

Land seismic multiparameter full waveform inversion in elastic VTI media by simultaneously interpreting body waves and surface waves with an optimal transport based objective function

Weiguang He¹, Romain Brossier¹, Ludovic Métivier^{1,2} and René-Édouard Plessix³

¹Univ. Grenoble Alpes, ISTERre, 38000 Grenoble, France. E-mail: weiguang.he@univ-grenoble-alpes.fr

²Univ. Grenoble Alpes, CNRS, ISTERre, 38000 Grenoble, France

³Shell Global Solutions International, Grasweg 31, 1031HW Amsterdam, The Netherlands

Accepted 2019 September 12. Received 2019 September 2; in original form 2018 November 26

SUMMARY

Land seismic multiparameter full waveform inversion in anisotropic media is challenging because of high medium contrasts and surface waves. With a data-residual least-squares objective function, the surface wave energy usually masks the body waves and the gradient of the objective function exhibits high values in the very shallow depths preventing from recovering the deeper part of the earth model parameters. The optimal transport objective function, coupled with a Gaussian time-windowing strategy, allows to overcome this issue by more focusing on phase shifts and by balancing the contributions of the different events in the adjoint-source and the gradients. We first illustrate the advantages of the optimal transport function with respect to the least-squares one, with two realistic examples. We then discuss a vertical transverse isotropic (VTI) example starting from a quasi 1-D isotropic initial model. Despite some cycle-skipping issues in the initial model, the inversion based on the windowed optimal transport approach converges. Both the near-surface complexities and the variations at depth are recovered.

Key words: land seismic data; multi-parameter inversion; elastic VTI model; optimal transport.

1 INTRODUCTION

Full waveform inversion (FWI) aims at reconstructing the sub-surface material properties by fitting synthetic seismic waveforms to observed waveforms (Lailly 1983; Tarantola 1984). Waveform inversion is nowadays commonly applied in seismology with Earthquake data (Tromp *et al.* 2005; Fichtner *et al.* 2008; Tape *et al.* 2009; Komatitsch *et al.* 2016; Sager *et al.* 2017) or in applied geophysics with marine active data (Sirgue *et al.* 2008; Plessix & Perkins 2010; Routh *et al.* 2011; Warner *et al.* 2013; Vigh *et al.* 2013; Operto *et al.* 2015). On-shore acquisitions record high-amplitude surface waves related to the free surface condition with strong elastic effects. The applications of FWI with land active seismic data are thus more challenging than offshore. These surface waves, also known as ground roll, are often considered as noise and are removed for the interpretation of deeper reflection phases. However, removing surface waves may not be an easy task, in particular in complex environments.

Fundamental mode of Rayleigh waves propagates along the free surface with a velocity slower than the shear velocity. Surface waves are energetic at the surface, but their depth of penetration is shallow because their amplitudes decay exponentially with depth. The

information about the near-surface complexities along the source–receiver path is accumulated and mixed in the surface-wave events recorded in time. This information is contained in their dispersive nature, and also in the possible conversion to body waves. Surface waves propagate long distance while still being energetic, as their geometrical spreading is associated to a two-dimension propagation in 3-D (and a one-dimension propagation in 2-D). On the contrary, body waves sample the whole volume suffering a faster amplitude decay by the geometrical spreading. Consequently, a typical land seismogram where the receivers are located at the surface is dominated by the energetic but shallow penetrating surface waves, making the deep imaging a challenge. Standard FWI based on least-squares misfit function containing surface waves thus faces three major difficulties. (1) The body-wave influence is rather weak and the model update will be essentially driven by surface waves. (2) Surface waves easily induce cycle-skipping effects at intermediate to long offsets due to their strongly dispersive and oscillatory nature. On the other hand, surface waves are preferentially confined to low frequencies, which reduces the potentiality of cycle skipping. (3) The dominant forward-scattering propagation prevents simple localization of scatterers; note however that this issue is less severe with denser acquisition.

Because of the above mentioned difficulties, the first approach to deal with land seismic data is to avoid using surface waves. The observed seismic data are pre-processed to remove them, and a time window can be further used to select the early body waves during the inversion, such as in the real data application presented in Bharti *et al.* (2016). An acoustic forward modelling engine can be used in the low frequency range to accelerate the computation and reduce the memory requirement (Pratt & Worthington 1990; Operto *et al.* 2004; Ravaut *et al.* 2004; Plessix & Perkins 2010; Adamczyk *et al.* 2014; Stopin *et al.* 2014). In complex area, acoustic approaches may reach their limits due to the presences of large elastic contrasts (Plessix *et al.* 2013). However, focusing on the phase information (such as by weighting each trace with its root mean square value) helps extending the validity domain of the acoustic approximation (Shen 2010). Elasticity should be taken into account in the forward modelling engine (Barnes & Charara 2008; Mulder & Plessix 2008; Marelli *et al.* 2012; Cance & Capdeville 2015) when (1) high-contrast velocity structure exists and/or (2) high resolution subsurface image is needed. An intermediate method between acoustic and elastic FWI is obtained by modifying the free surface boundary conditions in the modeling engine of synthetic fields (Plessix & Pérez Solano 2015; He *et al.* 2018). In such a case, the surface waves are not generated, but the PS and SP mode conversions inside the Earth are taken into account. The inversion workflow is similar with acoustic FWI but the forward modelling engine is based on this modified elastic wave propagator. In the frame of isotropic elastic FWI, Plessix & Pérez Solano (2015) compare the inversions with free surface boundary condition and modified boundary condition, and identify the superiority of the latter approach.

The second strategy to deal with land seismic data is to incorporate the surface waves and consider sophisticated inversion strategies or wide-attraction-basin objective functions. Real data applications have been presented recently by Wittkamp (2016), Dokter *et al.* (2017) and Köhn *et al.* (2019). Surface waves are more sensitive to shear wave velocity parameter v_s than any other parameters such as P -wave velocity parameter v_p and density parameter ρ . Therefore, v_s is the main retrieved parameter when surface waves are incorporated. Trinh *et al.* (2018) propose a hierarchical data-based strategy to mitigate the surface wave issue. In each frequency band, a time window is applied in the first inversion step to interpret only the early body waves. Both of v_p and v_s are updated. Then in a following inversion step, the time window is removed and the entire seismograms are used to further improve the model. This hierarchical strategy to separate the different events in the data originates from a detailed FWI synthetic study with land seismic data conducted by Brossier *et al.* (2009), in which an exponentially decaying time window focusing on the first arrival is taken as a weighting function in the data domain (Shin & Cha 2008). At first, diving waves are selected to retrieve the long-wavelength of the Earth parameters (Jannane *et al.* 1989; Neves & Singh 1996). Gradually, reflections and surface waves are incorporated. The non-linearity of elastic FWI is greatly alleviated by smoothly incorporating late-arrival events. Besides modifying the inversion strategy to alleviate the cycle-skipping issue, other objective functions rather than the conventional least-squares objective function can be considered. Borisov *et al.* (2016) apply an envelope objective function (Bozdağ *et al.* 2011) to the SEAM II model (Oristaglio 2016) and obtain better inversion results than with the least-squares objection function. Furthermore, Yuan *et al.* (2015) decompose the seismic data using wavelet transform before applying the envelope function. Another strategy relies on the comparison of observed and synthetic data in the FK domain, by considering the absolute difference (Masoni

et al. 2013; Pérez Solano *et al.* 2014). The attraction basin of such FK domain objective functions is wider than the least-squares function in the data space, and it is shown to be robust to interpret the surface waves.

More recently, optimal-transport-based (OT) objective functions have been proposed in the frame of FWI (Engquist & Froese 2014; Engquist *et al.* 2016; Métivier *et al.* 2016a,b, 2018; Yang *et al.* 2018). The original motivation for the application of OT to FWI is the convexity of the objective function with respect to shifted patterns, seen as a proxy for convexity with respect to velocity changes, to mitigate cycle-skipping issues. In this study, we find that the OT formulation helps to balance the contribution of the different phases. The amplitude distribution of the OT adjoint source is much more uniform than that of the least-squares function, suggesting that the high amplitudes of surface waves will have less impact into the gradient build-up. Besides, the OT adjoint source is less oscillating than the synthetics or observed data. For this reason, we investigate the behaviour of OT for this problem of imaging complex near-surface anisotropic targets.

In this study, we investigate the combined interpretation of both surface waves and body waves from synthetic land data sets. In Section 2, we consider the least-squares objective function in the scheme of isotropic FWI in order to focus only on the surface wave problem. We discuss two approaches. In the first experiment, we present a two-stage inversion strategy by considering a well defined time window to separate body waves and surface waves. At first, we target the body waves by selecting the first 1.5 s seismic signals at offsets longer than 3.5 km. Then the time windows are removed and the entire seismograms are taken into account. In the second experiment, we switch to use a smooth Gaussian time window. In Section 3, we repeat these experiments with the OT function, which shows more robust performance. We thus propose a strategy to deal with surface waves based on this OT function. In Section 4, we illustrate our strategy with a different geological setting. In Section 5, we present a challenging VTI (vertical transverse isotropic) model. We conclude in the last Section 6.

2 LEAST-SQUARES WAVEFORM INVERSION

In this section, we discuss the behaviour of the least-squares objective function for land seismic FWI. In the first inversion experiment, we apply a time window to fit the body waves at the first stage and then the surface waves at the second stage. In the second test, we apply a Gaussian time window to mainly include the early body waves and the short-offset surface waves.

2.1 Inversion settings

In classical waveform inversion (Lailly 1983; Tarantola 1984; Virieux & Operto 2009), we estimate the earth parameters, \mathbf{m} , by minimizing the data residuals. In a time domain formulation, the objective function reads:

$$J(\mathbf{m}) = \frac{1}{2} \sum_{s,r} \int dt [W_{s,r}(t)(d_{s,r}^{syn}(t) - d_{s,r}^{obs}(t))]^2 \quad (1)$$

with $d_{s,r}^{syn}$ and $d_{s,r}^{obs}$ the synthetic and observed data traces at the receiver indexed by r due to a source indexed by s , $W_{s,r}$ a trace time window to potentially select the events of interest. With long-offset data, this window can for instance be used to select only the early arrivals as discussed later (He *et al.* 2018).

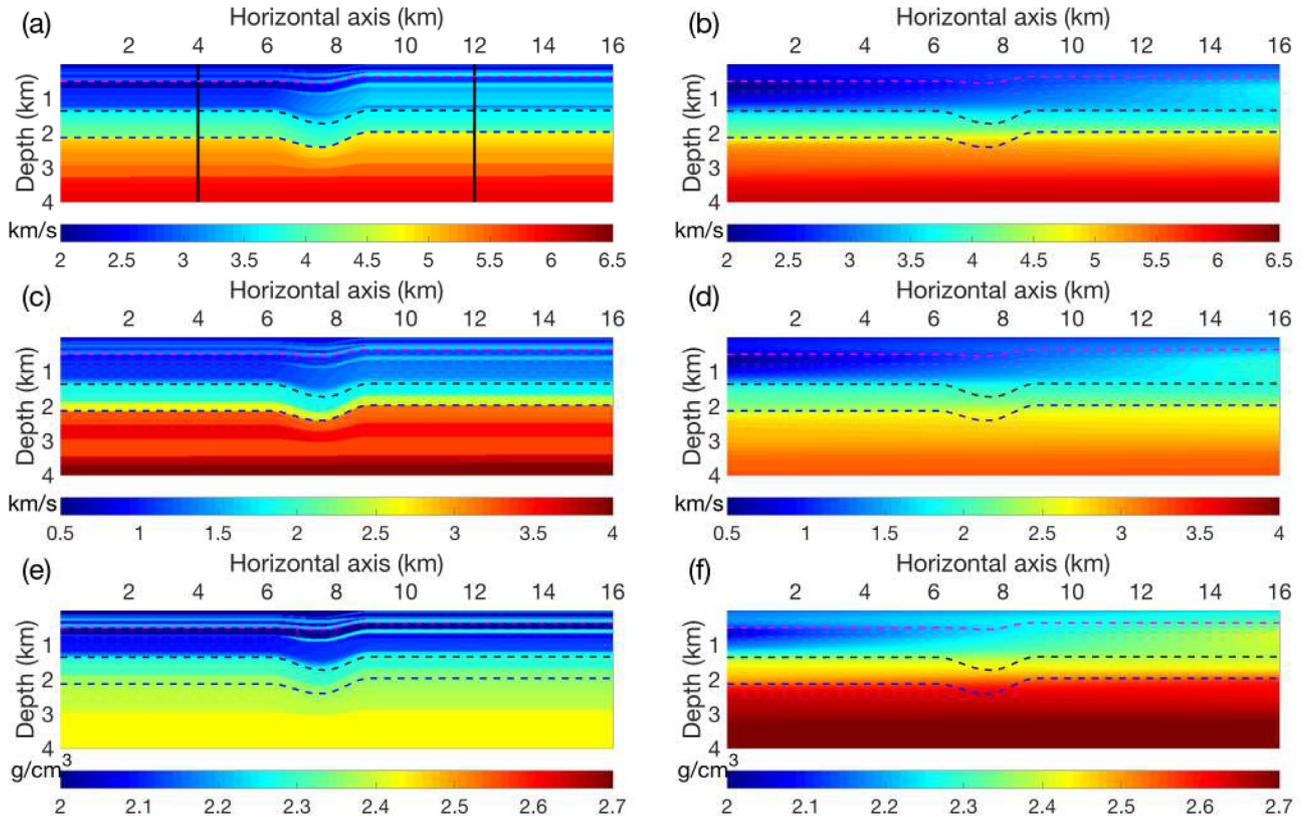


Figure 1. The true and initial isotropic models for the synthetic Middle-East example. (a) True v_p , (b) initial v_p , (c) true v_s , (d) initial v_s , (e) true ρ and (f) initial ρ . The dashed lines are the interfaces where the velocity changes rapidly.

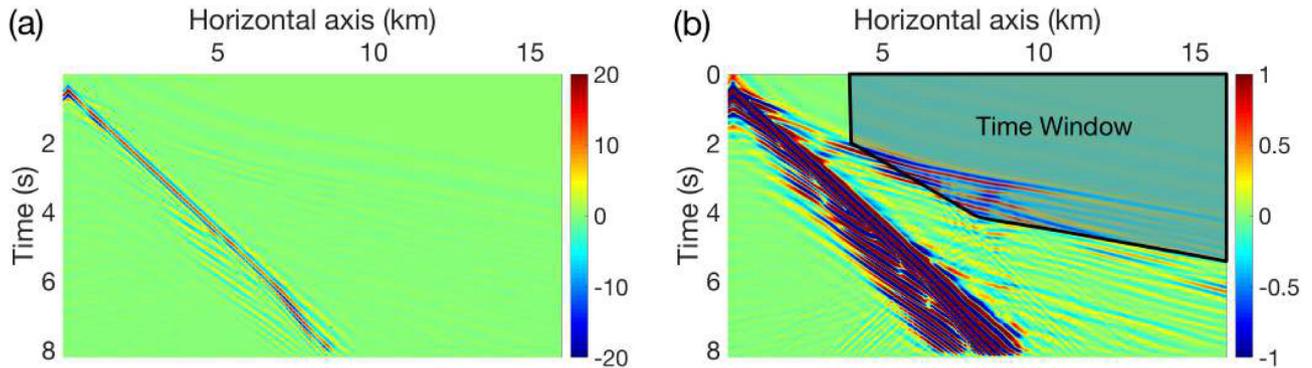


Figure 2. The observed data in different clippings. (a) The body waves are weak compared to surface waves. (b) To highlight the body waves, we need to clip the data with a factor 10–20.

In a multiparameter inversion, the choice of the earth parametrization plays an important role due to the non-linear behaviour of the objective function (Tarantola 1986). Here, we shall consider the VTI medium defined by the NMO (normal moveout) P -wave velocity v_n , the shear velocity v_s , the anelliptic parameter η , the stretched parameter δ and density ρ . This parameterization is adequate when we focus the inversion on the transmitted waves with the goal of retrieving the NMO velocity for further reflection imaging. This is a standard setting in exploration geophysics. Even if the analysis to select this parameterization (Plessix & Cao 2011; He & Plessix 2017) focuses mainly on the transmitted waves, we still use it when fitting the surface waves.

In the inversion results presented here, we do not invert for density, we recompute it from the NMO velocity through a heuristic

Gardner's relation, $\rho = 1.74v_n^{0.25}$. This relation may not be fully adequate for land and does not correspond to the $v_n - \rho$ relation of the true model of the synthetic example. However, we use it to update the density and absorb some of the amplitude effects. In addition, we do not directly invert for v_n , v_s , η and δ . We use a change of variables (Tromp *et al.* 2005). The unknown variables in the optimization stage are $(\log v_n, \log v_s, \log \sqrt{1 + 2\delta}, \log \sqrt{1 + 2\eta})$. The differential of the objective function J in eq. (1) is:

$$dJ = g_{v_n} d \log v_n + g_{v_s} d \log v_s + g_{\delta} d \log \sqrt{1 + 2\delta} + g_{\eta} d \log \sqrt{1 + 2\eta}. \quad (2)$$

with g_{v_n} , g_{v_s} , g_{δ} and g_{η} the different gradients, which are the partial

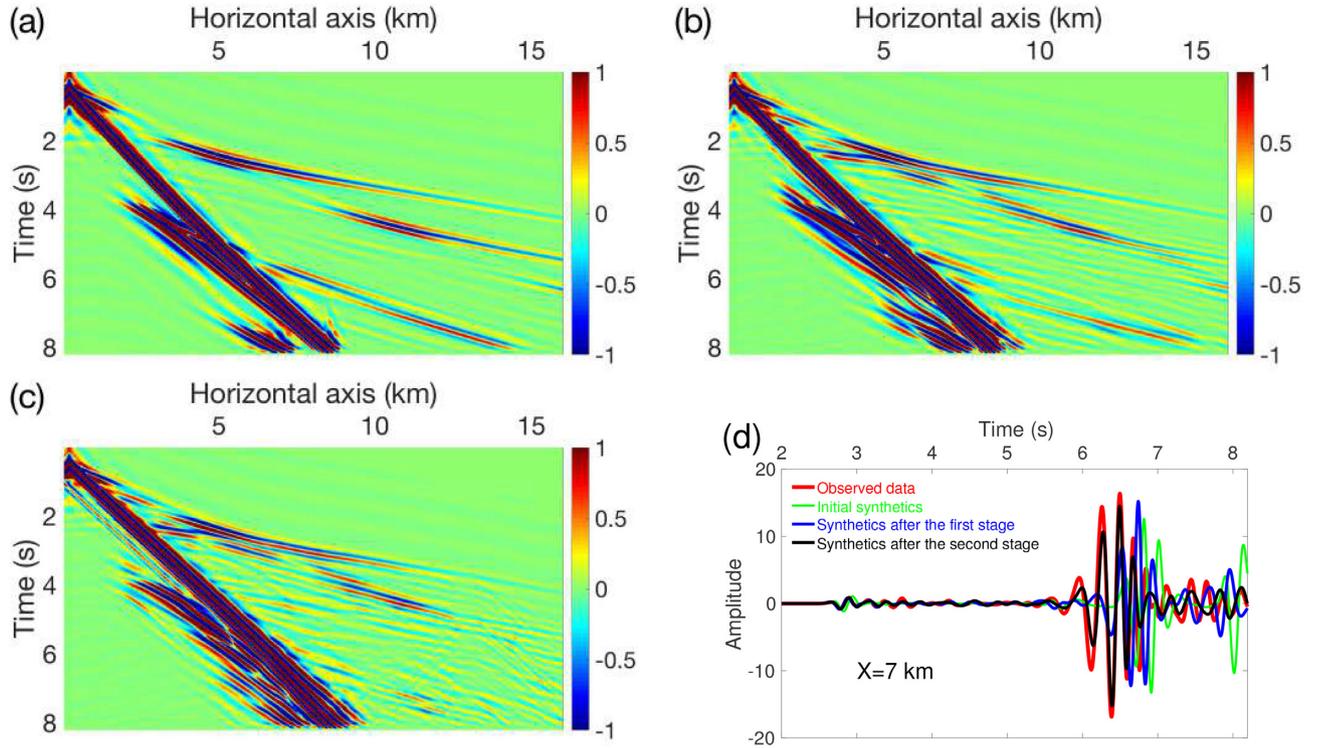


Figure 3. (a) Initial synthetics. (b) Synthetics after the first inversion stage, that is fitting the body waves. (c) Synthetics after the second inversion stage, that is fitting the entire seismograms. (d) Trace comparison.

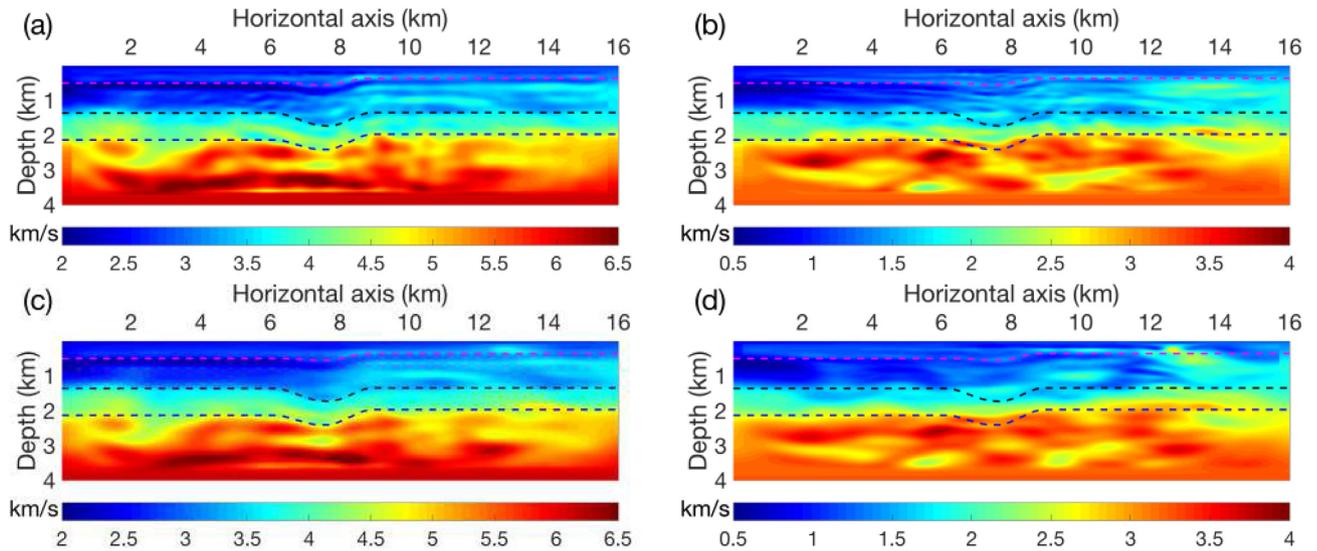


Figure 4. Inversion results for the (a) v_p and (b) v_s models after the first inversion stage (only the first 1.5 s after the first break are included). Panels (c) and (d) display the results after the second inversion stage (the entire seismogram is inverted).

derivatives of the objective function J with respect to the corresponding unknown variables.

Our 2-D forward simulator solving seismic wave propagation in elastic VTI media is based on a staggered grid finite difference scheme (Virieux 1986), with free surface boundary condition on top. We apply the same simulator, source wavelet, and grid size during the inversion. In eq. (1), the synthetics and observed data are computed with a same wide-band spectrum source wavelet (in this study a 3 Hz Ricker wavelet). We implement a multiscale inversion strategy by filtering the synthetic and observed data. In our

two-step inversion experiment presented in Section 2.3, we apply the time windows first and then apply the bandpass filter. Because the inter-receiver distance is large, spatial aliasing occurs in FK domain. To remove the surface waves, we have to apply a time window when the events are well focused in time (the broader the spectrum of the source wavelet, the more focused the seismic event becomes). In other inversion experiments, we apply the bandpass filter before applying the Gaussian windows. At each frequency band, we filter the seismograms with a zero-phase bandpass filter, which is extracted from the code SPEC3D (available online

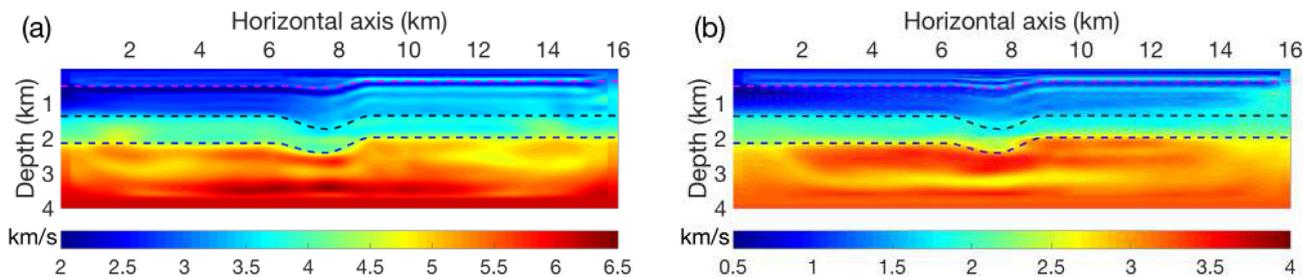


Figure 5. Inversion results with a least-squares function and a Gaussian window with $\sigma = 0.4$ s. (a) Final v_p . (b) Final v_s .

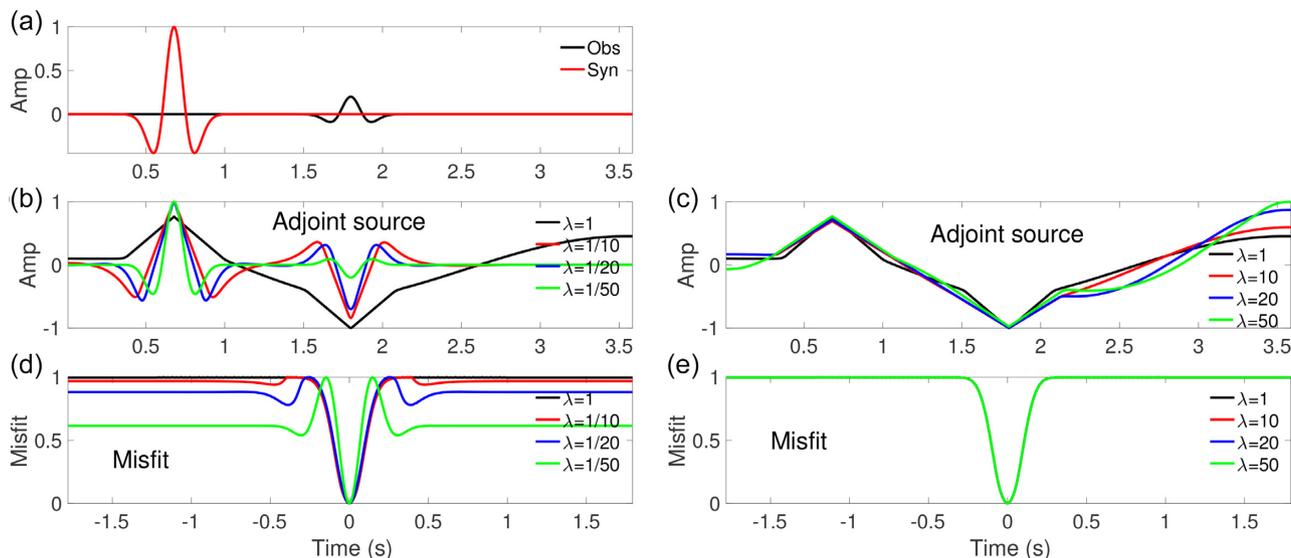


Figure 6. Adjoint sources for different bound constraints λ . (a) The observed and synthetic trace. (b) The KR-OT adjoint sources when $\lambda \leq 1$. (c) The KR-OT adjoint sources when $\lambda \geq 1$. (d) The misfit when $\lambda \leq 1$. (e) The misfit when $\lambda \geq 1$.

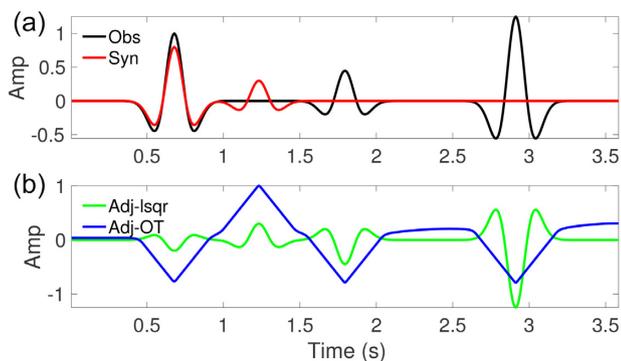


Figure 7. Behaviour of the KR-OT approach when three events are considered. (a) Observed and synthetic traces, (b) adjoint sources of least-squares function and KR-OT function. These events are balanced with KR-OT function.

<https://geodynamics.org>). The gradients are smoothed with an elliptic Gaussian filter. At the first scale, the standard deviation of the Gaussian smoothing filter is 750 m and the vertical one is 150 m. At the final scale, the standard deviation is 150 m and the vertical one is 30 m. To compensate for the decay with depth of the gradients, we multiply the gradient by $\sqrt{z + 300}$. Finally, the minimization of the objective function is performed with the *l*-BFGS quasi-Newton algorithm (Nocedal 1980).

2.2 Isotropic synthetic example: description

To study the waveform inversion of land data and the strategy to invert both surface and body waves, we first consider an isotropic synthetic case mimicking a Middle-East geology (He *et al.* 2018). This example contains large elastic parameter variations due to an alternation between sand-shale layers and carbonate layers (Fig. 1). The thicknesses of the different layers are several hundred meters. To help the analysis of the results, we have drawn three dashed lines that correspond to the main layers. From two v_p logs at $X = 4$ km and $X = 12$ km, we build an initial model for inversion. First, we heavily smooth the logs, then we interpolate them to obtain a 2-D v_p model. The shear velocity v_s is deduced from v_p with the following relation, $v_s = (v_p - 1360)/1160$ for $v_p < 3500$ and $v_s = 0.53v_p$ for $v_p > 3500$. The initial parameters are displayed in Fig. 1.

In this work, we assume that the data contains low frequencies down to 1.5 Hz as performed in different field acquisitions (Baeten *et al.* 2013; Plessix *et al.* 2013; Stopin *et al.* 2014). We shall invert the data in the frequency band 1.5 to 6.5 Hz and consider six scales: 1.5–2.5 Hz, 1.5–3 Hz, 1.5–3.5 Hz, 1.5–4.5 Hz, 1.5–6.0 Hz and 1.5–6.5 Hz. The spatial discretization of the model parameters is 20 m. The time discretization is 1 ms with a total recording length of 8.192 s. We consider fixed-spread land seismic surface acquisition (so the sources and receivers are at zero depth). We consider 48 equidistant sources (velocity force in the vertical direction), and 300 equidistant receivers (recording vertical velocity component). A typical shot gather is displayed in Fig. 2. We also display the shot gather obtained in the initial model (Fig. 3a). The waveform recorded below the

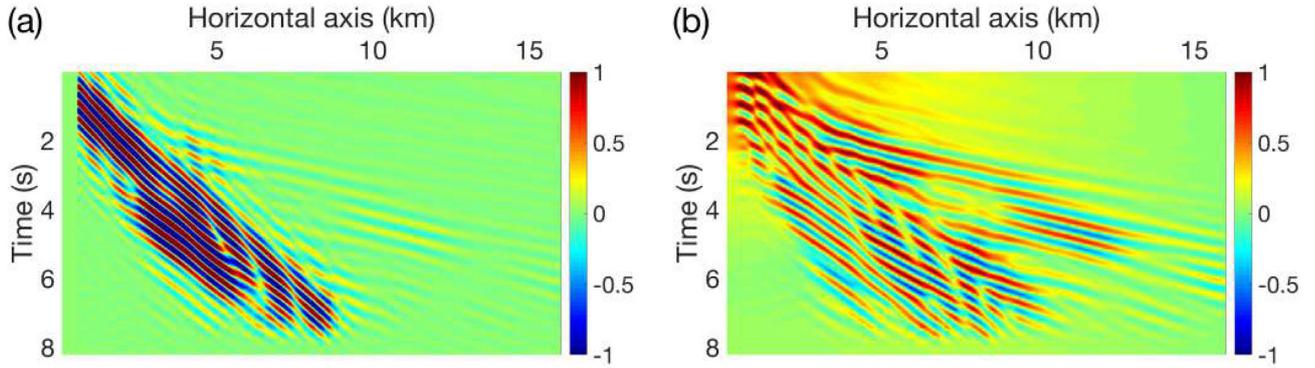


Figure 8. (a) Least-squares adjoint source and (b) KR-OT adjoint source at the first frequency band from 1.5 to 2.5 Hz.

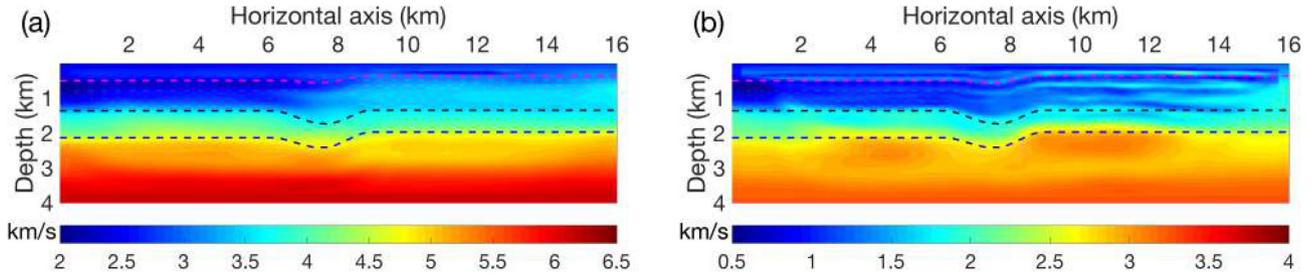


Figure 9. Inverted (a) v_p and (b) v_s with the KR-OT formulation without time windows.

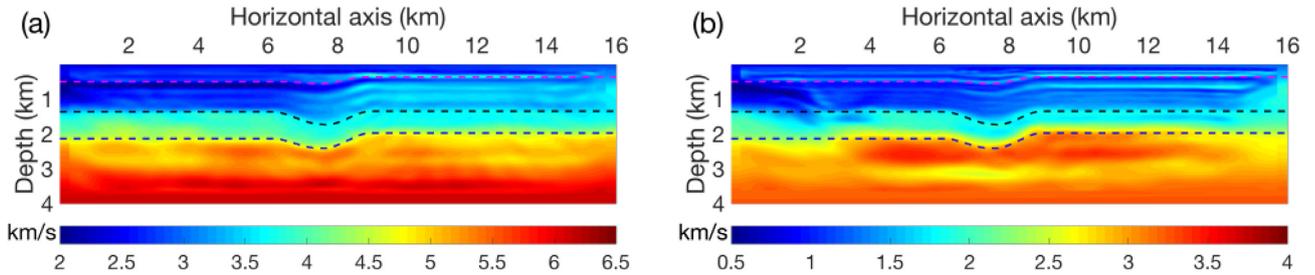


Figure 10. Inverted (a) v_p and (b) v_s with the KR-OT formulation and a Gaussian window with $\sigma = 1.6$ s.

first break in the *observed* shot gather is much more complicated than in the gather modeled in the initial model. The initial model being smooth, no mode conversion occurs in the earth leading to clearly distinguishable transmitted/diving events in the gather. We can identify the first diving P waves, the second diving P waves after a bounce on the free surface, the diving S waves in Fig. 3(a). Mode conversions do occur at the free surface and the ground-roll is the most energetic event. In the *observed* data, it is much more difficult to identify specific events below the first break, except obviously the ground roll. This complex waveform is due to the layering and the large parameter contrasts. Energy conversion occurs. The thicknesses of the layers are of the order of the wavelength in the frequency band we consider. The phases of the different events are tuned (constructive or destructive interference of various waves produce a single event). This is a challenge for inversion (Plessix & Pérez Solano 2015). This situation is also typical in fault and thrust settings, see for instance the SEAM example in Section 4.

2.3 Two-stage least-square inversion results

A possible workflow to invert the data is to apply a two-stage approach. We first focus on the long offset data to update the long

wavelengths and then we include the short-to-intermediate offsets that contain surface waves and reflections and use the full trace length to increase the resolution. In the first stage we focus on the early arrivals by applying a time window that only keeps 1.5 s below the first break (a 0.1 s long cosine taper is used to avoid sharp cut), as highlighted in Fig. 2(b). The velocities retrieved after the first stage of the inversion are displayed in Figs 4(a) and (b). In v_p , the continuous thin layer along the first dashed line is retrieved, and the variations inside each zone (separated by the dashed lines) are revealed. The shear velocity v_s is also partially retrieved. This is possible because in the 1.5 s time window of the observed data, the amplitudes and apparent phases of the P waves are affected by energy conversion even at the first scale (this is also the reason why an acoustic inversion would not perform well in this case). There are also shear events with amplitudes increasing with frequency. Starting with these retrieved velocities, we invert the full data set in a second-stage including surface waves, using the same frequency continuation starting at 1.5 Hz. To illustrate the impact of the ground-roll, we display in the appendix A the gradients with and without accounting for the surface waves (Fig. A1). The final velocities are displayed in Figs 4(c) and (d). Since the ground roll is much more energetic than the body waves, only the shallow part

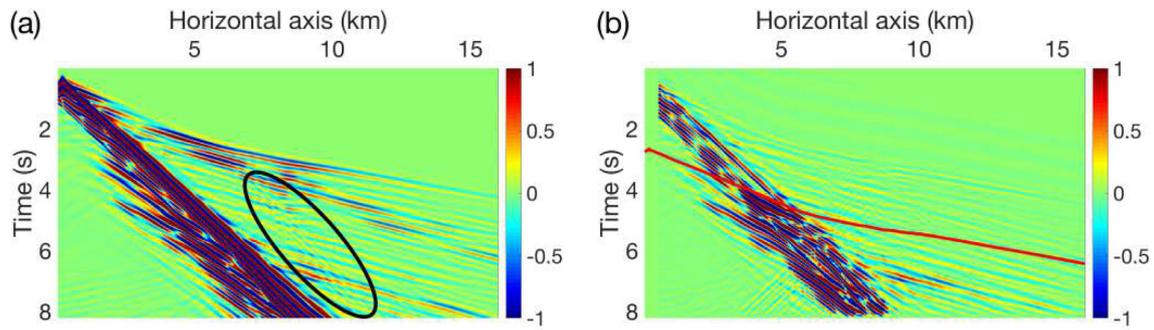


Figure 11. (a) The final synthetic and (b) the data residual (difference between synthetics and observed data) computed with the final model displayed in Fig. 10. The red line is 2.3 s away from the first break.

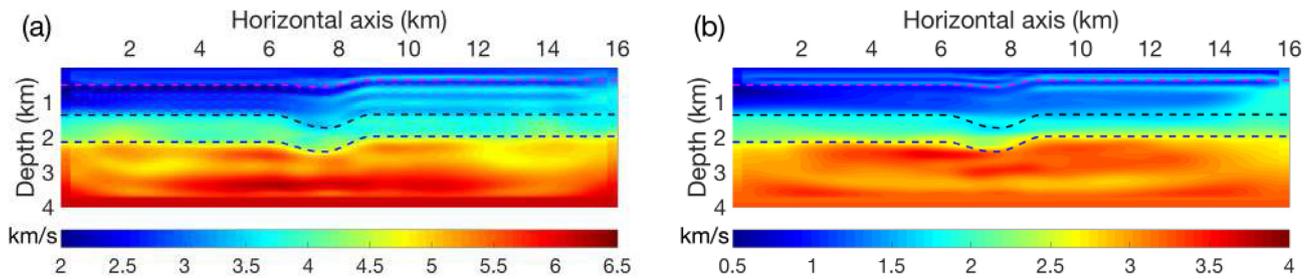


Figure 12. Inverted (a) v_p and (b) v_s with the KR-OT formulation and a Gaussian window with $\sigma = 0.4$ s.

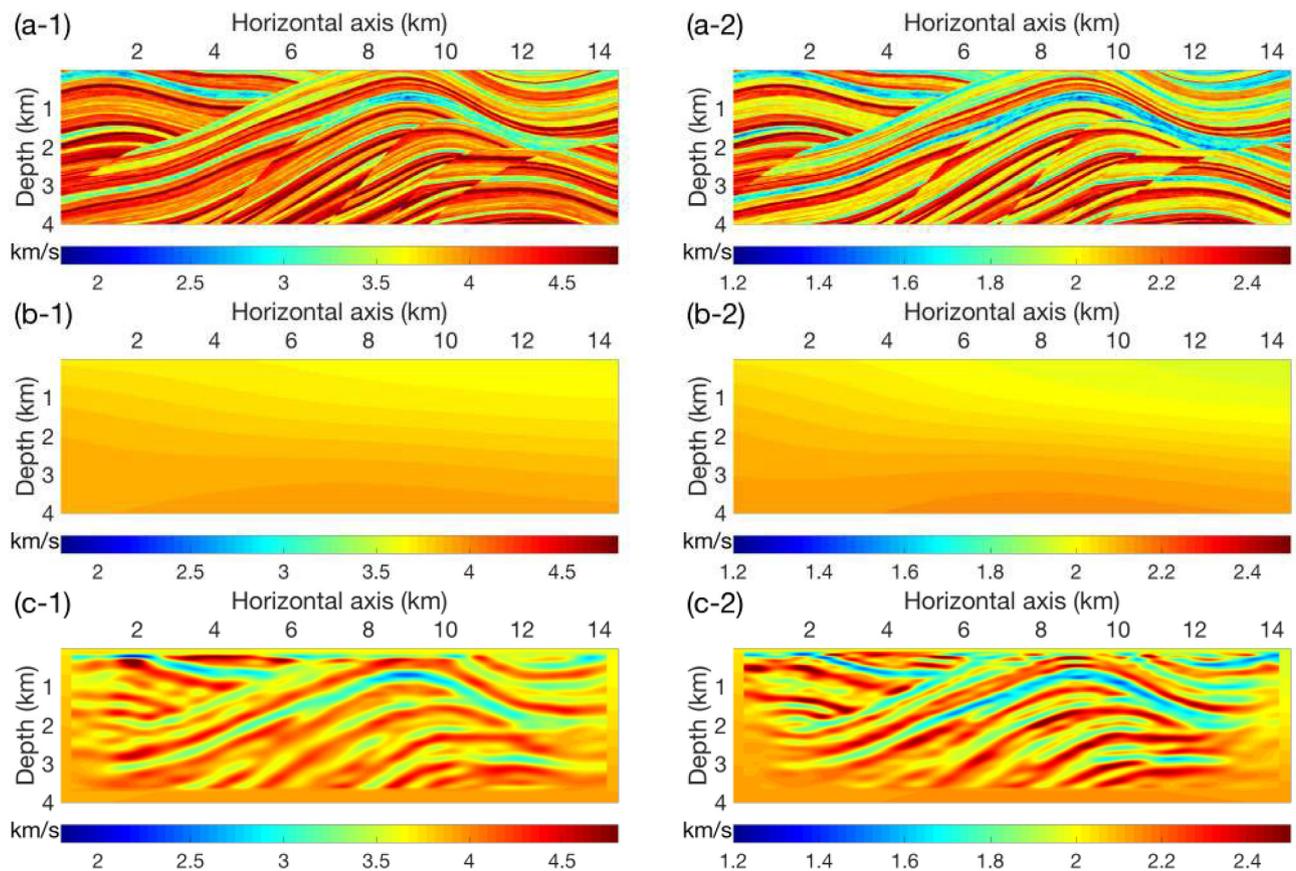


Figure 13. Isotropic inversion results with the windowed KR-OT approach with the SEAM II Foothill model. (a-1) True v_p , (a-2) True v_s , (b-1) Initial v_p , (b-2) Initial v_s , (c-1) Inverted v_p , and (c-2) Inverted v_s .

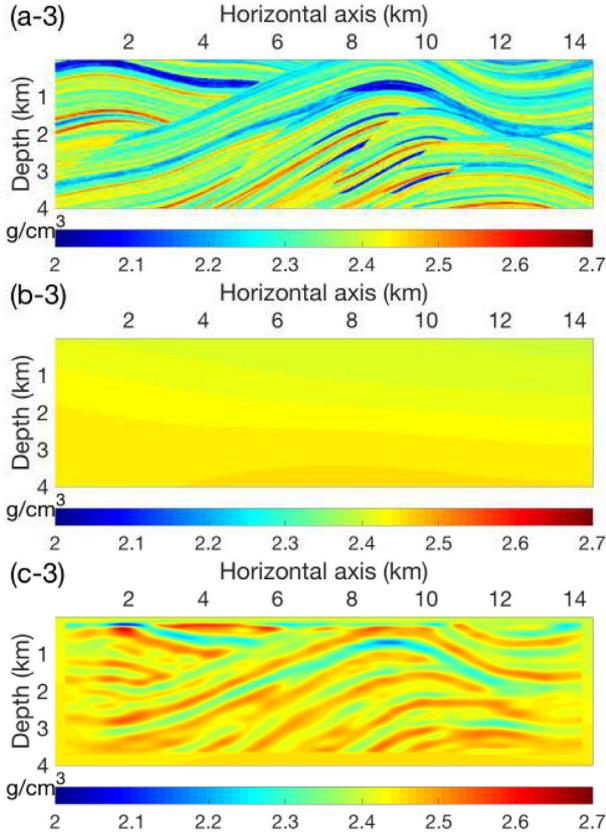


Figure 14. Density models in the isotropic SEAM II Foothill experiment. (a-3) True ρ , (b-3) Initial ρ , and (c-3) Inverted ρ .

of the model is updated, mainly the shear velocity. In Figs 3(b) and (c), we display a shot gather computed on the model obtained after the first inversion stage and after the second one. As expected, after the first inversion stage, the main transmitted waves (the first and second transmitted P waves) are interpreted. After the second inversion stage, the ground roll are better matched. This is illustrated in Fig. 3(d) where we compare a trace located at $X = 7$ km.

However, the second synthetics in Fig. 3 (c) shows that the body waves are not better fit. Instead, anomalous scattered events around $X = 13$ km emerge. This is the drawback of abruptly including surface waves: in the second inversion stage, the variations of body waves do not affect the misfit much since the surface waves are dominant. The inversion is driven by surface waves. Therefore, it is possible that the fitting of body waves becomes worse when surface waves are abruptly included.

2.4 Least-squares inversion with a Gaussian window

Instead of including abruptly the events through a two-step approach, we can use a smooth time window strategy defined through the Gaussian function:

$$W_{s,r}(t) = \exp\left(-\frac{(t - \tau_{s,r}^0)^2}{2\sigma^2}\right), \quad (3)$$

where $\tau_{s,r}^0$ corresponds to the peak of the first arrival wavelet of the trace s , r and σ is a chosen standard deviation. A small standard deviation gives a narrow window around the first break. The Gaussian function damps long-offset easily cycle-skipped surface waves while keeping the short-offset surface waves and the early

body waves. We should carefully choose a suitable time window to balance the contributions from body waves and short-offset surface waves. In the isotropic Middle East example, a window with $\sigma = 1.6$ s includes too many surface waves with the least-squares function. The inversion ends up in a local minimum and no update can be found when we start the second inversion stage. After trials and errors, we found that $\sigma = 0.4$ s gives a good recovery. The corresponding results are presented in Fig. 5. With this Gaussian window, surface waves are constrained to short-offset ranges and body waves to long-offset ranges. Body waves are mainly sensitive to v_p while surface waves are mainly sensitive to v_s . Therefore, v_p and v_s are decoupled which makes the inversion easier and more robust. This is the reason why we can recover them. The deep penetration of diving waves makes possible to recover the deeper structure of the model.

This approach could be challenging in practice for two reasons. First, it is a trial and error process to find a suitable Gaussian window. Secondly, if the macrovelocity increases slowly with depth, then the body waves are not dominated by diving waves, instead could be dominated by reflections. The early body waves then would not penetrate at depth. Thus early body waves could be helpful for reconstructing shallow structures but not for deep area. We may need a second level of iterations by enlarging the Gaussian windows besides the frequency continuation strategy. It may then be difficult to define the best approach for any particular inversion. In addition, the above two considerations are valid only for isotropic FWI. In case of anisotropic FWI, body waves are not only sensitive to velocity parameters but also sensitive to the anisotropy parameters. However, the Gaussian window should be sufficiently small to ensure body waves play a role. In turn, there may not be sufficient events to avoid the crosstalk between parameters, making the anisotropic case much more challenging for such a study.

3 OPTIMAL TRANSPORT APPROACH: KANTOROVICH–RUBINSTEIN STRATEGY

In order to mitigate the difficult selection of the Gaussian window with the different event strengths, we now consider the optimal transport approach proposed in Métivier *et al.* (2016a,b), referred to as Kantorovich–Rubinstein (KR) approach in the following. The objective function based on the least-squares norm do not connect the different samples. The coherency is found in the gradient thanks to the stacking. In the KR approach, the spatial coherency is also looked in the data (shot gather). The corresponding adjoint source is computed as a solution of a constrained convex problem. We explain how these constraints, achieving the smoothness of the adjoint source, help to balance the amplitudes between various events. This behaviour may render the choice of the windowing less crucial.

3.1 Theory

Following Métivier *et al.* (2016b), the OT objective function with KR approach is based on the dual of the 1-Wasserstein distance. For a shot indexed by s , the corresponding objective function h_s between a synthetic shot gather d_s^{syn} and an observed shot gather d_s^{obs} is

$$h_s(d_s^{syn}, d_s^{obs}) = \max_{\varphi_s \in \text{BLip}_1} \sum_r \int_0^T \varphi_s(x_r, t) W_{s,r}(t) \delta d_{s,r}(t) dt \Delta x_r, \quad (4)$$

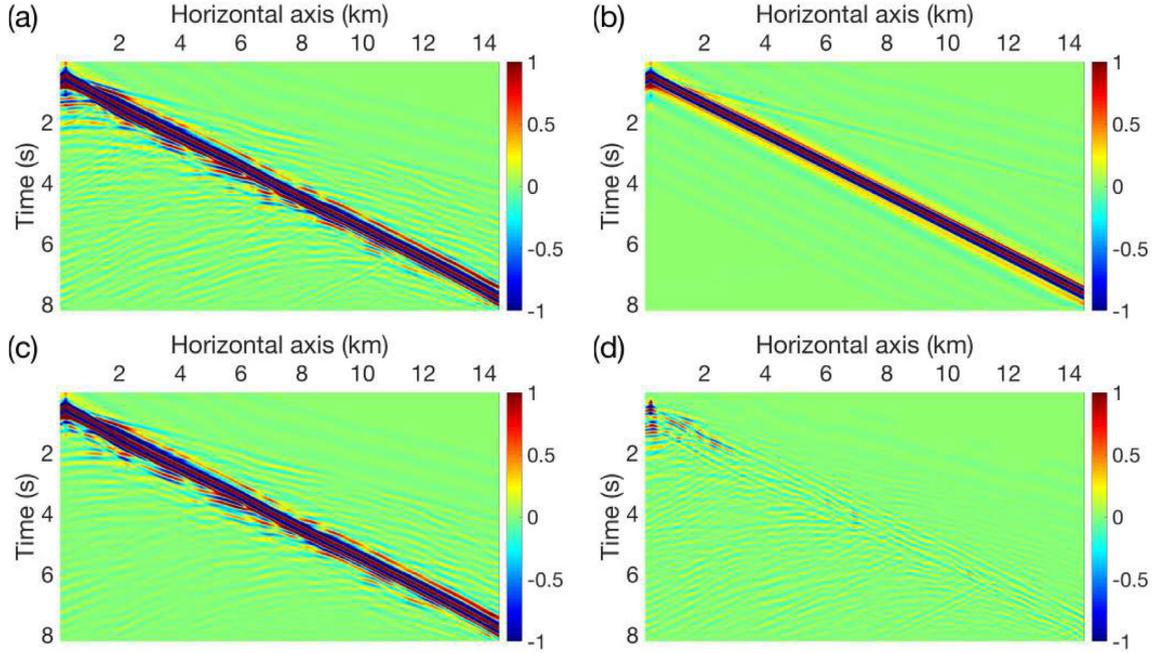


Figure 15. Isotropic SEAM II Foothill example in Figs 13 and 14. (a) Observed data, (b) Initial synthetics, (c) Final synthetics and (d) Final residuals.

where Δx_r is the inter-receiver distance, $W_{s,r}(t)$ is a weighting function introduced to represent the time-windowing technique, and the signed residual $\delta d_{s,r} = d_{s,r}^{syn} - d_{s,r}^{obs}$ is the difference between synthetic and observed shot gather. BLip_1 is the space of Bounded 1-Lipschitz functions defined in the (x_r, t) shot-gather space (Métivier *et al.* 2016b; Poncet *et al.* 2018) by

$$\begin{aligned} \forall(x_r, t) : |\varphi_s(x_r, t)| &< \lambda, \\ \forall(x_r, t) : |\varphi_s(x_r + \delta x_r, t) - \varphi_s(x_r, t)| &< \delta x_r, \\ |\varphi_s(x_r, t + \delta t) - \varphi_s(x_r, t)| &< \delta t, \end{aligned} \quad (5)$$

where λ is a pre-defined constant, δx_r is any positive increment in spatial direction and δt is any positive temporal increment. The first inequality ensures the amplitudes of φ_s is bounded; the last two inequalities prevent abrupt variations.

The maximization problem (eq. 4) can be solved efficiently through proximal splitting techniques (Métivier *et al.* 2016a). Once the optimal $\bar{\varphi}_s(x_r, t)$ is determined, the corresponding adjoint source is simply

$$\mathcal{D}_d h_s = \bar{\varphi}_s(x_r, t) W_{s,r}(t). \quad (6)$$

This is a 2-D algorithm, and we repeat it for each shot panel in our following inversion examples. Thanks to the adjoint state method (Liu & Tromp 2006; Plessix 2006), switching from a least-squares objective function to the KR-OT objective function in an existing FWI code only affects the adjoint source.

3.2 Role of the bound constraints

We use 1-D traces to explain how and why the various events in the KR-OT adjoint source are balanced. The bounded 1-Lipschitz space shown in eq. (5) now contains two conditions:

$$|\varphi| < \lambda, \quad |\mathcal{T}\varphi| < 1, \quad (7)$$

where the differential operator \mathcal{T} is introduced to shorten the expression ($\mathcal{T}\varphi = |\varphi(t + \delta t) - \varphi(t)|/\delta t$). One moving Ricker acts as

a synthetic trace, and one fixed Ricker acts as an observed trace. The moving Ricker test here aims to understand the role of the box constraint bound λ . In principle, we could also study the smoothing constraint bound 1 in eq. (7) ($|\mathcal{T}\varphi| < 1$). However, assuming the smoothing constraint is μ^{-1} ($|\mathcal{T}\varphi| < \mu^{-1}$), we could substitute φ by $\mu\varphi$ and we would come to the same eq. (7). Hence, there is only one independent control parameter and we can analyse the role of λ only.

We use a 3 Hz Ricker wavelet (time increment $\Delta t = 7$ ms, 512 sampling points). The synthetic trace is normalized to 1, and the maximum amplitude of the observed trace, 0.2, is intended to be smaller. Two groups of λ depending on whether λ is bigger or smaller than 1 are selected (Fig. 6). In the first group $\lambda = [1, 1/10, 1/20, 1/50]$, the adjoint source tends to the least-squares adjoint source as λ becomes smaller. The local minima become narrower with decreasing λ . The second group chooses the inverse value of the first group, which gives $\lambda = [1, 10, 20, 50]$. The adjoint sources and the misfit in the second group are similar to each other, indicating that there is a wide range of λ providing the same result.

This phenomena could be explained by the relative weight between the box constraint $|\varphi| < \lambda$ and the smoothing constraint $|\mathcal{T}\varphi| < 1$. The KR-OT function itself involves a maximization problem. The solution φ is iteratively computed via the alternating direction method of multipliers (ADMM) algorithm (Métivier *et al.* 2016a). Starting from $\varphi = 0$, the second φ is proportional to the residual δd . When the box constraint λ is very small, multiplying δd by a small value could easily satisfy the smoothing constraint. Later iterations do not really update φ because bound 1-Lipschitz constraints are already saturated. This explains why the KR-OT adjoint source is similar to the least-squares adjoint source with near-zero λ . On the other hand, once the box constraint is not activated (with large λ values), the smoothing constraint determines the shape of the KR-OT adjoint source. Therefore, the solution of KR-OT really depends on whether the differential operator \mathcal{T} or the identity operator λ^{-1} plays the main role. We conclude that the box constraint λ characterizes the ability of KR-OT to balance the events.

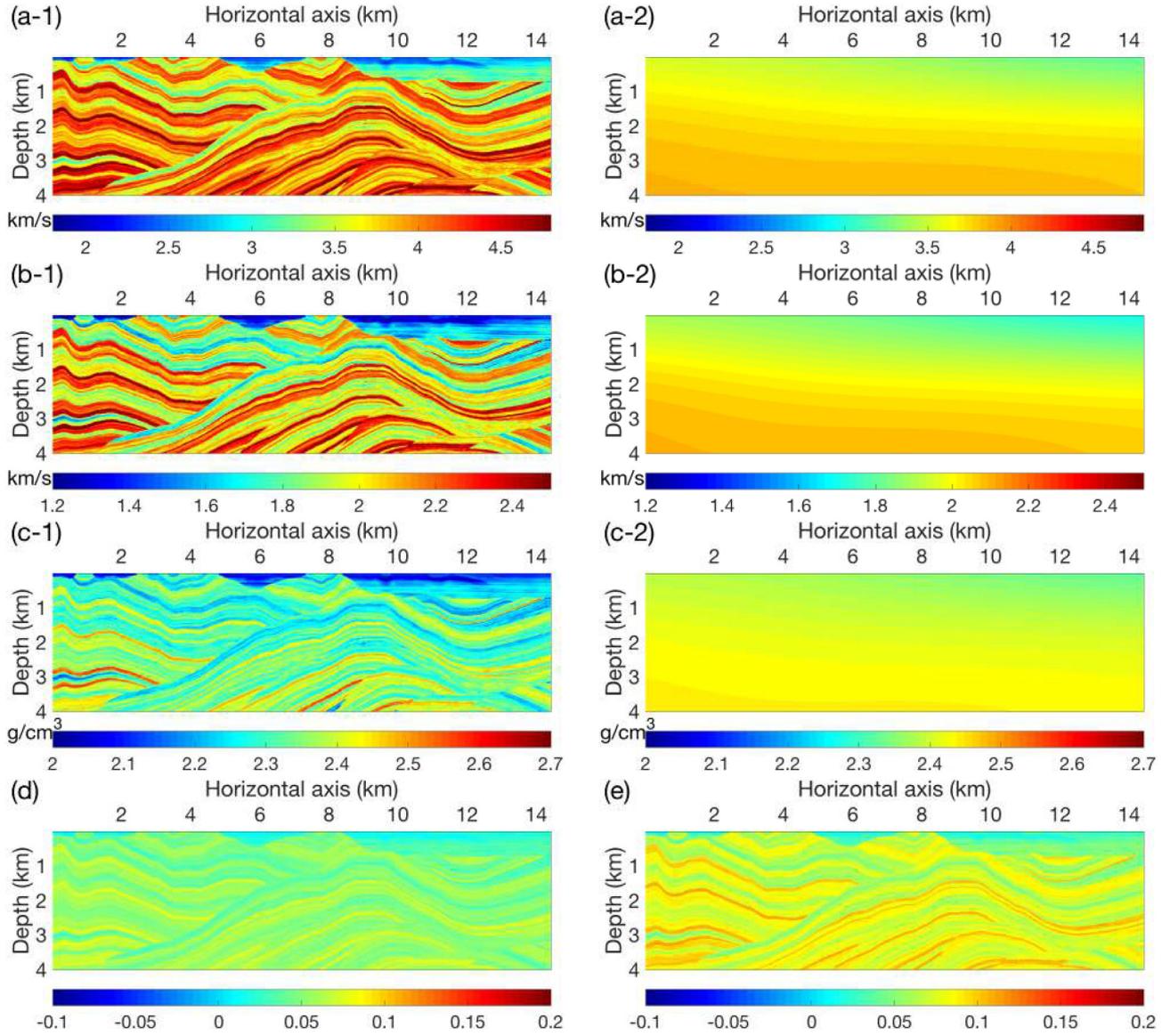


Figure 16. True 2-D VTI SEAM II Foothill model and initial smooth isotropic model. (a-1) True v_n , (a-2) Initial v_n , (b-1) True v_s , (b-2) Initial v_s , (c-1) True ρ , (c-2) Initial $\rho = \rho(v_n)$ is calculated from initial v_n , (d) True η , (e) True δ . Initial η and δ are 0.

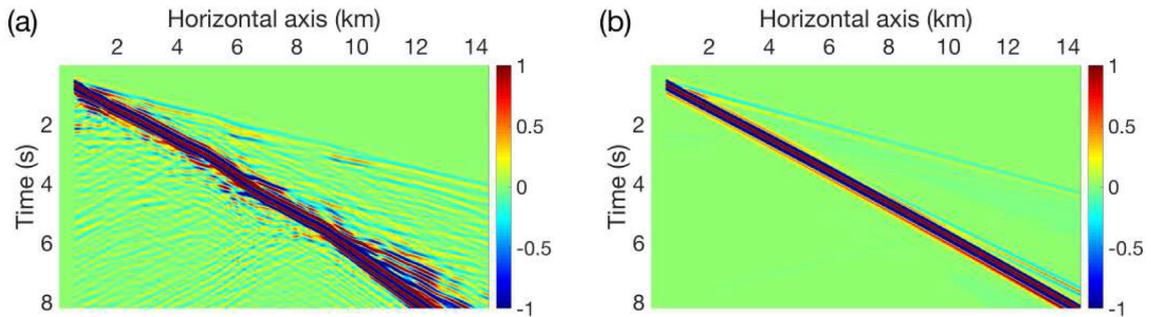


Figure 17. An observed and initial data shot with the anisotropic SEAM II Foothill model shown in Fig. 16. (a) Observed data and (b) Initial synthetics. Cycle-skipping occurs for both body waves and surface waves.

In realistic situations, we could have missing events and misaligned events. To illustrate the KR-OT behaviour in this more complicated situation, we now consider three events with different amplitudes in the observed trace shown in Fig. 7. For the synthetic

trace, the first arrival is assumed to arrive at the same time as the observed first event but with a different amplitude, the second synthetic event exhibits a time shift and has a different amplitude, and the third synthetic event is missing. The residual or adjoint source

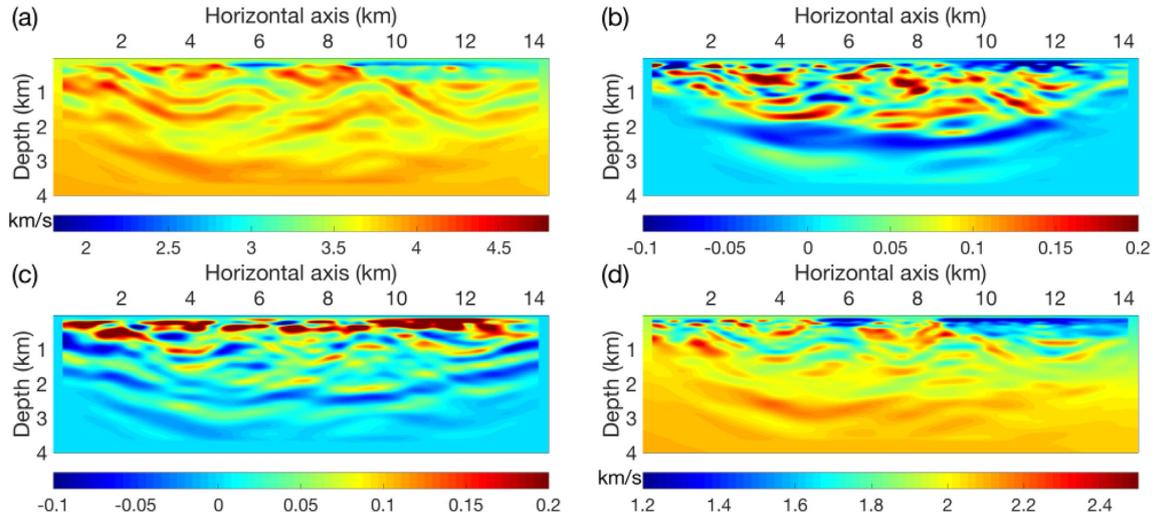


Figure 18. Inversion results of the anisotropic SEAM II Foothill example with a least-squares function and a Gaussian window with $\sigma = 0.4$ s. (a) v_n , (b) η , (c) δ , (d) v_s .

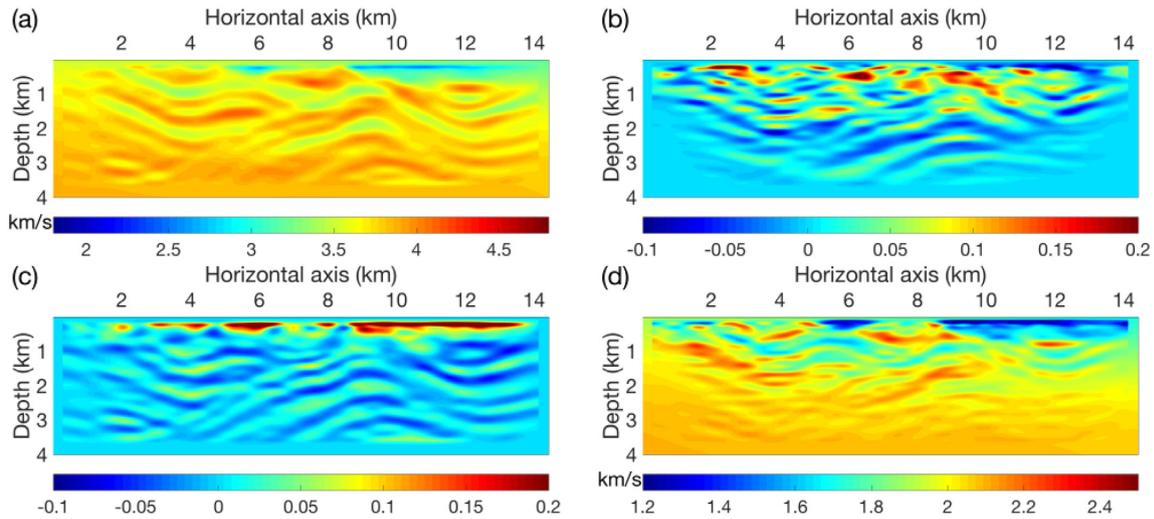


Figure 19. Inversion results of the anisotropic SEAM II Foothill example with a least-squares function and a Gaussian window with $\sigma = 1.6$ s. (a) v_n , (b) η , (c) δ , (d) v_s .

under the least-squares function is dominated by the third event. In this situation, the KR-OT adjoint source shows more balanced events, where the small signals are enhanced and the large amplitudes are attenuated.

3.3 Two-stage inversion approach and full-data inversion at once with KR-OT function

We repeat the two-stage inversion in Section 2.3 (first body waves then entire seismograms) with the KR-OT function. After the body waves are fitted, we get a similar model as with the least-squares function. Interestingly, once the body waves are fitted, the residuals contain negligible body waves. Consequently, there is no body waves to balance with surface waves, in other words, there is no body waves to be enhanced. We then retrieve a model similar to the final model obtained with the least-squares function.

A promising inversion is obtained by applying KR-OT function on the entire seismograms. In this test, we do not use any time window. In the first frequency band, the surface waves dominate

the observed data and the synthetic data. The resulting KR-OT adjoint source is shown in Fig. 8(c). For comparison, the residuals (adjoint source of least-squares function) are also shown. A striking feature is that KR-OT function balances the amplitudes of the surface waves and that of body waves; the amplitudes of all the events are distributed into comparable ranges. Since the surface waves are completely included, as a result, the v_s -gradient is dominant compared to the v_p -gradient. Eventually the v_s parameter is reconstructed (Fig. 9): the lateral variations as well as the layered structure are reconstructed. The inverted v_p does not show layered structures, suggesting that the contribution of the diving waves in the objective function is still weaker than the one of the surface waves.

3.4 KR-OT inversion with Gaussian window

To improve this result, we test the usage of the Gaussian time-window approach to further weaken the surface waves at long offsets. We choose an intermediate Gaussian window with $\sigma = 1.6$ s to enhance the contribution of the body waves. The resulting final

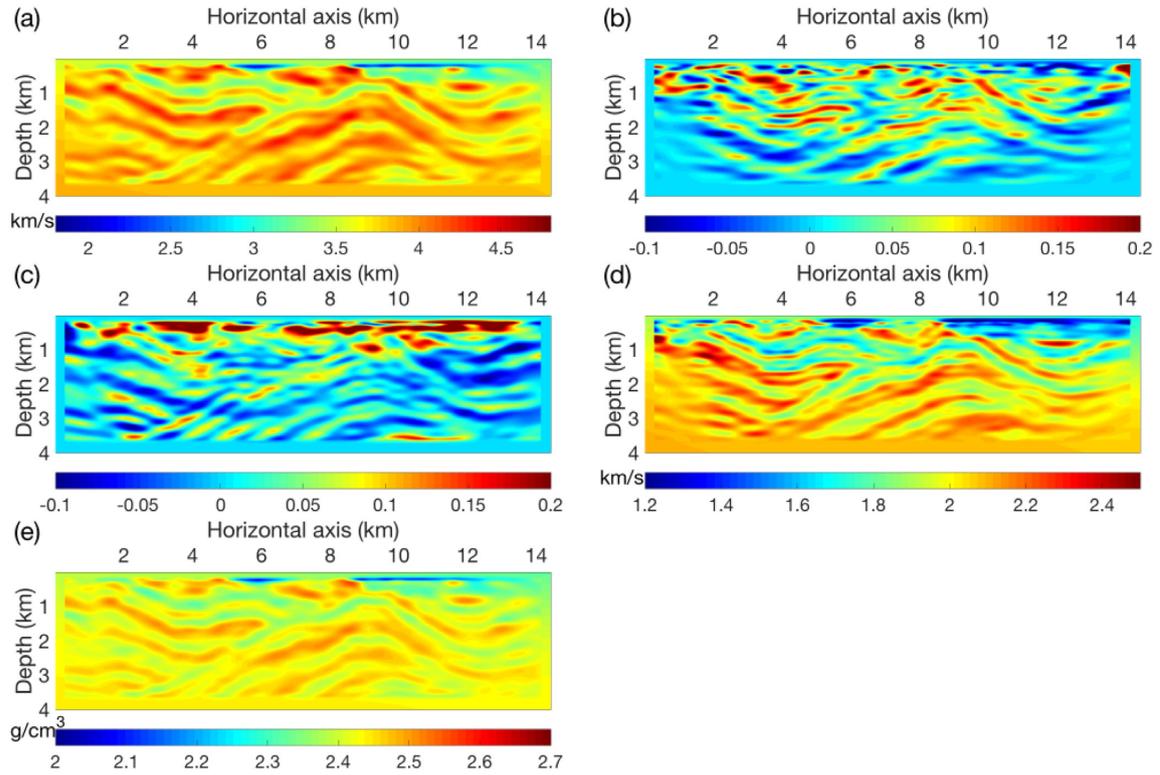


Figure 20. Inversion results of the anisotropic SEAM II Foothill example with the KR-OT function and a Gaussian window with $\sigma = 1.6$ s. (a) v_n , (b) η , (c) δ , (d) v_s , (e) ρ .

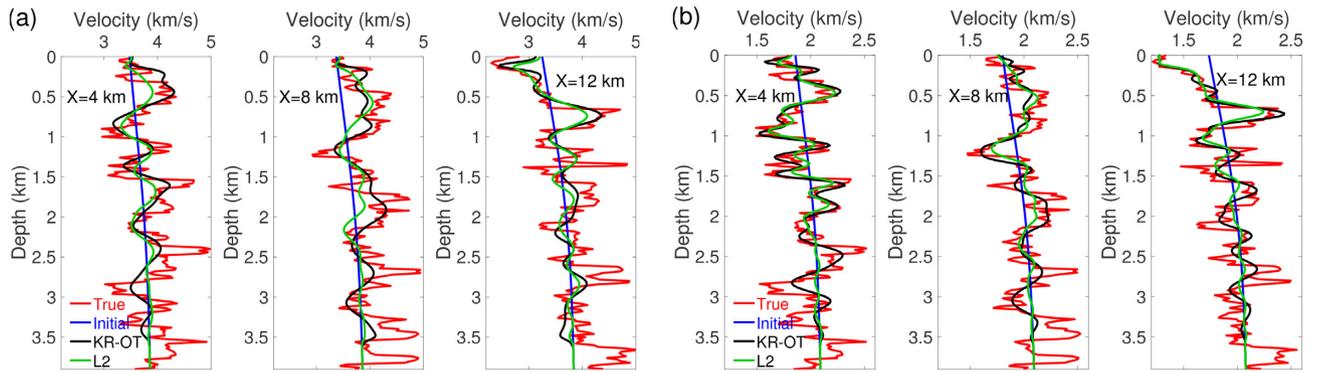


Figure 21. Velocity profiles of the KR-OT inversion in Fig. 20 and L2 inversion in Fig. 19 shown at three locations 4, 8 and 12 km for (a) v_n and (b) v_s .

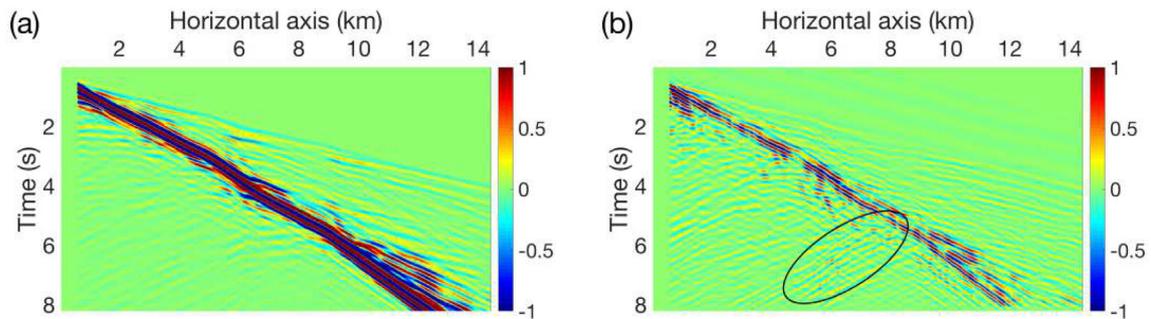


Figure 22. (a) Final synthetics and (b) data residuals with the windowed KR-OT function and the anisotropic SEAM II Foothill model.

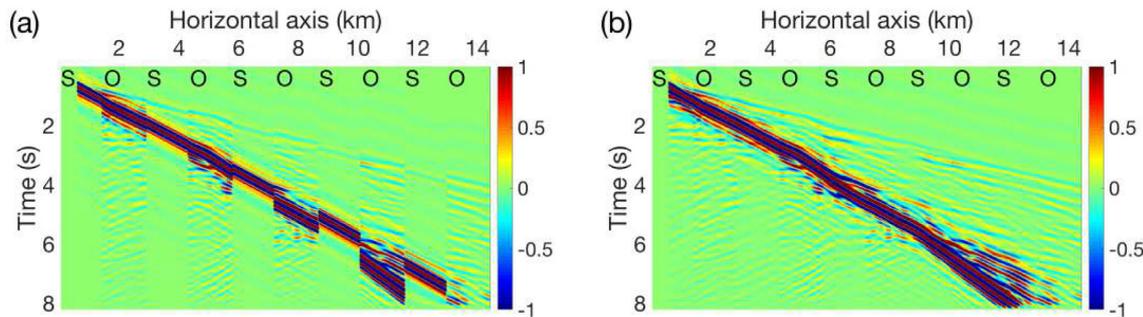


Figure 23. Comparisons between observed data (O) and initial synthetics in plot (a) or final synthetic in plot (b). In the plots, we interleave synthetic and observed data to highlight potential cycle skipping and mismatches.

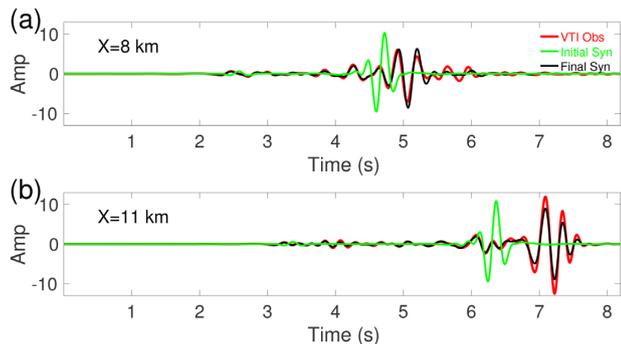


Figure 24. Data trace comparisons (a) at $X=8$ km and (b) at $X=11$ km. Initially cycle-skipped surface waves and body waves are fit after inversion.

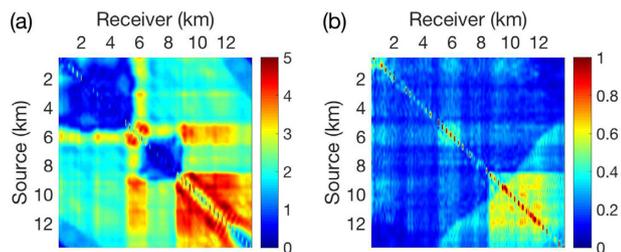


Figure 25. rms distribution (a) in the initial model and (b) in the final model. The high-rms belts at 6 km and [9,14] km correspond to the low-velocity near-surface structures; while the low-rms in the diagonal blocks correspond to the high-velocity structures.

model shown in Fig. 10 is much better than without a window. Both of v_p and v_s show the lateral variations and the thin layered structures. The final synthetic in Fig. 11 matches the observed data. Even the weak scattered body waves are fitted (denoted by a black ellipse in Fig. 11a). Inside the influence zone of the Gaussian window (2.3 s away from the centre, or 1.4σ), the body waves and the surface waves are effectively fitted. We do not design a second level loop over the Gaussian windows to further fit the long-offset surface waves. Whatever the offset range is considered for the surface wave, the same information is carried out and is related to shallow structures.

We also apply the Gaussian window with $\sigma = 0.4$ s. The final model is shown in Fig. 12. This window is a bit narrow for the KR-OT function. The inversion quickly converges in the final stage. The recovery is good. In v_s , we see the decoupled contribution from short-offset surface waves, reflections and long-offset diving waves: the thin structures along the first dashed line are well recovered (short-offset waves); the lateral variations along the third dashed

line are correctly recovered (long-offset waves); but the structures along the second dashed line are not well recovered.

4 ISOTROPIC SEAM II FOOTHILL EXAMPLE

In the Middle East example, the body waves are dominated by diving waves and near-surface interbedding reflections. Also, the geology is mainly flat. We would like to understand the feasibility of our proposed strategy of combining the KR-OT function and a Gaussian window in other geological settings. We extract one slice from the SEAM II Foothill model (elastic isotropic), shown in Fig. 13 (the density models are shown in Fig. 14). The macrovelocity of the SEAM II Foothill model increases much slower along depth than that of the Middle East model. The longest-offset diving waves penetrate to about 1 km deep. The initial v_p and v_s models are strongly smoothed from the true models, while the initial density is estimated from the initial v_p .

The inversion setting is similar to that in the Middle East example. We still consider a 3 Hz Ricker wavelet as source wavelet, and limit the frequencies to [1.5 6.5] Hz. There are 48 vertical point-forces and 300 v_s -receivers. The frequency bands are the same as in the Middle East example. We also apply the same smoothing parameters as for the Middle East example to the gradients in this isotropic SEAM II Foothill example. We choose the intermediate Gaussian window with $\sigma=1.6$ s. We obtain high-resolution high-fidelity structures in Fig. 13. The variations of the structures are reconstructed. These low-velocity channels (blue belts) are recovered both in v_p and v_s . Since density ρ is obtained via empirical relationship from v_p , the structure of the density is similar to that of v_p , but the value of density is overestimated. We present the observed data and synthetics in Fig. 15. The initial synthetics are quite simple. At the end, the entire seismograms are well interpreted including reflections and surface waves in Fig. 15(c). The final residuals in Fig. 15(d) are uniformly low.

5 INVERSION OF A CHALLENGING VTI EXAMPLE

A more realistic representation of the Earth than isotropic elastic is obtained by adding anisotropy parameters. Therefore, a VTI subsurface structure has been created from SEAM II Foothill model. The original varying topography is flattened but the near-surface complexity is kept. Five parameters v_n , v_s , ρ , η and δ are shown in Fig. 16. The initial model is quasi-1-D isotropic (Fig. 16). We show the observed and synthetic seismograms in Fig. 17. The observed

data contain a lot of scatterings. The effects of low-velocity zones on the early body waves could be seen around $X = 6$ km and $X = 9$ km in Fig. 17(a). The surface waves are strongly affected by the near-surface complexity, and they are strongly scattered around $X = 10$ km. Surface waves are not recorded at offsets longer than 12 km. In comparison, the initial synthetics are much simpler. Both body waves and surface waves more or less lie in straight lines in Fig. 17(b), and their amplitudes are more uniform than in observed data. We can recognize that cycle-skipping occurs for both body waves and surface waves (Fig. 23a).

Four parameters (v_n , η , δ , v_s) are simultaneously updated while density is again calculated from v_n at each iteration. First, we present the inversion obtained by applying the least-squares function and a Gaussian time window with $\sigma = 0.4$ s in Fig. 18. For the velocity parameters, we can compare with the true model and recognize the structures above 2 km. For the anisotropy parameters, it is difficult to recognize meaningful structures. We repeat the least-squares inversion with a wider Gaussian time window with $\sigma = 1.6$ s. The final model shown in Fig. 19 is better. The penetration is deeper and we can recognize the structures of the anisotropy parameters.

We apply the KR-OT function and a Gaussian window with $\sigma = 1.6$ s, and get a subsurface structure shown in Fig. 20. The velocity parameters v_n and v_s are well recovered. We can recognize the low-velocity zone, the near-surface complexity, and some interior low-velocity channels. The velocity traces at $X = 4, 8, 12$ km are shown in Fig. 21, along with the inversion with the L2 objective function (Fig. 19). The anisotropy parameters are not recovered with the same quality as for the velocity parameters. But, we can recognize the main structures. The final synthetics shown in Fig. 22(a) capture most of the features of the true seismogram. The final residuals displayed in Fig. 22(b) show the mismatches are uniformly decreased. We do notice that the back-scattered body wave events due to surface waves around $X = 8$ km are not well simulated. The improvement of data fitting before and after inversion could be further seen in Fig. 23. To show a detailed comparison, traces at $X = 8$ and 11 km are extracted and shown in Fig. 24. Even when both body and surface waves are cycle-skipped in the initial model, the inversion converges and the events are well in phase with the observed data in the final model. The rms for the whole shot gathers are shown in Fig. 25. In the initial rms distribution in Fig. 25(a), the rms is high when the sources or receivers are located in the near-surface low velocity zones. After inversion, the rms becomes uniformly low. The highest-rms-blocks in Fig. 25(b) correspond to near-surface low velocity zones.

6 DISCUSSION AND PERSPECTIVE

A two-stage inversion strategy where the surface waves are inverted during the second stage could be insufficient, as we illustrated in Middle East example. Although the transmitted body waves are fitted in the first inversion stage, in the second inversion stage, when the time windowing is removed to include the ground roll, the inversion is immediately driven to fit the energetic surface waves. The use of the Gaussian window allows us to weaken the role of the ground roll in the inversion of the full traces. With a Gaussian time window, we manage to obtain a good recovery with the least-squares function after trial-and-error tests. The Middle East example is successfully inverted with the L2 objective function and the Gaussian windows strategy because: first, the data contain diving waves that penetrate the deep part of the model, therefore structures at depth

are reconstructed; second, a narrow Gaussian time window separates the influence zone of body waves and surface waves, this in some sense is equivalent to decoupling the effects of v_p and v_s . The Gaussian window should be carefully chosen. A too large window could include too many surface waves, and the inversion may have difficulties finding a suitable step-length.

The KR-OT strategy is very attractive for elastic FWI mainly because this objective function is able to balance energy among various events. In the gradient, the contributions from surface waves and body waves are balanced. Therefore, the two-stage inversion strategy is not required with the KR-OT function. With a Gaussian window, the KR-OT function performs quite robustly. It is nevertheless important to consider a smoothing window for three reasons. First, KR-OT algorithm compares patterns. With smoothed windows, the shape of events are preserved. Secondly, damping the long-offset surface waves reduces the risk of cycle-skipping. Third, most importantly, keeping the near-offset events (shear waves and surface waves) helps to balance contributions between body waves and surface waves. The choice of the Gaussian window length becomes less crucial with the KR-OT function. We recommend to use an intermediate Gaussian window in a multiparameter inversion.

The design of the smoothing window is similar to the strategy of Brossier *et al.* (2009). We choose Gaussian functions instead of the exponential function $e^{-|t-t_0|}$ because the Gaussian window is differentiable and more focused around the centre. There could be other choices for the time windows besides the Gaussian function. We suggest to design time windows based on smoothness and compactness criteria. Although in principle we could include a second-level loop over the Gaussian windows besides the frequency continuation strategy, we choose not to do so in this study because of the narrow frequency band [1.5, 6.5] Hz. With much broader frequency band, a second loop with widening Gaussian window may be necessary, and the window length may need to be adapted at each stage. To illustrate the relevance of our inversion strategy, we carried out some inversions with the elastic SEAM II Foothill model, which is a different geological complexity than the Middle East model. While the Middle East model produces abundant diving waves and near surface reflections, the SEAM II Foothill model produces significant reflections and diffractions without deep-penetrating diving waves. The initial model is quasi-1-D. With our KR-OT based strategy, the main features of the velocity model are retrieved.

The L2 norm is known to enhance the contribution of the outliers. We could consider different objective functions. The L1 norm could be used to better balance the contribution of the different events in the trace. We can also think of applying an automatic gain control (agc) on the trace to equal the different contributions. However, those objective functions are less satisfactory than the KR-OT approach because they do not mitigate the issue of cycle-skipping. Moreover, the rebalancing with the L1 norm is just partial and not sufficient with ground roll. The agc has the strong drawback to also significantly enhance the noise.

In multiparameter FWI, time windowing may reduce the wavenumber information content and hence reduce our ability to resolve some parameters or crosstalk. With the KR-OT function, we can use smoother time windows than with the least-squares function in the isotropic examples, making the KR-OT approach attractive in the multiparameter anisotropic case. The anisotropic inversion of an adapted SEAM II Foothill model shows that we could recover v_p and v_s , since the two parameters have a rather isotropic radiation pattern in our chosen parametrization. As expected, the anisotropic parameters are less well recovered because of their limited-angle radiation patterns.

We may raise three questions related to the VTI inversion shown in Fig. 20. One question is whether it is necessary to invert the anisotropy parameters. η is one of the parameters that govern the kinematic of the P-body waves, notably at long offsets. Meanwhile, the role of δ in our inversion is more related to a garbage collector aiming at absorbing non-predicted effects. The long wavelengths of δ cannot be really recovered. Therefore, inverting this parameter allows to stabilize inversion more than aiming at retrieving its quantitative information. Instead of including δ in the inversion, we could have included density. Since v_s is the best recovered parameter in Fig. 20, another question is whether v_s leaked to v_n . In order to assess such leakage, additional tests (not shown here) show that the leakage between these two parameters remains weak, showing that our chosen parametrization is well suited for such setting. In the shallow part, v_s is recovered thanks to the ground roll. In those cases, some shear converted events are present because of the large contrasts helping us to limit the leakage of v_s into the other parameters. We conclude that the parametrization (v_n , η , δ , v_s , ρ) is suitable for those land seismic multiparameter inversions. The choice of parameters may however be case dependent since there are no general optimal parametrization and the leakage between parameters might still occur since the seismic data may not provide sufficient information to recover all the parameters. A third question concerns the inversion strategy. In this study, we have chosen simultaneous inversion (updating all four parameters at once) over sequential inversion. Sequential inversion is well suited when we can decompose the seismic data set into subsets that are mainly governed by a single component. If this hierarchization is not doable, a sequential inversion is prone to leakage and overfitting. In the cases discussed in this manuscript, the large elastic parameter variations make the decomposition/hierarchization of the seismic data almost impossible. This is why we favor simultaneous inversion.

The proposed procedure seems promising and should be considered in more realistic situations including 3-D and topography effects as well as attenuation. For a 3-D acquisition across vast surface area, how to apply the KR-OT approach is still an open question. A straightforward way consists in applying the method to multiple 2-D slices of the 3-D shot gathers, as performed by Poncet *et al.* (2018) and Messud & Sedova (2019). However, if the receiver sampling along the cross-line is sufficiently dense, it might be interesting to apply fully 3-D KR-OT (Métivier *et al.* 2016a). The varying topography would scatter the incident waves into later arriving times, it would be a challenging task to interpret this kind of seismograms. As an extension of our inversion strategy, we show a 3-D inversion experiment in the Appendix B. It is a thin 3-D model with one line of sources and one line of receivers. As expected, the 3-D results of Fig. B1 show that the v_p and v_s parameter are well retrieved with our KR-OT strategy even for complex 3-D topography setting.

ACKNOWLEDGEMENTS

This study was partially funded by the SEISCOPE consortium (<http://seiscope2.osug.fr>), sponsored by AKERBP, CGG, CHEVRON, EQUINOR, EXXONMOBIL, JGI, PETROBRAS, SCHLUMBERGER, SHELL, SINOPEC, SISPROBE and TOTAL. This study was granted access to the HPC resources of SIGAMM infrastructure (<http://crimson.oca.eu>) and CINES/IDRIS/TGCC under the allocation 046091 made by GENCI. Thanks to Satish SINGH and SHELL for providing the first author a PhD position in IPG Paris. Numerical computations were partly performed on the SCAPAD platform, IPGP, France, where the first author established

the inversion strategy. Thanks to Miao YU, Pengliang YANG and Jean VIRIEUX for helpful discussions. We thank the two reviewers Daniel Köhn and Anatoly Baumstein and the editor Herve Chauris for their comments and suggestions that help us to improve this manuscript.

REFERENCES

- Adamczyk, A., Malinowski, M. & Malehmir, A., 2014. High-resolution near-surface velocity model building using full-waveform inversion—a case study from southwest Sweden, *J. geophys. Int.*, **197**(3), 1693–1704.
- Baeten, G., de Maag, J.W., Plessix, R.-E., Klaassen, R., Qureshi, T., Kleemeyer, M., ten Kroode, F. & Rujie, Z., 2013. The use of low frequencies in a full-waveform inversion and impedance inversion land seismic case study, *Geophys. Prospect.*, **61**(4), 701–711.
- Barnes, C. & Charara, M., 2008. Full-waveform inversion results when using acoustic approximation instead of elastic medium, *SEG Tech. Prog. Expanded Abstracts*, **27**(1), 1895–1899.
- Baumstein, A., Ross, W. & Lee, S., 2011. Simultaneous source elastic inversion of surface waves, in *Expanded Abstracts*, p. C040, European Association of Geoscientists and Engineers.
- Bharti, S. *et al.*, 2016. Application of an anisotropic elastic multiparameter waveform inversion on a land data set from north Kuwait, in *2016 SEG International Exposition and Annual Meeting*, Society of Exploration Geophysicists.
- Bleibinhaus, F. & Rondenay, S., 2009. Effects of surface scattering in full-waveform inversion, *Geophysics*, **74**(6), WCC69–WCC77.
- Borisov, D., Modrak, R., Rusmanugroho, H., Yuan, Y., Simons, F., Tromp, J. & Gao, F., 2016. Spectral-element based 3D elastic full-waveform inversion of surface waves in the presence of complex topography using an envelope-based misfit function, in *SEG Technical Program Expanded Abstracts 2016*.
- Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *J. geophys. Int.*, **185**(2), 845–870.
- Brossier, R., Operto, S. & Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, **74**(6), WCC105–WCC118.
- Cance, P. & Capdeville, Y., 2015. Validity of the acoustic approximation for elastic waves in heterogeneous media, *Geophysics*, **80**(4), T161–T173.
- Castellanos, C., Métivier, L., Operto, S., Brossier, R. & Virieux, J., 2015. Fast full waveform inversion with source encoding and second-order optimization methods, *J. geophys. Int.*, **200**(2), 720–744.
- Dokter, E., Köhn, D., Wilken, D., De Nil, D. & Rabbel, W., 2017. Full waveform inversion of sh- and love-wave data in near-surface prospecting, *Geophys. Prospect.*, **65**, 216–236.
- Engquist, B. & Froese, B.D., 2014. Application of the Wasserstein metric to seismic signals, *Commun. Math. Sci.*, **12**(5), 979–988.
- Engquist, B., Froese, B.D. & Yang, Y., 2016. Optimal transport for seismic full waveform inversion, *Commun. Math. Sci.*, **14**(8), 2309–2330.
- Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.P., 2008. Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain, *J. geophys. Int.*, **175**, 665–685.
- He, W. & Plessix, R., 2017. Analysis of different parameterisations of waveform inversion of compressional body waves in an elastic transverse isotropic earth with a vertical axis of symmetry, *Geophys. Prospect.*, **65**, 1004–1024.
- He, W., Plessix, R.-É. & Singh, S., 2018. Parameterization study of the land multi-parameter VTI elastic waveform inversion, *J. geophys. Int.*, **213**(3), 1662–1674.
- Jannane, M. *et al.*, 1989. Wavelengths of Earth structures that can be resolved from seismic reflection data, *Geophysics*, **54**(7), 906–910.
- Köhn, D. *et al.*, 2019. Comparison of time-domain sh waveform inversion strategies based on sequential low and bandpass filtered data for improved resolution in near-surface prospecting, *J. appl. Geophys.*, **160**, 69–83.
- Komatitsch, D., Xie, Z., Bozdağ, E., de Andrade, E.S., Peter, D., Liu, Q. & Tromp, J., 2016. Anelastic sensitivity kernels with parsimonious storage

- for adjoint tomography and full waveform inversion, *J. geophys. Int.*, **206**(3), 1467–1478.
- Krebs, J., Anderson, J., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A. & Lacasse, M.D., 2009. Fast full-wavefield seismic inversion using encoded sources, *Geophysics*, **74**(6), WCC105–WCC116.
- Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering, Theory and application*, pp. 206–220, Society for Industrial and Applied Mathematics.
- Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.*, **96**(6), 2383–2397.
- Marelli, S., Maurer, H. & Manukyan, E., 2012. Validity of the acoustic approximation in full-waveform seismic crosshole tomography, *Geophysics*, **77**(3), R129–R139.
- Masoni, I., Brossier, R., Virieux, J. & Boelle, J., 2013. Alternative misfit functions for FWI applied to surface waves, in *EAGE Technical Program Expanded Abstracts 2013*, pp. Th P10 13, EAGE.
- Messud, J. & Sedova, A., 2019. Multidimensional optimal transport for 3D FWI: demonstration on field data, in *Expanded Abstracts, 81th Annual EAGE Meeting*, London.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. & Virieux, J., 2016a. An optimal transport approach for seismic tomography: application to 3D full waveform inversion, *Inverse Problems*, **32**(11), 115008.
- Métivier, L., Brossier, R., Mérigot, Q., Oudet, E. & Virieux, J., 2016b. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion, *J. geophys. Int.*, **205**, 345–377.
- Métivier, L., Allain, A., Brossier, R., Mérigot, Q., Oudet, E. & Virieux, J., 2018. Optimal transport for mitigating cycle skipping in full waveform inversion: a graph space transform approach, *Geophysics*, **83**(5), R515–R540.
- Mulder, W. & Plessix, R.E., 2008. Exploring some issues in acoustic full waveform inversion, *Geophys. Prospect.*, **56**(6), 827–841.
- Neves, F.A. & Singh, S.C., 1996. Sensitivity study of seismic reflection/refraction data, *J. geophys. Int.*, **126**(2), 470–476.
- Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage, *Math. Comput.*, **35**(151), 773–782.
- Operto, S., Ravaut, C., Improta, L., Virieux, J., Herrero, A. & Dell’Aversana, P., 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltimes and waveform inversions: a case study, *Geophys. Prospect.*, **52**, 625–651.
- Operto, S., Miniussi, A., Brossier, R., Combe, L., Métivier, L., Monteiller, V., Ribodetti, A. & Virieux, J., 2015. Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: application to Valhall in the visco-acoustic vertical transverse isotropic approximation, *J. geophys. Int.*, **202**(2), 1362–1391.
- Oristaglio, M., 2016. Land seismic challenges wraps up with Foothills model, *Leading Edge*, **35**(3), 292–293.
- Pérez Solano, C., Donno, D. & Chauris, H., 2014. Alternative waveform inversion for surface wave analysis in 2-d media, *J. geophys. Int.*, **198**, 1359–1372.
- Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *J. geophys. Int.*, **167**(2), 495–503.
- Plessix, R.E. & Cao, Q., 2011. A parametrization study for surface seismic full waveform inversion in an acoustic vertical transversely isotropic medium, *J. geophys. Int.*, **185**, 539–556.
- Plessix, R.-E. & Pérez Solano, C., 2015. Modified surface boundary conditions for elastic waveform inversion of low-frequency wide-angle active land seismic data, *J. geophys. Int.*, **201**, 1324–1334.
- Plessix, R.E. & Perkins, C., 2010. Full waveform inversion of a deep water ocean bottom seismometer dataset, *First Break*, **28**, 71–78.
- Plessix, R.E., Milcik, P., Rynja, H., Stopin, A., Matson, K. & Abri, S., 2013. Multiparameter full-waveform inversion: marine and land examples, *Leading Edge*, **32**(9), 1030–1038.
- Poncet, R., Messud, J., Bader, M., Lambaré, G., Viguier, G. & Hidalgo, C., 2018. Fwi with optimal transport: a 3D implementation and an application on a field dataset, in *Expanded Abstracts, 80th Annual EAGE Meeting*, Copenhagen.
- Pratt, R.G. & Worthington, M.H., 1990. Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method, *Geophys. Prospect.*, **38**, 287–310.
- Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A. & dell’Aversana, P., 2004. Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt, *J. geophys. Int.*, **159**, 1032–1056.
- Routh, P. *et al.*, 2011. Full-wavefield inversion of marine streamer data with the encoded simultaneous source method, in *Expanded Abstracts, 73th Annual Meeting*, EAGE.
- Sager, K., Ermert, L., Boehm, C. & Fichtner, A., 2017. Towards full waveform ambient noise inversion, *J. geophys. Int.*, **212**(1), 566–590.
- Shen, X., 2010. Near-surface velocity estimation by weighted early-arrival waveform inversion, in *SEG Technical Program Expanded Abstracts 2010*, pp. 1975–1979, Society of Exploration Geophysicists.
- Shin, C. & Cha, Y.H., 2008. Waveform inversion in the Laplace domain, *J. geophys. Int.*, **173**(3), 922–931.
- Sirgue, L., Etgen, J.T. & Albertin, U., 2008. 3D frequency domain waveform inversion using time domain finite difference methods, in *Proceedings 70th EAGE Conference and Exhibition*, p. F022, Roma, Italy.
- Stopin, A., Plessix, R.-E. & Al Abri, S., 2014. Multiparameter waveform inversion of a large wide-azimuth low-frequency land data set in Oman, *Geophysics*, **79**(3), WA69–WA77.
- Tape, C., Liu, Q., Maggi, A. & Tromp, J., 2009. Adjoint tomography of the southern California crust, *Science*, **325**, 988–992.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- Tarantola, A., 1986. A strategy for non linear inversion of seismic reflection data, *Geophysics*, **51**(10), 1893–1903.
- Trinh, P.T., Brossier, R., Métivier, L., Virieux, J. & Wellington, P., 2017. Bessel smoothing filter for spectral element mesh, *J. geophys. Int.*, **209**(3), 1489–1512.
- Trinh, P.T., Brossier, R., Métivier, L., Tvard, L. & Virieux, J., 2018. Data-windowing hierarchy in multi-parameter elastic FWI: 3D synthetic foothills case study, in *Expanded Abstracts, 80th Annual EAGE Meeting*, Copenhagen.
- Trinh, P.T., Brossier, R., Métivier, L., Tvard, L. & Virieux, J., 2019. Efficient 3D time-domain elastic and viscoelastic full waveform inversion using a spectral-element method on flexible Cartesian-based mesh, *Geophysics*, **84**(1), R75–R97.
- Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *J. geophys. Int.*, **160**, 195–216.
- Vigh, D., Moldoveanu, N., Jiao, K., Huang, W. & Kapoor, J., 2013. Ultralong-offset data acquisition can complement full-waveform inversion and lead to improved subsalt imaging, *Leading Edge*, **32**(9), 1116–1122.
- Virieux, J., 1986. P-SV wave propagation in heterogeneous media: velocity-stress finite difference method, *Geophysics*, **51**, 889–901.
- Virieux, J. & Operto, S., 2009. An overview of full waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Warner, M. *et al.*, 2013. Anisotropic 3D full-waveform inversion, *Geophysics*, **78**(2), R59–R80.
- Wittkamp, F., 2016. Individual and Joint 2D elastic full-waveform inversion of Rayleigh and Love waves, *PhD thesis*, Karlsruhe Institute of Technology.
- Yang, Y., Engquist, B., Sun, J. & Hamfeldt, B.F., 2018. Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion, *Geophysics*, **83**(1), R43–R62.
- Yuan, Y.O., Simons, F.J. & Bozdag, E., 2015. Multiscale adjoint waveform tomography for surface and body waves, *Geophysics*, **80**, R281–R302.

APPENDIX A: ILLUSTRATION OF SURFACE WAVE PROBLEM

Inverting surface land seismic data is challenging notably because of the ground rolls. The Rayleigh waves are much more energetic

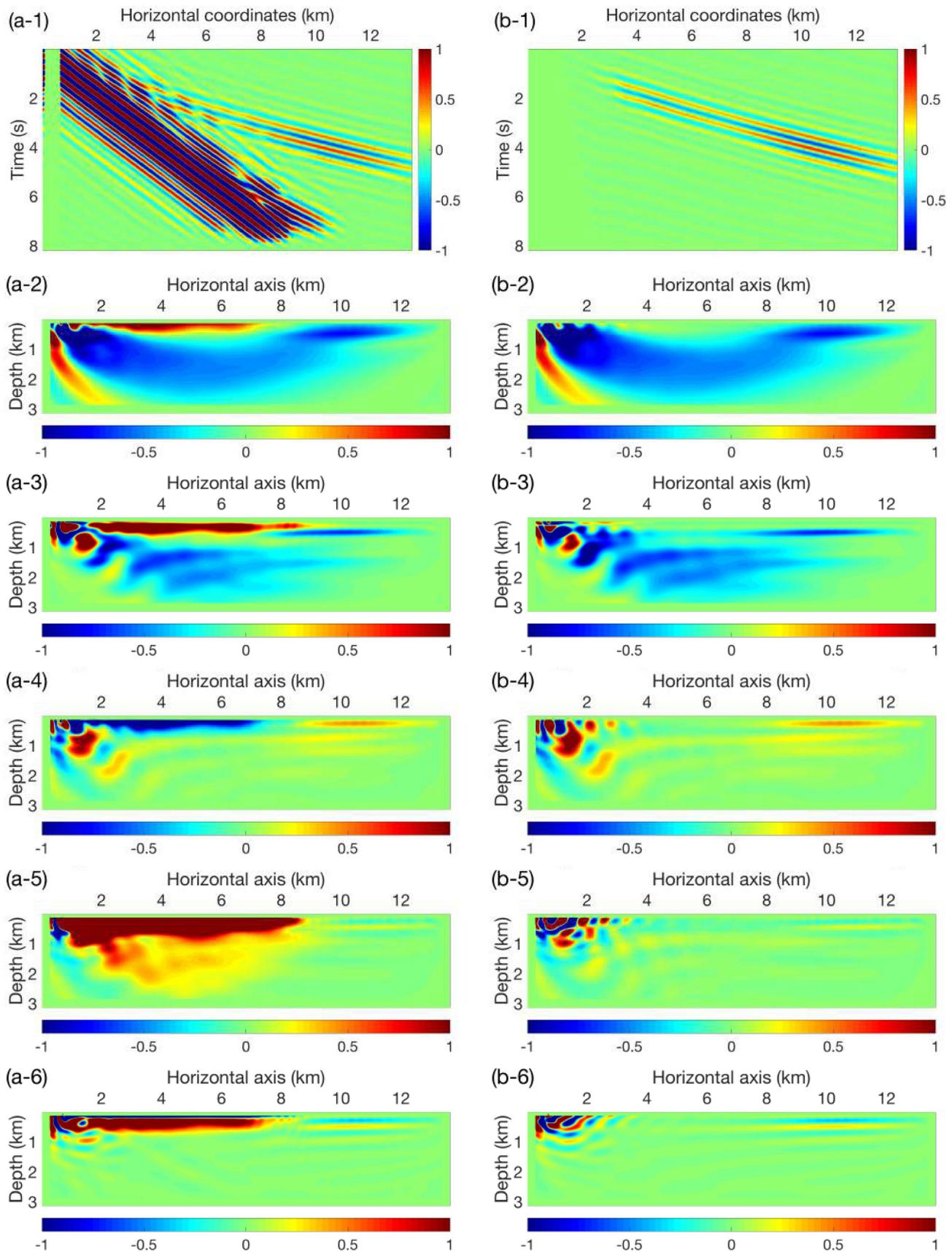


Figure A1. In the left-hand column, the entire seismograms are included in the objective function; in the right-hand column, only the P diving waves are kept. (a-1) Whole data, (b-1) Body waves, (a-2) g_{v_n} with whole data, (b-2) g_{v_n} with body waves, (a-3) g_η with whole data, (b-3) g_η with body waves, (a-4) g_s with whole data, (b-4) g_s with body waves, (a-5) g_{v_s} with whole data, (b-5) g_{v_s} with body waves, (a-6) g_ρ with whole data, (b-6) g_ρ with body waves.

and hence their contribution dominates in the gradients of the least-squares objective function. To illustrate this aspect, we generate a shot gather in a linearly increasing velocity model under the transverse vertical isotropic assumption. The true model is given by: $v_h = 2500 + 0.62z$, $v_n = 2500 + 0.57z$, $v_v = 2300 + 0.48z$, $v_s = 0.5v_v$, and $\rho = 2100$, where v_h , v_n , v_v and v_s are the horizontal (quasi) P -wave velocity, the nmo P -wave velocity, the vertical (quasi) P -wave velocity and the (quasi) S -wave velocity in m s^{-1} , z the depth in m and ρ the density in kg m^{-3} . We consider the unknown variables v_n , $\sqrt{1+2\eta} = v_h/v_n$, $\sqrt{1+2\delta} = v_n/v_v$, v_s and ρ . The shot contains 300 vertical geophones and is generated with a vertical force source. To estimate the gradients, we only change the horizontal and nmo-velocities to $v_h = 2300 + 0.6z$ and $v_n = 2300 + 0.55z$. The gradients obtained when the ground roll is included are plotted in the first column of Fig. A1 and the gradients obtained without including the ground rolls in the second column. As expected the two v_s -gradients are completely different. The ground roll are mostly sensitive to v_s and in a smooth medium, the transmitted P waves are not sensitive to the shear velocity (there are some artifacts near the source). In the four other gradients, we notice that including the ground roll add a strong shallow contribution. It is this strong contribution that prevents the deeper update of v_n when the ground roll are included. The P -wave velocity will be updated only when we would have correctly matched the ground roll to reduce its weight in the data residuals. The v_n -gradient in Fig. A1(b-2) is more regular than the η -gradient in Fig. A1(b-3). This is a consequence of the parametrization choice. As explained in He *et al.* (2018) by analysing the transmission radiation patterns, v_n has a full-angle sensitivity while η has an anisotropic, wide-angle sensitivity being mainly sensitivity to the horizontal propagation and not sensitivity to the vertical propagation. This wide-angle sensitivity together with the wave path trajectories explain the irregularity of the η -gradient. In this parametrization δ has a narrow-angle sensitivity. The transmitted body waves in surface acquisition system are almost insensitive to δ .

APPENDIX B: ISOTROPIC FWI IN THE PRESENCE OF VARYING TOPOGRAPHY

Land environment usually exhibits varying topography, which greatly enhances the non-linearity of FWI (Bleibinhaus &

Rondenay 2009). Surface waves could even be scattered by the varying topography into body waves. Furthermore, in 3-D space, the amplitude distribution between body waves and surface waves in a seismogram is much different from that in 2-D space due to geometrical spreading. In this section, we apply our inversion strategy to a thin-slab 3-D isotropic model.

The true model shown in Figs B1(a) and (b) is extracted from 3-D SEAM Foothill model. Density is constant $\rho = 1.8 \text{ g cm}^{-3}$ and is fixed during the inversion. The initial velocity v_p and v_s shown in Figs B1(c) and (d) is obtained by smoothing the true models. Layered structures are removed in the initial model. To save computation time, we use eight sources (vertical force) at the surface. The receiver distances are 50 m, totally there are 280 receivers recording vertical velocity component. We consider the same frequency band at each inversion stage as we designed in the 2-D inversion tests. As we did in the 2-D inversion tests, we assume known source wavelet, and we use the same mesh to do the inversion (inverse crime).

To honour the varying topography, we rely on our spectral element solver SEM46 (Trinh *et al.* 2019) to generate the observed data as shown in Fig. B2(a). Time interval is $\Delta t = 0.6 \text{ ms}$, and total recording time is around 4.8 s. Surface waves are scattered; and body waves at offsets longer than 7 km are negligible compared to surface waves. The initial synthetics shown in Fig. B2(b) exhibits phase shift compared to observed data (Fig. B2c). During the inversion, we smooth the gradients with Bessel filter (Trinh *et al.* 2017) (the effects after applying twice are similar to Laplacian filter); and the local filter length is adaptive to the velocity values.

The final model is shown in Figs B1(c) and (d). The complexities are revealed in the v_s model, albeit some irregularities. v_p is recovered down to almost 3 km, and the resolution is inferior to that of v_s . The final synthetics fit the observed data very well from short offset to long offset as shown in Fig. B2(d). Based on the data fitting, we believe the irregularities in the final v_s model are due to sparse sources. Since the 3-D elastic FWI based on spectral element method is expensive, source encoding (Krebs *et al.* 2009; Castellanos *et al.* 2015) could be an option to accelerate the process, as demonstrated by Baumstein *et al.* (2011) in a real case study.

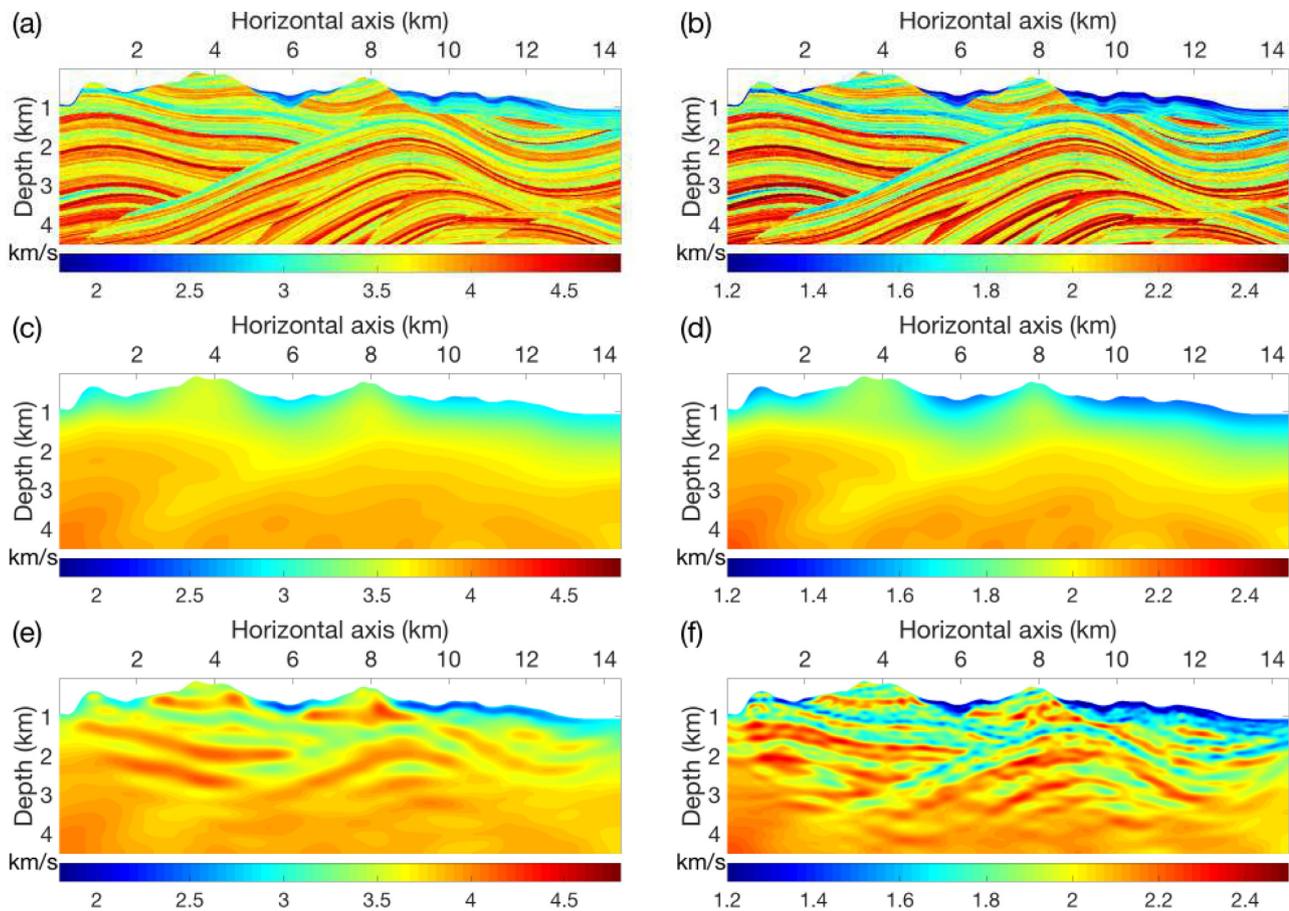


Figure B1. Isotropic elastic inversion experiments in the presence of varying topography. (a) True v_p , (b) true v_s , (c) initial v_p , (d) initial v_s , (e) final v_p and (f) final v_s . Density is constant $\rho = 1.8 \text{ g cm}^{-3}$.

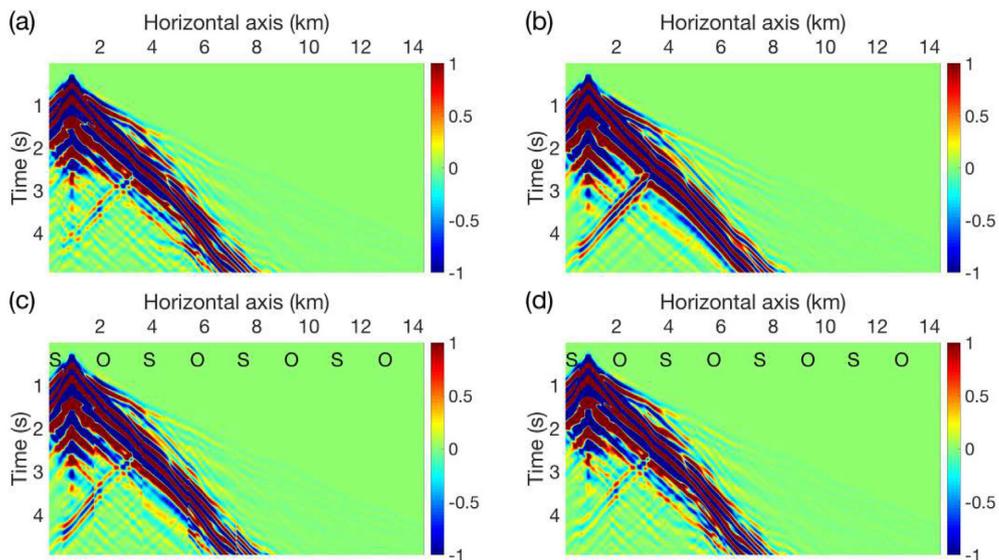


Figure B2. (a) Observed data, (b) initial synthetics, (c) comparison between observed (O) and initial synthetics (S), (d) comparison between observed (O) and final synthetics (S).