmatrix} \tag{A11}$$

and

$$\begin{aligned}&\nabla R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right)\nabla R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right) \\ &= \begin{bmatrix} \dfrac{\partial R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right)}{\partial x}\dfrac{\partial R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right)}{\partial x} & \dfrac{\partial R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right)}{\partial x}\dfrac{\partial R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right)}{\partial z} \\ \dfrac{\partial R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right)}{\partial z}\dfrac{\partial R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right)}{\partial x} & \dfrac{\partial R_i\left(\boldsymbol{\xi}-\frac{\mathbf{u}}{2}\right)}{\partial z}\dfrac{\partial R_{i+1}\left(\boldsymbol{\xi}+\frac{\mathbf{u}}{2}\right)}{\partial z} \end{bmatrix}\end{aligned} \tag{A12}$$

are 2×2 matrices of second derivatives and products of first derivatives, respectively. Both λ and $\mathbf{u}$ are spatial vectors; thus, the gradient with respect to them is identical to the spatial gradient of the image with respect to the spatial coordinate vector $\mathbf{x}$ except for a scaling factor. Notice that, since $(\partial_{\mathbf{u}} F)^{-1}$ is a 2×2 matrix-valued function defined at every image point $\mathbf{x}$ and $\partial_{p_{s_i}} F$ is a 2×1 vector-valued function at every point $\mathbf{x}$, their product is a 2×1 vector-valued function, and then the dot product with the spatial image gradient ∇R_i in equation (A5) makes mathematical sense. The generalization to 3D is straightforward. This expression is similar to in (Luo and Schuster 1991, equations (7) and (8(a)).

The analysis of the differential of objective function in equation (2) with respect to the model parameters shows the mathematical complexity of implementing this solution for velocity analysis. The multi-generalization of the procedure developed by Luo and Schuster (1991) poses a number of implementation challenges: in the time domain, we have 1D shifts and in (Luo and Schuster 1991, equation(7a)) involves *scalar* functions. Moreover, Luo and Schuster (1991) focused only on transmitted waves and thus measure a single scalar-valued travel-time error $\Delta\tau$. Ma and Hale (2013) generalized the travel-time inversion approach to reflections by using dynamic image-warping but still dealt with scalar-valued travel-time error functions. In the image domain, the shifts are vectors, and we have to deal with *vector* and *matrix*-valued functions, which considerably increases the complexity of the gradient computation. The local correlation implementation increases the dimensionality of the problem and then the cost, and picking of specific location in the image is necessary for an efficient implementation. The approach developed by Yang *et al.* (2014) relies on scalar shifts and would require additional processing to capture the multi-dimensional nature of image-domain apparent displacements. The increase in computational cost can be easily estimated by considering the extra floating-point operations needed to invert the matrices in equation (A10). Both our approach and the adjoint-state calculations require the computation of the apparent shifts between the input images and first-order spatial derivatives. However, in 3D, the adjoint-state calculations require the inversion of a 3×3 system constructed from higher order derivatives of the migrated images for every point in the image. Even neglecting the computational cost of the higher order derivatives, inverting a 3×3 system of equations has a computational cost of approximately $3^3 = 27$ floating-point operations. This means that, in 3D, the adjoint-state calculations are more computationally expensive than our proposed methodology by a factor of 27, which is more than an order of magnitude.

The increased complexity is self-evident by comparing the equations in this appendix with equations (10), (11), and (12).