

Elastic wave-equation reflection travelttime inversion wave mode decomposition

Tengfei Wang¹, Jiubing Cheng², Qiang Guo³ and Chenlong Wang⁴

¹ *School of Ocean and Earth Science, Tongji University, Shanghai, China. E-mail:*
1110701@tongji.edu.cn

² *State Key Laboratory of Marine Geology, Tongji University, Shanghai, China. E-mail:*
cjb1206@tongji.edu.cn

³ *School of Ocean and Earth Science, Tongji University, Shanghai, China. E-mail:*
1110701@tongji.edu.cn

⁴ *State Key Laboratory of Marine Geology, Tongji University, Shanghai, China. E-mail:*
cjb1206@tongji.edu.cn

(July 6, 2017)

Running head: **Elastic WERTI based on mode decomposition**

ABSTRACT

Elastic reflection waveform inversion (ERWI) utilize the reflections to update the low and intermediate wavenumber in the deeper part of elastic model, which can provide good initial models for elastic full waveform inversion (EFWI). Though ERWI can mitigate the nonlinearity to some extent, it is still stuck with the cycle-skipping problem due to the objective function of waveform fitting. Building initial P and S wave velocity models for EFWI by using elastic wave-equation reflections travelttime inversion (ERTI) would be effective and robust since travelttime information relates to the background model more linearly. The reflection kernels in elastic media are complicated and difficult to use, especially when cal-

culating the gradient of S-wave velocity. The investigation of reflection kernels show that mode decomposition can suppress the artifacts in gradient calculation. Besides, P/S separation of multicomponent seismograms can isolate the P or S recordings for individual data residuals. Thus, we implement the ERTI by using the L_2 norm of the travelttime residual extracted by the Dynamic image warping (DIW) as objective function. A two-step inversion strategy is adopted to effectively reduce the nonlinearity of inversion, in which PP reflections are first used to invert V_p , followed by V_s inversion with PS reflections based on the well recoverd V_p . To avoid the mismeasurement of PS travelttime residual, the well recovered PP image are used to generate the PS reflections. Numerical example of Sigsbee2A model validates the effectiveness of the algorithms and strategies for ERTI.

INTRODUCTION

With the emergence of long-offset wide-azimuth acquisitions and broad-band sources, full waveform inversion (FWI) has been recognized as an efficient tool for constructing velocity models and quantitative seismic imaging (see Virieux and Operto (2009) for a review). Though the acoustic FWI, primarily focusing on P-wave velocity inversion, has been widely studied in the past decades (Tarantola, 1984; Pratt et al., 1998; Shipp and Singh, 2002), people are paying more attention to the waveform inversion under the elastic assumption, referred to as elastic full waveform inversion (EFWI) (Tarantola, 1986). Though waveform inversion provides high-resolution model estimation of the elastic properties, it will suffer from the cycle-skipping easily because of its insensitivity to the low and intermediate wavenumber components of the model when the acquisition illumination is poor and/or good initial models are missing during the inversion(Sears et al., 2008; Brossier et al., 2009). Besides, multi-parameter trade-off effects and more complicated elastic wave phenomena will further increase the difficulties for EFWI. To deal with the nonlinearities and parameter trade-offs, more preconditioning, hierarchical strategies and parameterization investigation should be considered during EFWI (Sears et al., 2008; Prieux et al., 2013; Operto et al., 2013; Wang et al., 2015b; Oh and Alkhalifah, 2016).

In the classical FWI, the long-offset data corresponding to diving waves are very important to build the long-to-intermediate wavelengths of the model. However, the penetration depths of diving waves are always far from sufficient to reach the target in the deeper part even using the wide-aperture surveys. In addition, the low signal-to-noise ratio at the far offset is also a limit for the classical FWI relying on the diving waves. Therefore, people tried to utilize the reflections to help build the macro-model containing low-to-intermediate

wavenumber in the deep part(Chavent et al., 1994; Clement et al., 2001; ?)(more reference). This process can be implemented in the image domain or the data domain. Actually, the image-domain ray-based tomography (Woodward, 1992; Woodward et al., 2008; Jones, 2010) has already been a standard workflow to obtain the background velocity by flattening the common image gathers. But when the lateral velocity variation is strong, the ray-based method would fail to present the wave propagation underground, which will probably lead to the failure of tomography. To overcome the limit of ray theory, the reflection inversion employed waveform or traveltime based on wave-equation draws a lot of consideration(Xu et al., 2012; Ma and Hale, 2013; Wu and Alkhalifah, 2015; Zhou et al., 2015; Wang et al., 2015a; Chi et al., 2015).

The misfit function of reflection inversion can be built in image domain in the manner of wave equation migration velocity analysis (WEMVA), which tries to maximize the energy at zero offset location in the extended image space (Symes, 2008; Almomin and Biondi, 2012; Sun and Symes, 2012; Biondi and Almomin, 2013). Raknes and Weibull (2016) developed the image-domain method to invert P-wave velocity (V_p) in the 3D elastic media. Wang et al. (2017a) proposed to employ the extended PS image in WEMVA to update the S-wave velocity (V_s) with the help of elastic wave mode decomposition. However, the extended-domain methods are limited due to its prohibitively computational cost, especially in 3D case. While in the data domain, Xu et al. (2012) suggested using a reflection waveform inversion (RWI) method to suppress the nonlinearity in FWI, which aim to invert the long-wavelength components of the model by using the reflections predicted by migration/demigration process. RWI highly relies on the accurate reflectivity to generate the reflections that can match the observed data. However, it is very challenging and also expensive to obtain a good reflectivity model through least-square migration when initial

model is far away from the true value. Zhou et al. (2015) proposed a joint FWI method to combine the diving and reflected waves to utilize both RWI and the conventional FWI. Recently, the RWI method is investigated by several works (Wu and Alkhalifah, 2015; Zhou et al., 2015), and recently extended to elastic case by Guo and Alkhalifah (2016).

However, compared with the waveform information, traveltime is more sensitive and linearly related to the low-wavenumber components of the model. Therefore, traveltime inversion will be more robust and helpful to build good initial models for conventional FWI (Wang et al., 2014). Ma and Hale (2013) introduced a wave equation reflected travelttime inversion (WERTI) method based dynamic image warping (DIW) to build the low wavenumber of the model. Chi et al. (2015) and Wang et al. (2015a) employed correlation-based method to extract the temporal or spatial lag to implement the reflection inversion. Elastic reflections carry the background information of the P and S wave velocities, which can help to rebuild the good initial velocity models for EFWI. Unfortunately, in elastic case, traveltimes of a particular wave modes are difficult to extract due to the complicated wave phenomena, such as mode-conversions. Therefore, the estimated time shifts would be inaccurate if using the original multicomponent seismograms. Besides, since the multi-parameter trade-offs will increase the nonlinearity of inversion, more hierarchical strategies should be considered to deal with this problem. Wang and Cheng (2017) explained the mechanism that the wave mode decomposition can mitigate parameter trade-offs in EFWI by preconditioning the gradients. As a natural tool to obtain the separated data subsets, wave mode decomposition similarly has the potential to precondition the elastic wave-equation reflection travelttime inversion (ERTI) with more flexible hierarchical strategies.

In this paper, we will tackle the travelttime misfits of PP and PS reflections to implement the ERTI approach with the aid of wave mode decomposition and DIW (Hale, 2013). First,

the elastic reflection wavepath and its components of different wave modes are calculated and the implications to suppress the artifacts in the gradient calculation are investigated. Then, P/S separation of multicomponent seismograms is applied to the observed and predicted reflection data to extract the individual traveltimes residuals through DIW. Based on the analysis of elastic reflection kernels and the separated traveltimes residuals, we design a two-stage workflow to implement the ERTI method, in which the traveltime of PP is firstly used to recover the background V_p model followed by inverting the background V_s model through the traveltime of PS reflections. During the V_s inversion, the PS reflections are predicted by the PP image to avoid the wrongly estimated time shifts when using PS image. Moreover, we precondition the V_s gradient through spatial wave mode decomposition. Finally, the numerical example of Sigsee2A model proves the robustness and validity of our ERTI method.

THEORY OF ERTI

Assume that there is a perturbation c_{ijkl}^1 in the background elastic media c_{ijkl}^0 , using the Born approximation, the background wavefields u_i and perturbed wavefields \hat{u}_i satisfy:

$$\rho \frac{\partial u_i^2}{\partial t^2} - \frac{\partial}{\partial x_j} \left[c_{ijkl}^0 \frac{\partial u_k}{\partial x_l} \right] = f_i, \quad (1)$$

and

$$\rho \frac{\partial \hat{u}_i^2}{\partial t^2} - \frac{\partial}{\partial x_j} \left[c_{ijkl}^0 \frac{\partial \hat{u}_k}{\partial x_l} \right] = \frac{\partial}{\partial x_j} \left[c_{ijkl}^1 \frac{\partial u_k}{\partial x_l} \right], \quad (2)$$

in the sense of first-order elastic scattering, where \hat{u}_i can be seen as the demigrated reflection data using the image perturbation c_{ijkl}^1 obtained from RTM or other imaging method.

In ERTI, we aim to minimize the traveltime differences between observed data \mathbf{d}^o and

calculated data \mathbf{d}^c , then the objective function can be expressed as:

$$\begin{cases} \tau(\mathbf{x}_r, t) = \arg \min_{\tau} \int_0^T \sum_r \| \mathbf{d}^c(\mathbf{x}_r, t) - \mathbf{d}^o(\mathbf{x}_r, t + \tau) \|^2 \\ E = \frac{1}{2} \int_0^T \sum_r \tau^2(\mathbf{x}_r, t) dt, \end{cases} \quad (3)$$

where the time differences $\tau(\mathbf{x}_r, t)$ can be extracted through DIW (The readers can refer to Hale (2013) for detail). After a similar derivation as in Ma and Hale (2013) (see Appendix A), the gradients of equation (3) can be expressed as:

$$\frac{\partial E}{\partial c_{ijkl}^0} = - \int \left(\frac{\partial u_i}{\partial x_j} \frac{\partial \hat{\psi}_k}{\partial x_l} + \frac{\partial \hat{u}_i}{\partial x_j} \frac{\partial \psi_k}{\partial x_l} \right), \quad (4)$$

where u_i and \hat{u}_i are the state variables, while ψ_i and $\hat{\psi}_i$ are the adjoint state variables satisfying:

$$\rho \frac{\partial \psi_i^2}{\partial t^2} - \frac{\partial}{\partial x_j} \left[c_{ijkl} \frac{\partial \psi_k}{\partial x_l} \right] = \tau(\mathbf{x}_r, t) \frac{\dot{d}_i^o(\mathbf{x}_r, t + \tau)}{h_i(\mathbf{x}_r, t)}, \quad (5)$$

and

$$\rho \frac{\partial \delta \psi_i^2}{\partial t^2} - \frac{\partial}{\partial x_j} \left[c_{ijkl} \frac{\partial \delta \psi_k}{\partial x_l} \right] = \frac{\partial}{\partial x_j} \left[\delta c_{ijkl} \frac{\partial \psi_k}{\partial x_l} \right], \quad (6)$$

with $h_i(\mathbf{x}_r, t) = \dot{d}_i^o(\mathbf{x}_r, t + \tau)^2 - \ddot{d}_i^o(\mathbf{x}_r, t + \tau)(d_i^c(\mathbf{x}_r, t) - d_i^o(\mathbf{x}_r, t + \tau))$. (The hat dot denotes the time derivative.) On the right-hand-side (RHS) of equation (4), the two terms indicate the source and receiver parts of the reflection wavepath, respectively. For simplicity, we derive the gradients of stiffness coefficients for inversion. But the velocity parameterization will be more reasonable to implement the travelttime inversion. Thus, we can get the gradients in terms of P- and S- wave velocities through the chain rule:

$$\begin{aligned} \frac{\partial E}{\partial V_p} &= 2\rho V_p \frac{\partial E}{\partial c_{ijkl}^0} \delta_{ij} \delta_{kl}, \\ \frac{\partial E}{\partial V_s} &= 2\rho V_s \frac{\partial E}{\partial c_{ijkl}^0} (-2\delta_{ij} \delta_{kl} + \delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk}). \end{aligned} \quad (7)$$

ELASTIC BORN REFLECTION KERNELS

Reflection traveltime and waveform inversions utilize different objective functions but share the same reflection wavepath information (reflection kernels). Therefore, the key point of reflection inversion is how to calculate and utilize the reflection kernel. In elastic case, due to the complex wave phenomena, the cross-correlations between different wave mode conversions will generate cross-talks in the kernel. Therefore, the wavepath of elastic reflections will be far more complicated than that in acoustic case. For simplicity, we rewrite (4) as follow:

$$\nabla E(\mathbf{m}_0) = - \int (\mathbf{u} \otimes \hat{\psi} + \hat{\mathbf{u}} \otimes \psi) \quad (8)$$

with \mathbf{m}_0 is the background model, \mathbf{u} and $\hat{\mathbf{u}}$ are the incident and perturbed forward wavefields, while ψ and $\hat{\psi}$ are the incident and perturbed adjoint wavefields, respectively. The operator \otimes denotes the cross correlation between two wavefields. Note that, equation (8) just schematically shows the manner of cross correlation. The detailed formulas should be derived according to the parameterization through chain rule as in equation (7).

Since in the elastic case, the four wavefields in eq. (8) contain both P and S waves, the cross-correlation between different wave modes may cause interference to gradients (Wang et al., 2017b). To analyze the elastic reflection kernels, we decompose the original kernel into four components which correspond to the cross-correlation of different wave modes, respectively, as follows:

$$K_{m_0}^{MN} = - \int (\mathbf{u}^M \otimes \hat{\psi}^N + \hat{\mathbf{u}}^M \otimes \psi^N), \quad (9)$$

where $M, N \in \{P, S\}$. $K_{m_0}^{MN}$ represents the cross correlation between the M mode forward wavefields and the N mode adjoint wavefields. As illustrated in Fig. 1, this kernel can be considered as the cross-correlation between the forward wavefields emitted from a

virtual source at a mirror location and the adjoint wavefields back-propagated from the receiver. If the traveltime of the adjoint wavefields (injected at receiver) equals to the forward wavefields, the cross-correlation will produce the low-frequency wavepath. Otherwise, the cross-correlation should be a high-wavenumber migration impulse. For example, K^{PS} is the cross-correlation of P-mode forward wavefield and S-mode adjoint wavefields, but has little to do with the PS mode data.

Numerically, we calculate the kernels with the single-source-receiver data which are synthesized for a single reflector with a pure P-wave source. Since different objective functions only induce different types of adjoint sources from the perspective of gradient calculation, the investigation of kernels can be done by using a waveform data residual as the adjoint source for back-propagation. We use the $V_p - V_s$ parameterization as in equation (7). In the first test, we place a single V_p reflector in the homogeneous background (Fig 2a and b). Since there is no perturbation of V_s , only PP reflection exist in the data (almost the same as in acoustic media), which means mode decomposition is unnecessary in this case. As shown in Figure 2c and d, the reflection kernel consists of two “rabbit-ear”, the source and receiver parts. K_{V_p} produces the low-wavenumber PP reflection wavepath as expected. While in K_{V_s} , the energy focus on the edge of the wavepath rather than the first Fresnel-Zone as in the V_p kernel. One plausible reason is that the V_s kernel is relatively insensitive to the PP data generated by V_p reflector. Besides, we can see the migration impulse below the reflector due to the down-going perturbed wavefields (Zone D mentioned by Zhou et al. (2015)).

In the second test, we use the V_s reflector (Fig 3a and b) to generate both PP and PS reflections. The V_p kernel excludes S-wavefield automatically because of the divergence operator implied in the term $\delta_{ij}\delta_{kl}$ (see eq. (7)). However, $\hat{\psi}$ contains the converted SP

wavefields induced by the back-propagated ψ^S at the location of reflector. These non-physical wavefields make the V_p kernel slightly different from that in Figure 2c. To remove these non-physical artifacts in K_{V_p} , we should only inject the PP data during the back-propagation.

For the V_s kernel (Figure 3d), due to mode conversions, multi-wavepaths overlapping with each other make it much more difficult to find the correct reflection kernel. The straightforward utilization of this kernel in gradient calculation will very likely cause cross-talk during reflection inversion. According to equation (9), we calculate the components of K_{V_s} , as shown in Figure 4. We observe that the $K_{V_s}^{PP}$ is similar to $K_{V_p}^{PP}$ but with an opposite sign. This is because the term $\delta_{ij}\delta_{kl}$ implied divergence operator has a negative sign in the V_s gradient (eq. (7)). $K_{V_s}^{PS}$ and $K_{V_s}^{SP}$ mainly consist of high-wavenumber energy corresponding to the migration impulse of cross-mode wavefields. Figure 4d shows that the separated S-wavepath of PS reflection is little affected by the P-wavefields.

To further investigate the implication of the reflection kernel, we will analyze the generation of every parts by not only decomposing the spatial wavefields but also isolating the injection of PP or PS data. Taking the second test as example, we further decompose the kernels into several parts by using the manner shown in the schematic illustrations, respectively. The decomposition of K_{V_p} is easy to understand (Fig 5) because only K^{PP} is involved. In Fig 5e, only the PP data are injected as adjoint source, then the cross-correlation between perturbed wavefields $\hat{\psi}^{PP}$ and the background \mathbf{u}^P produces a clean PP wavepath at the source side. The schematic illustration with raypaths denotes the wave mode of the modelled and adjoint wavefields involved in calculation. Note that, when PS data are injected, an SP conversion will occur at the interface, which will produce a migration impulse, as shown in Fig. 5f. Moreover, in Fig. 5h, since a non-physical P wavefield is

generated in ψ due to the PS data injection (?), another migration impulse can be seen at the receiver side. Accordingly, the non-physical conversions brought about by the PS data injection will cause artifacts in the K_{V_p} calculation.

Since K_{V_s} is more complicated, we first decompose it into source and receiver parts as shown in Fig 6c and d. Then, the isolation of PP and PS data injection provides four different components (Fig 6e to h). Fig 6e and f show the decomposition of 6c while Fig 6g and h corresponds to the decomposition of 6d. These four components can be further decomposed with the help of mode decomposition. For short, we will only analyze the decomposition of Fig 6h, while the analyses of the rest three parts will be placed in Appendix B. In Fig 7, when PS data is injected P and S wavefields are generated simultaneously in ψ . Although 7c denotes a same-mode correlation, it generates a migration impulse due to the inconsistent traveltime of modelled $\hat{\mathbf{u}}^{PP}$ and non-physical converted ψ^P as if a PP event is injected at the location of PS event. The cross-mode correlations in Fig 7d and 7e also produce migration impulses. Luckily, the cross-correlation between ψ^S and $\hat{\mathbf{u}}^{PS}$ provides the low-wavenumber S-wavepath of the PS reflection.

According to the above investigation, we can find that the cross-correlation between different wave modes cause cross-talks in gradient calculation. The isolation of PP data when injecting the adjoint source helps to reduce the artifacts in V_p gradients, so does the isolation of PS data for V_s gradients. Most importantly, we expect to update the V_s model through the S-wavepath in PS reflection, which is exactly the component $K_{V_s}^{SS}$. Therefore, we recommend using $K_{V_s}^{SS}$ to mitigate the cross-talks in V_s gradient calculation.

WORKFLOW OF ELASTIC WERTI

In elastic case, it is common to observe that different mode-conversions, mainly PP and PS events, overlap and intersect with each other in the original multi-component seismograms . The cross points between events would be singularities for traveltime difference estimation through DIW. Therefore, the estimated $\tau(\mathbf{x}_r, t; \mathbf{x}_s)$ would be inaccurate due to these singularities. Besides, according to the reflection kernel analysis, individually injecting separated PP or PS seismograms can mitigate the cross-talk in gradient calculation. Thus, we decompose the observed and synthesized data into P- and S-wave parts through P/S separation of multi-component seismograms (Li et al., 2016). In this way, the traveltime differences can be decoupled into PP and PS parts by DIW. Accordingly, we implement the ERTI through a two-stage workflow, i.e. estimating V_p using PP reflections followed by the V_s estimation using PS reflections.

Stage I: PP reflection for V_p

In this stage, we use traveltimes of PP reflection to recover the low-to-intermediate wavenumbers of V_p model. The perturbation of V_p (δV_p) is obtained through elastic reverse time migration (ERTM). When the migration velocity model is inaccurate, people always assume the two way traveltime of zero-offset as an invariants during reflection inversion, such as reflection travelttime tomography or RWI, and thus use the zero or small offset data to generate the velocity perturbation (Zhou et al., 2015). Nonetheless, this method still may have problem to ensure a correct zero offset travelttime in the demigration when velocity anomaly is complex. A better method would be the demigration with extended image (??), but it is out of the scope of this paper. Besides, since the travelttime rather than the amplitude

is fitted in ERTI, ERTM (instead of its least-squares counterpart) is sufficient to obtain the image perturbation for the kinematic demigration. Thus, the objective function can be written as:

$$\begin{cases} \tau_{pp}(\mathbf{x}_r, t; \mathbf{x}_s) = \arg \min_{\tau} \| \mathbf{d}_{pp}^c(\mathbf{x}_r, t; \mathbf{x}_s) - \mathbf{d}_{pp}^o(\mathbf{x}_r, t + \tau; \mathbf{x}_s) \|^2 \\ E_{pp} = \frac{1}{2} \int \tau_{pp}^2(\mathbf{x}_r, t; \mathbf{x}_s) dt d\mathbf{x}_r d\mathbf{x}_s, \end{cases} \quad (10)$$

where d_{pp}^c and d_{pp}^o are synthesized and observed PP reflection through the P/S separation on the recording surface, respectively. According to the previous derivation, we can obtain the gradient of V_p , i.e. $\frac{\partial E}{\partial V_p}$. Note that, the divergence operator is implied in $\frac{\partial E}{\partial V_p}$, which makes the gradient calculation only involve PP reflection. Therefore, the PP reflection traveltime inversion for V_p in elastic media is similar to the acoustic case except that we need the isolation of PS component in the former one. In addition, the separated P-wave reflections may include some SP conversions due to the source-side effects in the elastic case. This may lead to the mismatch of the traveltime inversion.

Stage II: PS reflection for V_s

In the second stage, we utilize the PS reflections to retrieve the background component of V_s model. Similarly, the objective function becomes:

$$\begin{cases} \tau_{ps}(\mathbf{x}_r, t; \mathbf{x}_s) = \arg \min_{\tau} \| \mathbf{d}_{ps}^c(\mathbf{x}_r, t; \mathbf{x}_s) - \mathbf{d}_{ps}^o(\mathbf{x}_r, t + \tau; \mathbf{x}_s) \|^2 \\ E_{ps} = \frac{1}{2} \int \tau_{ps}^2(\mathbf{x}_r, t; \mathbf{x}_s) dt d\mathbf{x}_r d\mathbf{x}_s. \end{cases} \quad (11)$$

where d_{ps}^c and d_{ps}^o are synthesized and observed PS reflections, respectively. After the P-wave stage inversion, both the background model and the structural image perturbation of V_p have been well recovered. These are the good constrains for the prediction of PS reflection. Moreover, in most geological settings, V_p and V_s share the same structure in the

subsurface. It means that we can use the well-located δV_p instead of δV_s to generate the PS reflections. Therefore, to fit the traveltimes, there are two ways to predict the PS reflection after the P-wave stage:

- I:** Migrate the observed PS data with inverted V_p model and initial V_s model to get the high wavenumber of V_s (δV_s), then predict the PS reflection;
- II:** Start from the inverted V_p model and initial V_s model, use the well-located δV_p to generate the PS reflections.

Unfortunately, the predicted PS reflection with method **I** are very sensitive to the error of V_p model. This sensitivity makes the inversion more nonlinear. Under the high-frequency approximation, we will explain this situation using a single reflector example through map migration and demigration.

Figure 8 shows the asymmetric ray path of PS reflection. According to the Snell's law and geometrical relationship, we have:

$$\frac{\sin\theta_1}{V_p} = \frac{\sin\theta_2}{V_s}, \quad (12)$$

$$z(\tan\theta_1 + \tan\theta_2) = X,$$

with V_p and V_s are the correct velocity, X is the offset, z is the depth, and θ_1 and θ_2 are the incident and reflected angles, respectively. Therefore, the moveout of PS reflection is:

$$t = \frac{z}{\cos\theta_1 V_p} + \frac{z}{\cos\theta_2 V_s}. \quad (13)$$

The difference between method **I** and **II** is the migrated depth when velocity model is inaccurate. To show the difference, zero (near) offset map migration is implemented to obtain the migrated depth when velocity is wrong. Then, we use this depth and the wrong velocity through equation (13) to predict the travelttime of PS reflection.

In method I, since the zero-offset two-way traveltime keeps unchanged during map migration, the migrated depth (z_1) satisfies:

$$\frac{z_1}{V'_p} + \frac{z_1}{V'_s} = \frac{z}{V_p} + \frac{z}{V_s}, \quad (14)$$

where V'_p and V'_s are the wrong velocity. Easily, we can get:

$$z_1 = \frac{V'_s(V_p + V_s)}{V_s(V'_p + V'_s)} z. \quad (15)$$

Generally, $V_p > V_s$ and $V'_p > V'_s$. Therefore, a small perturbation of V'_p will lead to a relatively big change of $\frac{V_p + V_s}{V'_p + V'_s}$, which means that the migration depth z_1 is very sensitive to both V'_p and V'_s . While in method II, the depth after migration is:

$$z_2 = \frac{V'_p}{V_p} z. \quad (16)$$

Then, the demigration process can be implemented with:

$$t' = \frac{z'}{\cos\theta_1 V'_p} + \frac{z'}{\cos\theta_2 V'_s}. \quad (17)$$

where $z' \in \{z_1, z_2\}$ is the migrated depth and t' is the corresponding two-way travelttime.

To test the sensitivity of demigrated PS travelttime to V'_p and h , we compare the predicted travelttime curves of the wrong velocity with the observed ones. Here, $V_p = 2500m/s$, $V_s = 1500m/s$, $V'_s = 1300m/s$ are fixed and the value of V'_p are perturbed ($V'_p = 2450, 2500, 2550m/s$) to calculate the curves. As shown in Figure 9a, the predicted PS travelttime with method I (black) oscillates around the true value (red). Theoretically, the predicted curve should behave like the second black line ($V'_p = V_p$), where the zero-offset travelttime is correct while the long offset travelttime is larger than the true one. Especially, when $V'_p = 2450m/s$, the predicted curve is inconsistent with the S-wave velocity error ($V'_s < V_s$). While when $V'_p = 2550m/s$ the predicted curve is below the true value. This

means that the sign of the traveltime residual will be very sensitive to the error of V_p model, even if the error of V_p is as small as $\pm 50m/s$. If the depth of reflector decreases, the sensitivity of method I reduces (figure 9b). But the inconsistency between residual and velocity error still exists, especially in the small offset. This kind of sign change leads to the abnormal change of the gradient's direction during the inversion. Therefore, the nonlinearity of inversion increases obviously due to this sensitivity. Fortunately, if we use method II, the predicted curves (blue) are all below the true value, whose traveltime residual are consistent with the S-wave velocity error. According to the above analysis, we recommend method II to implement the PS stage inversion.

Nonetheless, the amplitude of PS will be inaccurate since we use the δV_p to generate PS reflection, even if δV_p is obtained by the ELSRTM workflows. Luckily, this has little effect on ERTI, because it does not rely on the amplitude information. As shown in figure 9, the traveltime residuals are large for the PS reflection in the deeper part, DIW may suffer from the cycle-skipping problem in this situation. We can utilize the “layer-stripping” strategy to tackle this, during which the shallow part is inverted using the early arrived reflection followed by the inversion of deep part using the late arrived reflections gradually. In this way, the incorrect traveltime residuals of the deeper part will not mislead the inversion of shallow part. Vice versa, the reliable inversion of the shallow part guarantee the correctness of traveltime residual for the deeper part.

In order to make sure that reflected S-wavepath is used to update V_s , wave mode decomposition is applied to calculate $K_{V_s}^{SS}$, namely:

$$\frac{\partial E_{ps}}{\partial V_s} = -2\rho V_s \int \left(\frac{\partial \hat{u}_i^S}{\partial x_j} \frac{\partial \psi_k^S}{\partial x_l} \right) (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \quad (18)$$

This is similar to the gradient preconditioning for EFWI proposed by Wang and Cheng

(2017). The mode decomposition based precondition can mitigate parameter trade-offs and suppress artifacts for the V_s inversion.

NUMERICAL EXAMPLE

We select a part of the Sigsbee2A model (Figure 10a and b) to test the inversion algorithm and strategy. The V_s model is generated through V_p model with a constant scale factor of 0.66. The initial model for E-WERTI are shown in Figure 10c and d, which increase linearly with depth. The initial model of V_p is generally lower (from 1500m/s to 1996m/s) while V_s is higher (from 990m/s to 1317m/s) than the true model. 48 shots are evenly deployed on the surface, in which 8 shots are beyond the right boundary of the model to gain a better illumination. 320 receivers are fixed at the calculation area. The main frequency of P-wave source is 15Hz. The spatial and time interval of forward modeling is 16m and 1.2ms, respectively.

Figure 11 is the results of ERTM and EFWI with the initial model. Since the initial V_p and V_s model are far from the true value, both the PP image (δV_p) and PS image (δV_p) are wrongly in the ERTM. Especially, the refracions are not correctly migrated in PP image and faults are almost disappear in PS image. Meanwhile, we try the EFWI with the above initial model as well. During the EFWI, a four-stage hierarchical strategy from low to high frequency is applied through the time-domain low-pass filter, in which the frequency bands are 0-2Hz, 0-4Hz, 0-6Hz and 0-8Hz. We can see the showllow parts of inverted results is acceptable due to the exsistence of diving wave and low-frequency data. However, the inversion of deep part suffer from the severe cycle-skipping problem because of the absence of low-to-intermediate components in the initial model, which leads to the failure of inversion.

Starting from this initial model, we implement the proposed two-stage ERTI workflow. During the inversion, the direct waves are muted to make sure that only the reflection data are involved. Figure 12a and 12b show the inverted V_p and V_s model of ERTI. After 40 iterations for each stage, ERTI provides a good recovery of the low-to-intermediate components of V_p and V_s model. Besides, we apply the structure-oriented smooth filter to the gradients of inversion to regularize the model update based on the seismic image and structure tensor (???). Note, the structure-oriented smooth filter is an appropriate regularization to stabilize and even accelerate the convergence. In each iteration of ERTI, we constrain the gradients of ERTI with the seismic image obtained by ERTM. Figure 15b shows the gradient of V_p smoothed by an isotropic filter in the first iteration. Though there is no high wavenumber artifacts, the illumination footprints of acquisition geometry are obvious, which will cause uncertainties during inversion. After the structure-oriented smooth, the gradient is more balanced for the inversion of background velocity.

Using the inverted results of ERTI as starting models, we also perform the conventional ERTM. As shown in Figure 13c and d, since the low-to-intermediate wavenumber components of V_p and V_s are well recovered, both of the PP and PS image are greatly improved. The location of reflectors and faults are almost correct and the refractions are migrated as well. In addition, figure 14 shows the comparison between the demigrated PP and PS reflections using initial and inverted model with the true model. We can see that the main reflections of both PP and PS are well matched after the ERTI. Therefore, the cycle-skipping problem will be mitigated greatly if using the starting model in figure 12 for EFWI. Figure 16 show the inverted results of conventional EFWI with the same hierarchical strategy as the previous example. Comparing with the Fig 11, EFWI with the new starting model improves quite a lot, especially in the deep part. But on the bottom of model, the

background components are not well recovered due to the lack of reflection coverage. That's why EFWI still cannot provide good recovery of the model. The vertical profiles at 1.4km and 3km validate the effectiveness of ERTI+EFWI workflow.

Usually, the low-frequency components in the data and the low-to-intermediate wavenumber in the initial model are two key points to mitigate the cycle-skipping problem in FWI. In real data, the low frequency components are difficult to obtain and also more easily contaminated by the noises. Good initial models can compensate for the lack of low frequency data. To check the robustness of the above ERTI models for EFWI, we will test the EFWI's dependency on the low-frequency data. The low-cut frequency threshold are 3Hz, 5Hz and 7Hz. The components lower than the threshold are filtered out when applying the similar hierarchical strategy from low to high frequency in time domain. As shown in figure 18, even starting from 5Hz, EFWI still provides the acceptable inverted results like the no-low-frequency-cut test (figure 16). When the starting frequency goes up to 7Hz, because of the insufficient wavenumber coverage, the inversion acts more like a least-squares migration process which concentrates on the high-wavenumber update, especially in the deep part. Nonetheless, the reconstructed low-to-intermediate components by ERTI can help to reduce the dependency of EFWI on low-frequency data.

DISCUSSION

Besides DIW method, the temporal or spatial correlation is also widely used to extract the traveltime differences (van Leeuwen and Mulder, 2010; ?; Wang et al., 2015a):

$$\begin{cases} \Delta t(h) = \arg \max_{\Delta t} \int d^o(t + \Delta t, h) d^c(t, h) dt, \\ E_t = \frac{1}{2} \sum \| \Delta t(h) \|^2 \end{cases} \quad (19)$$

or

$$\begin{cases} \Delta h(t) = \arg \max_{\Delta h} \int d^o(t, h + \Delta h) d^c(t, h) dt, \\ E_h = \frac{1}{2} \sum \| \Delta h(t) \|^2 \end{cases} \quad (20)$$

The traveltime residual extracted by correlation function is the global average delay of a single trace or time. If there are multi-arrivals in the seismograms, the correlation-based method can not accurately extract the local travelttime information, whcih may increase the difficulty of event-to-event travelttime fitting. Nevertheless, DIW can estimate the local variation of travelttime differences , which have a higher resolution than the correlation-based method.

To investigate the behaviors of DIW and correlation-based objective function, we use three models of different complexity as shown in figure 19. For simplicity, we only investigate the behavior of PP travelttime misfit function, which is almost the same as the acoustic case. A constant background velocity ($V_p = 2500m/s$) model and different reflectivity model are used to implement the Born modeling. We perturb the background velocity to calculate the objective function through migration/demigration process.

As shown in figure 19, we can see that for the simple one-reflector case, the temporal correlation objective function has a better convexity than the spatial correlation and DIW method. But when there are multi-reflectors, the temporal correlation fails to find the correct minimal due to the multi-arrivals in the seismograms. While the behavior of spatial correlation and DIW method are robust to locate the correct solution. In the third test, we choose a part of Sigsbee2A model to simulate a complicated case. Figure 19e shows that the temporal correlation still fails to find the correct solution and the DIW method even has a best convexity, especailly when the initail velocity is high. The above investigation indicates that DIW method is robust and effective to extract travelttime residual, especially

for the complicated model.

CONCLUSIONS

We extend the wave equation reflection traveltime inversion towards elastic media to build the low-wavenumber component of elastic model. RTI can recover the low-wavenumber components effectively, because time misfits are more sensitive and linearly related to the low-wavenumber model perturbation. Compared with the correlation based method, DIW method provides more stable estimation of traveltimes shifts especially when seismograms contain multi-reflections. The complicated mode conversions in elastic case make the reflection kernels very complex. The kernel analyses show that spatial mode decomposition and P/S separation of multicomponent data can suppress artifacts and recover the correct reflection wavepath in gradient calculation. With the aid of P/S separation of 3C seismograms, we can obtain the separated PP and PS seismograms and then get travel time differences of PP and PS through DIW, respectively. To build the long-wavelength component of the model, we introduce a two-stage ERTI workflow by firstly using PP then PS reflections, through which the nonlinearity of reflection inversion is reduced effectively. In the second stage, the wave mode decomposition is introduced to calculate the gradient of V_s to mitigate the trade-off between V_p and V_s . The Sigsbee2A model example shows that even starting with a bad initial model, the two-stage E-WERTI can provide reliable starting model for conventional EFWI.

ACKNOWLEDGEMENT

This work is supported by the National Natural Science Foundation of China (NO.41474099, 41674117 & 41630964). This paper is also based upon the work supported by the King Abdullah University of Science and Technology (KAUST) Office of Sponsored Research (OSR) under award NO. 2230. We appreciate the open-source package of DENISE from <https://github.com/daniel-koehn/> and Mines Java Toolkit from <https://github.com/dhale>. We thank the useful advice from Tariq Alkhalifah (KAUST), Zedong Wu (KAUST), and Benxin Chi (Los Alamos).

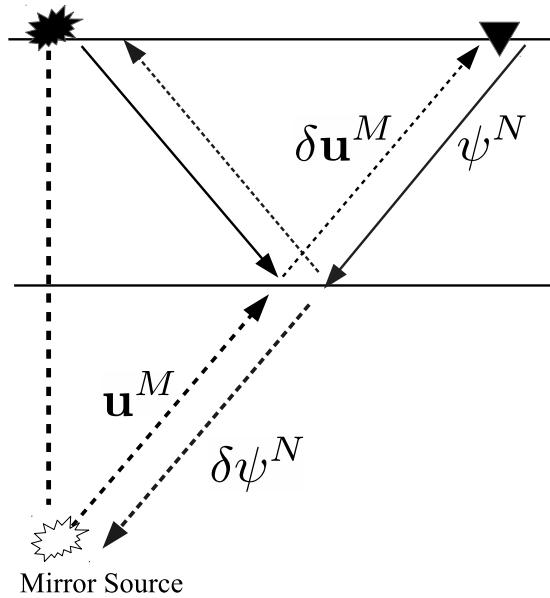


Figure 1: Schematic illustration of the reflection kernels K^{MN} . The cross-correlation can be regarded as the conventional FWI kernel between the mirror source and the receiver. The kernel can be a migration impulse or a transimission-type wavepath.

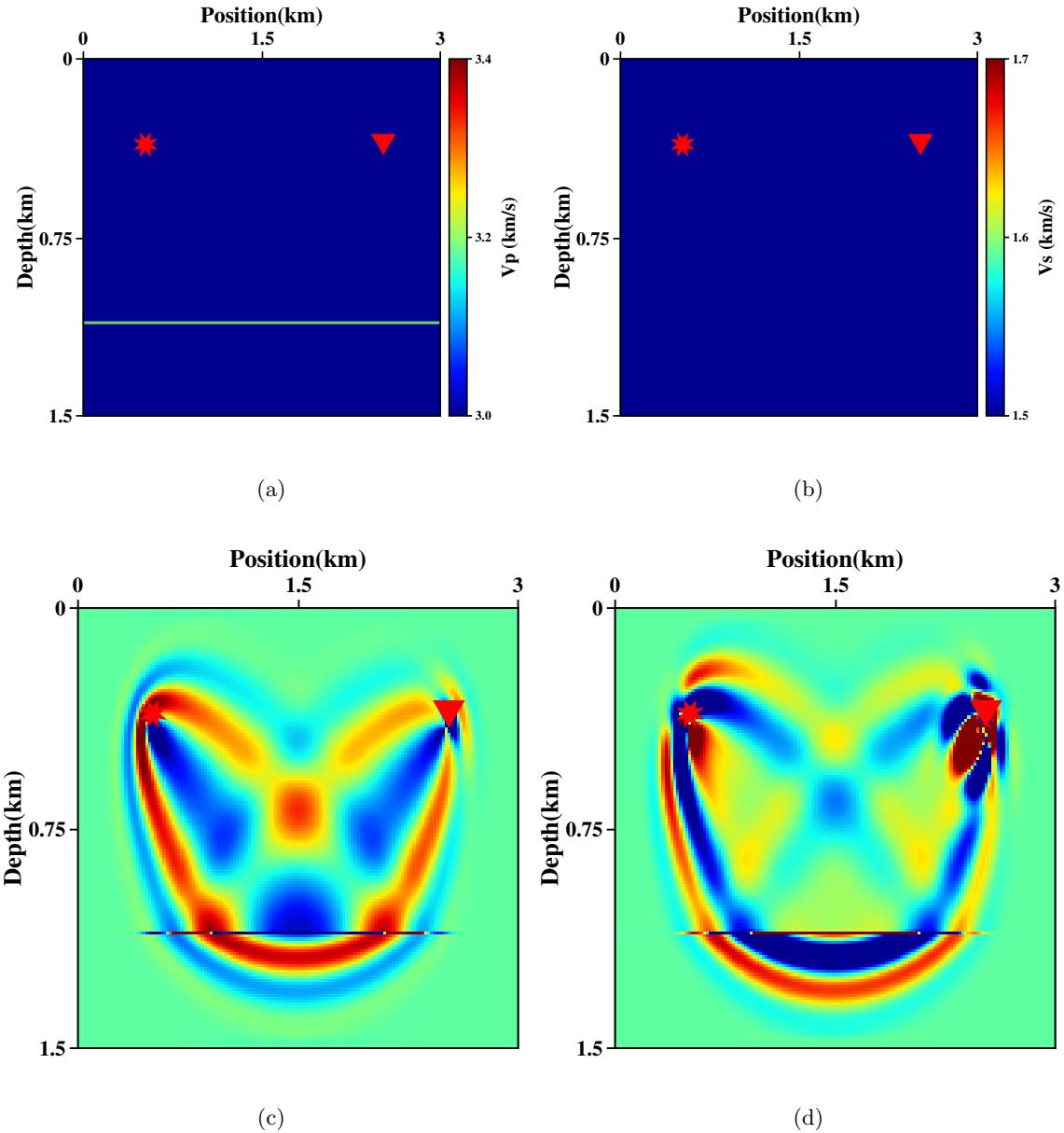


Figure 2: Kernels with single reflector in V_p model. (a) V_p model, (b) V_s model, (c) K_{V_p} , (d) K_{V_s} .

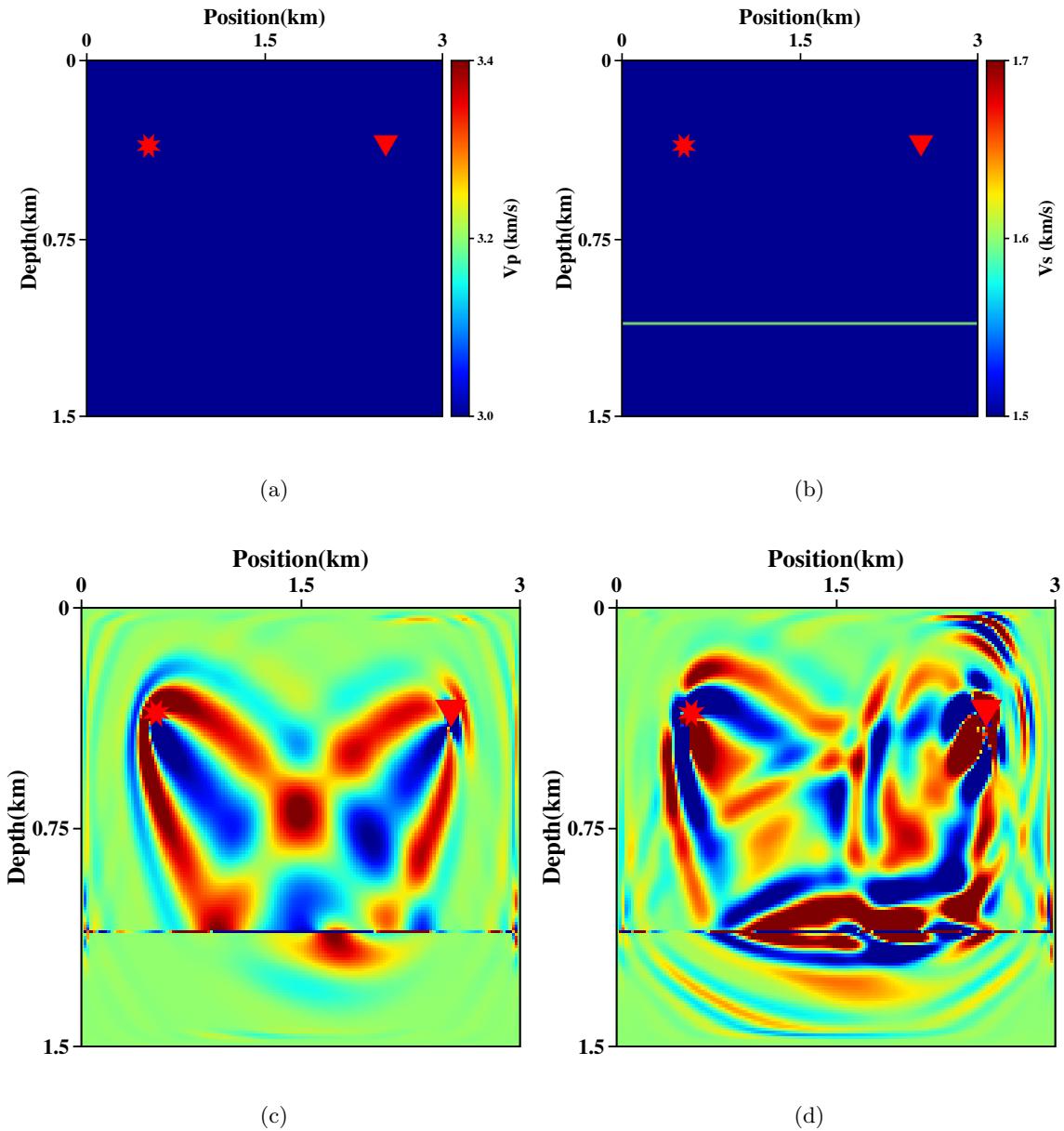


Figure 3: Kernels with single reflector in V_s model. (a) V_p model, (b) V_s model, (c) K_{V_p} , (d) K_{V_s} .

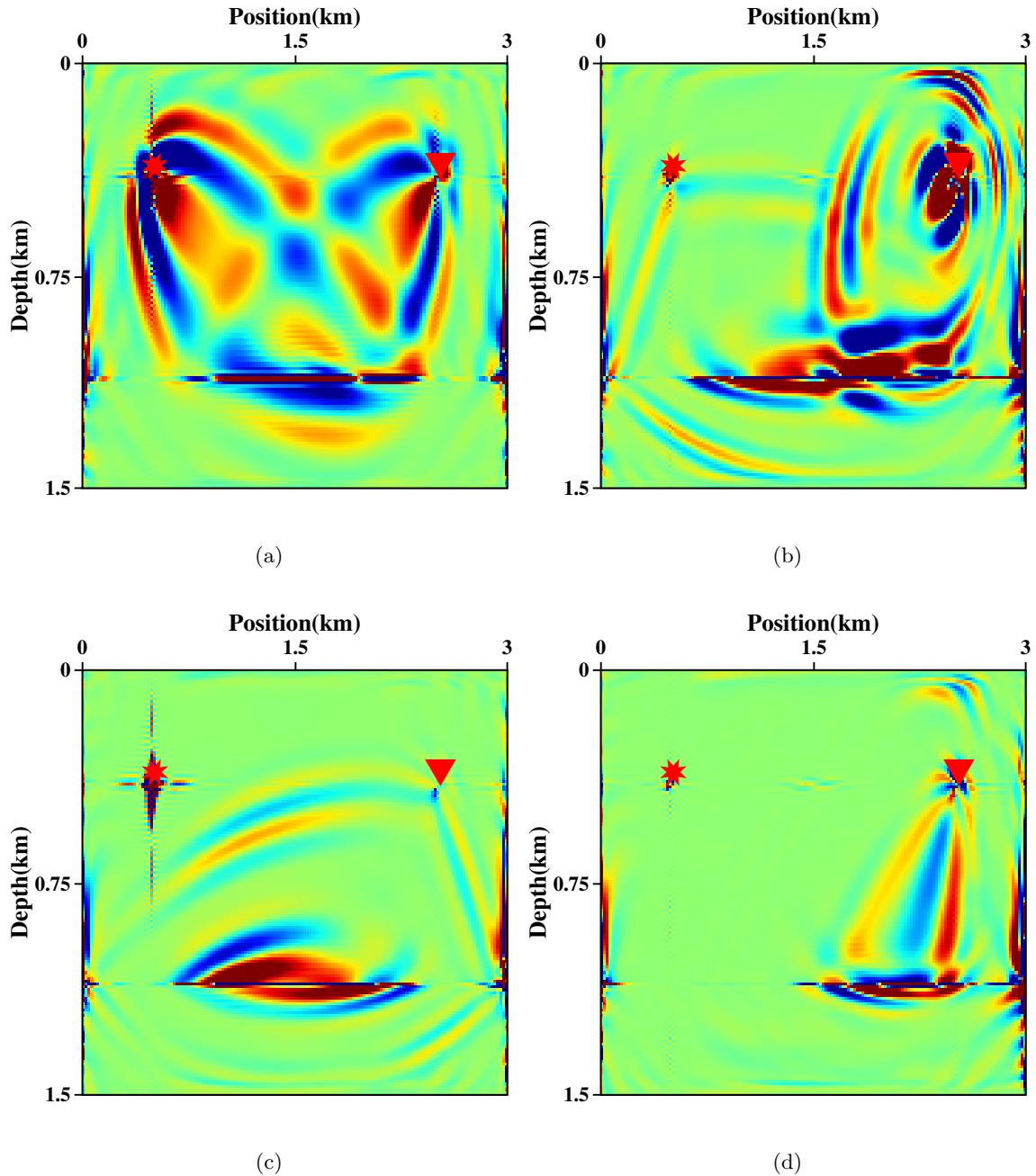


Figure 4: Four components of K_{V_s} . (a) $K_{V_s}^{PP}$, (b) $K_{V_s}^{PS}$, (c) $K_{V_s}^{SP}$, (d) $K_{V_s}^{SS}$.

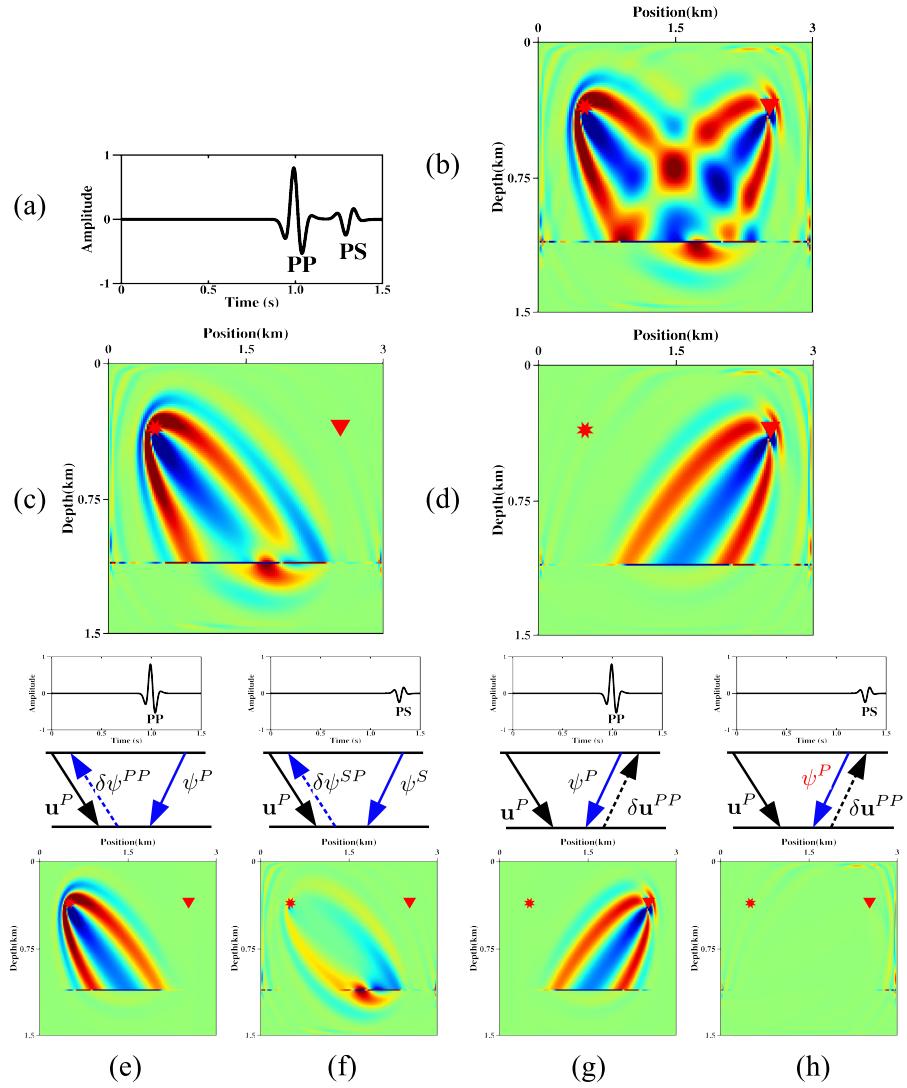


Figure 5: Further decomposition of K_{V_p} for the second test by injecting the PP and PS event separately as adjoint source and decomposing the wavefields with different manners. (a) Original adjoint source for back-propagation, (b) K_{V_p} , (c) Source side of K_{V_p} , (d) Receiver side of K_{V_p} . (e)-(h): The first line denotes the adjoint source. The second line denotes the manner of mode decomposition. The last line denote the corresponding kernel. Note that, when injecting PS data, a non-physical SP conversion (red) is generated in panel (h)

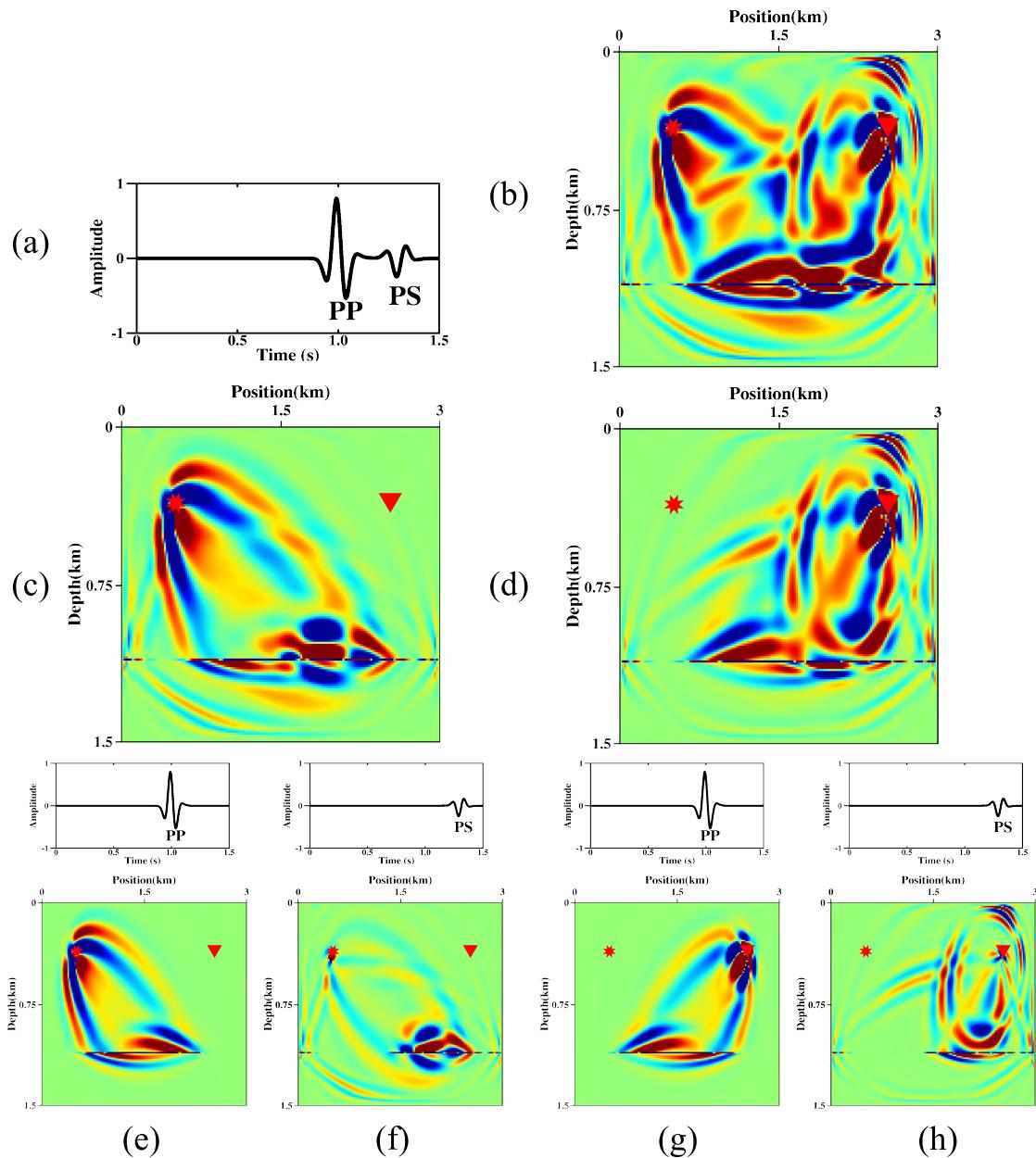


Figure 6: Further decomposition of K_{V_s} for the second test. (a): Original adjoint source , (b) K_{V_s} , (c) Source term of K_{V_s} , (d) Receiver term of K_{V_s} . (e)-(h): The first line denotes the adjoint source for injection. The second line denotes the decomposed kernels when separately injecting PP and PS data.

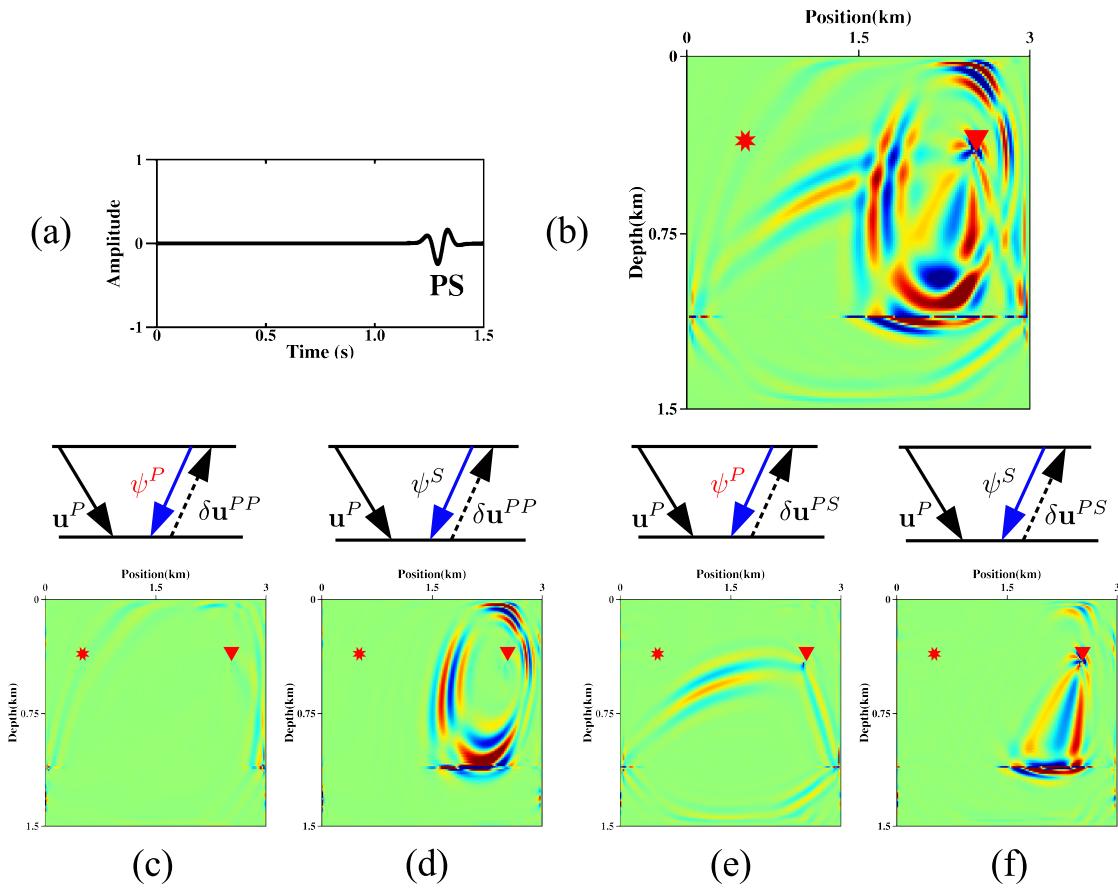


Figure 7: Further decomposed components of Fig 6f. (a) The adjoint source for injection, (b) Receiver term of K_{V_s} using (a) as adjoint source, (c)-(f) The first line denotes the manner and wavefield type for cross-correlation. The second line denotes the corresponding kernels.

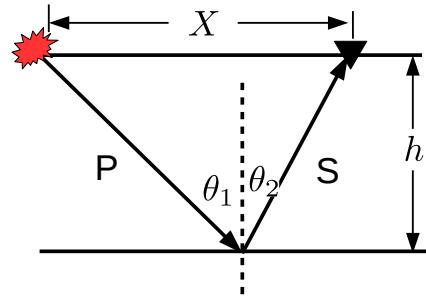


Figure 8: Ray path of PS reflection with a single reflector.

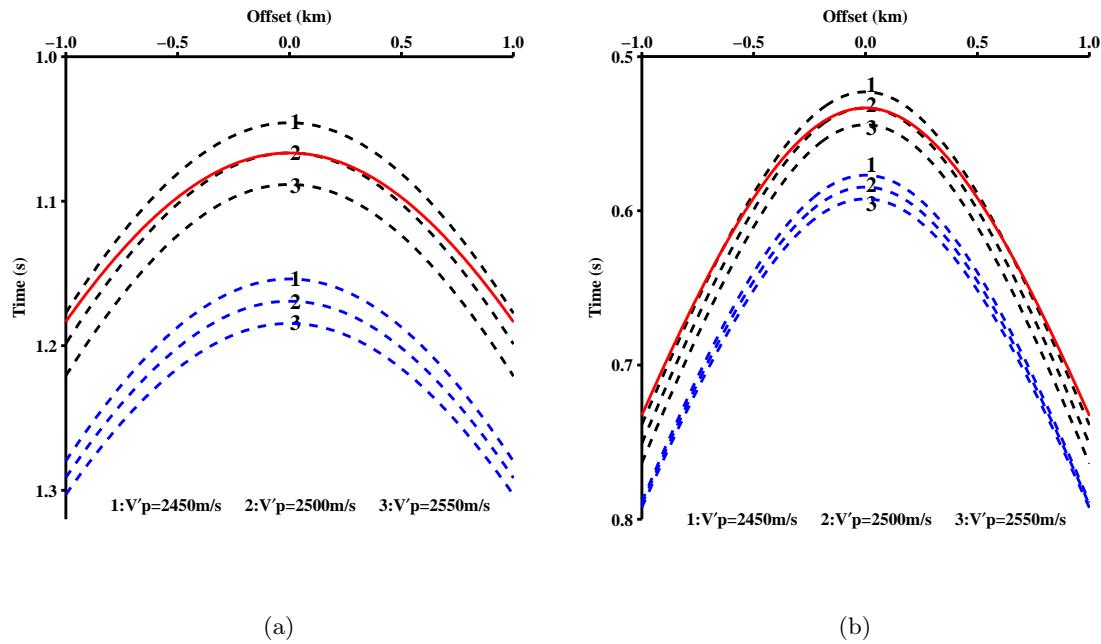


Figure 9: The comparison among demigrated PS traveltimes of method I (black), method II (blue) and the real (red) one. (a) $h = 1000m$, (b) $h = 500m$.

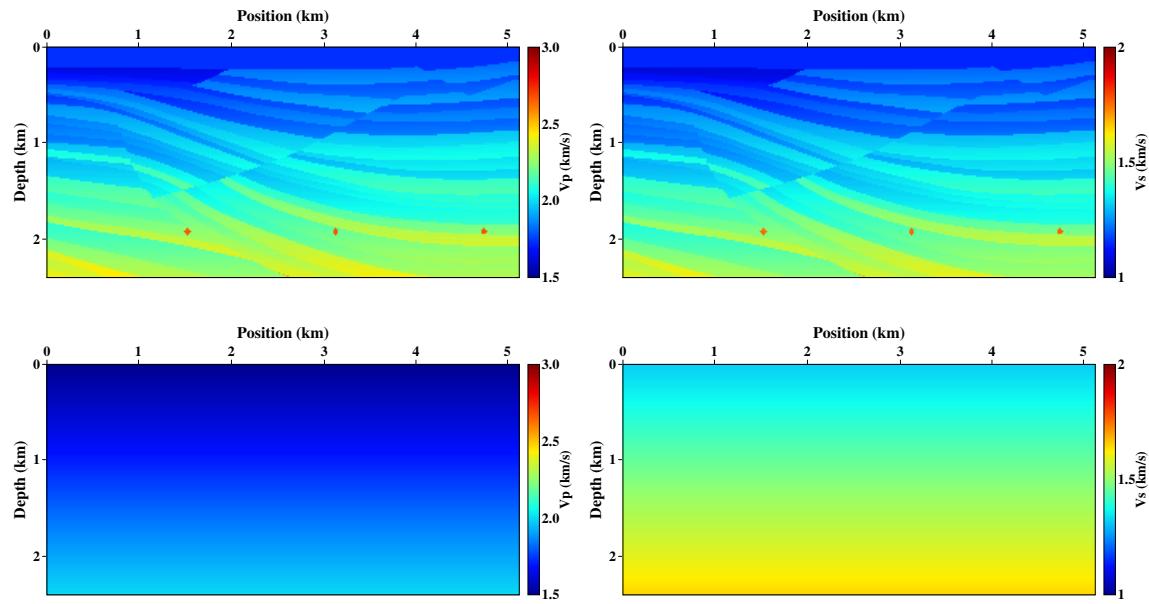


Figure 10: Sigbee2A model example. On the top are true models of V_p (a) and V_s (b). On the bottom are initial models of V_p (c) and V_s (d) linearly increasing with depth.

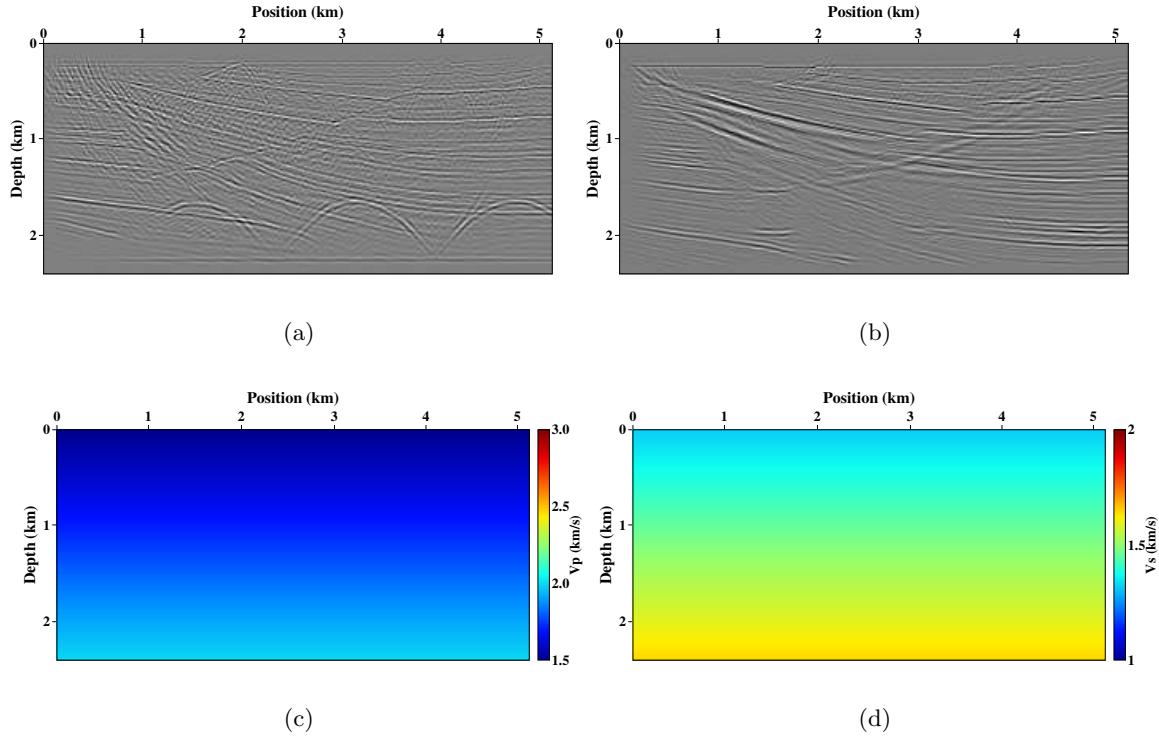


Figure 11: The results of ERTM and EFWI using initial model: (a) and (b) are PP and PS image of ERTM with near offset data, (c) and (d) are inverted V_p and V_s with EFWI.

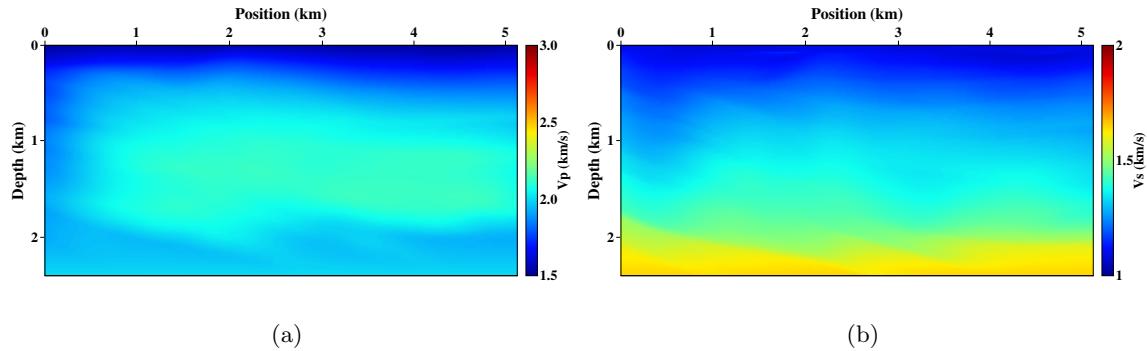


Figure 12: Inverted results of ERTI: (a) V_p , (b) V_s .

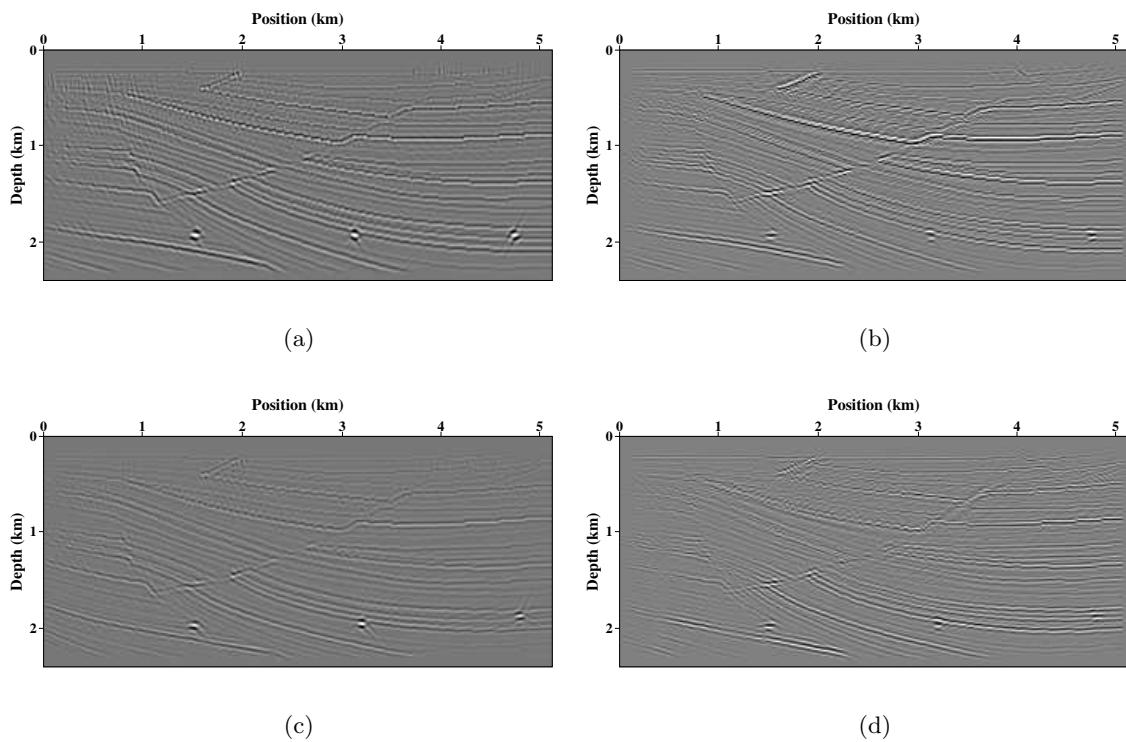


Figure 13: ERTM results using the true model (a, b) and inverted model (c, d). (a) and (c) are the PP image(b) and (d) are the PS image

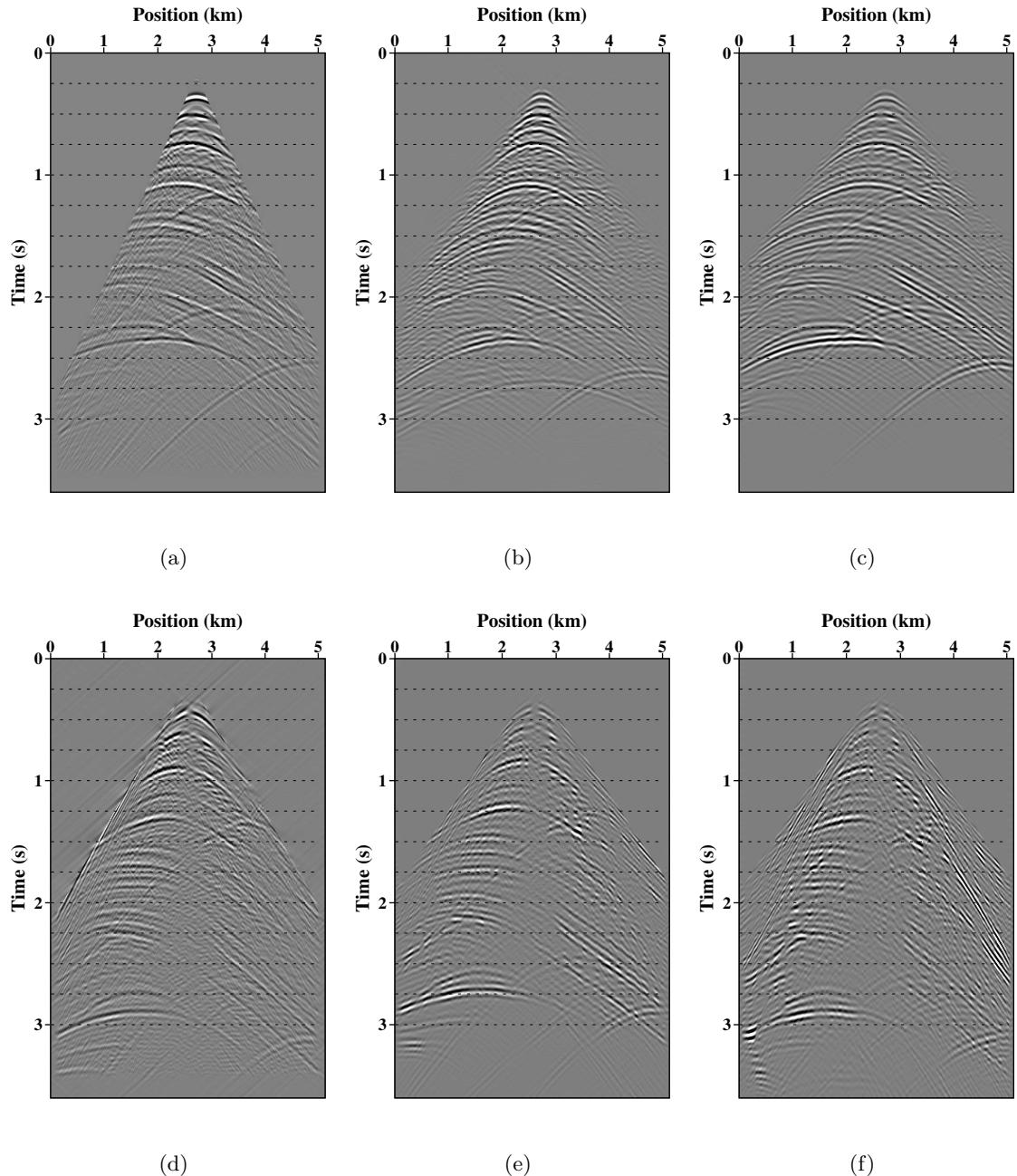


Figure 14: Comparison of the observed and the demigrated reflection data using initial model and the inverted model. The first row are the separated PP reflection, while the second row are the separated PS reflection. The left, middle and right column are the observed reflection data, the demigrated reflection data with initial model and the demigrated data with inverted model, respectively.

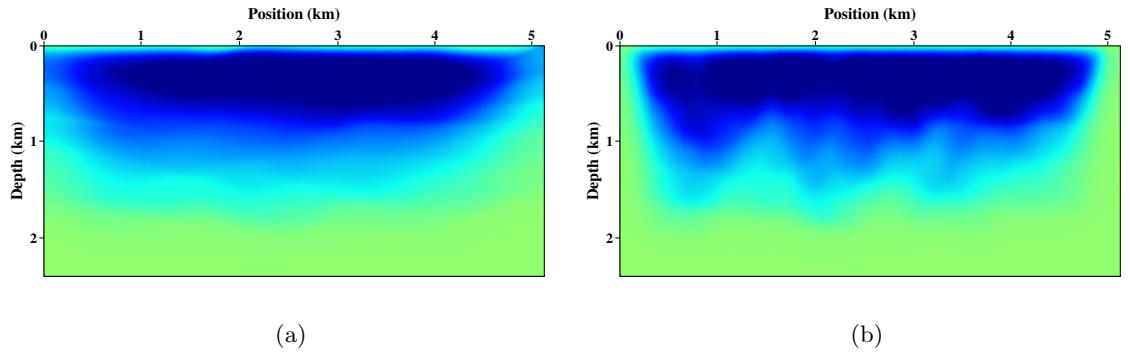


Figure 15: The comparison of gradients and final inverted results with (left) and without (right) the structure-oriented constrain: (a) and (c) are the V_p gradients in the first iteration, (b) and (d) are final inverted V_p model.

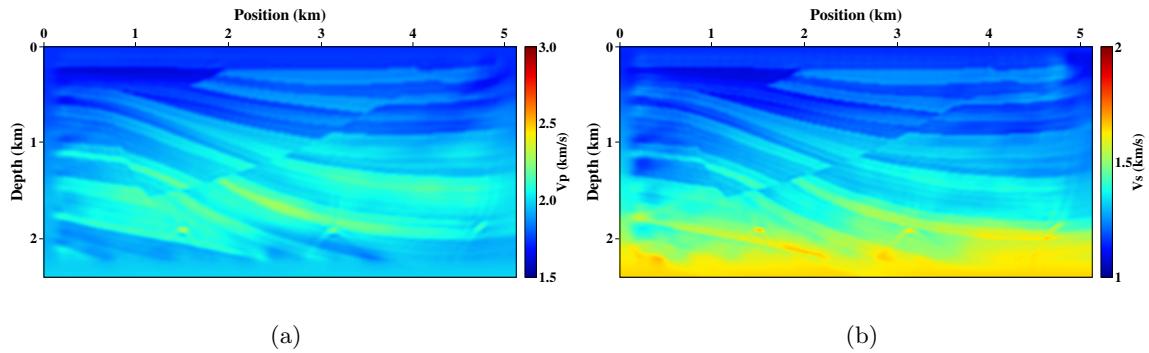


Figure 16: EFWI results using the ERTI model as starting model. (a) V_p , (b) V_s .

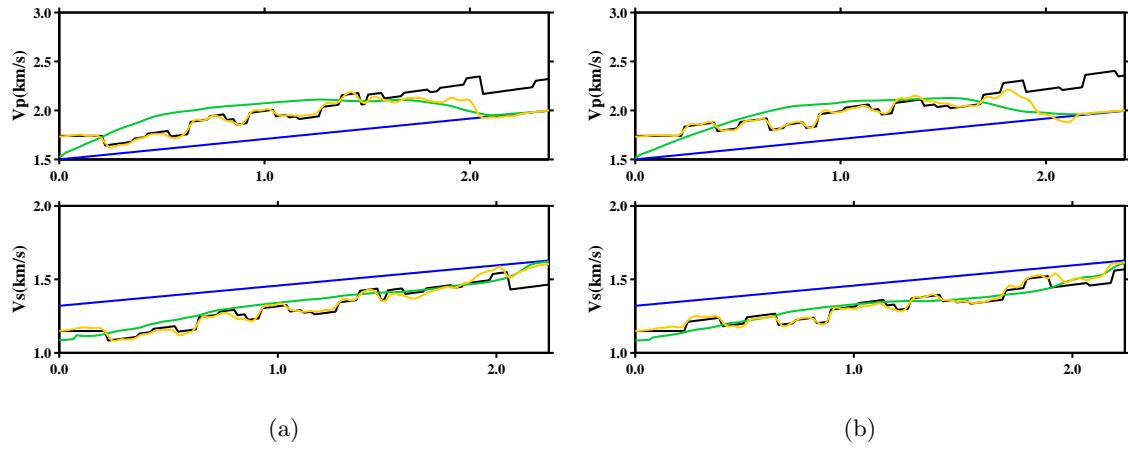


Figure 17: Vertical profile of ERTI and EFWI model at 1.4km (a) and 3km (b). The black and blue lines denote the true and initial model. The green and yellow denote the ERTI and EFWI model.

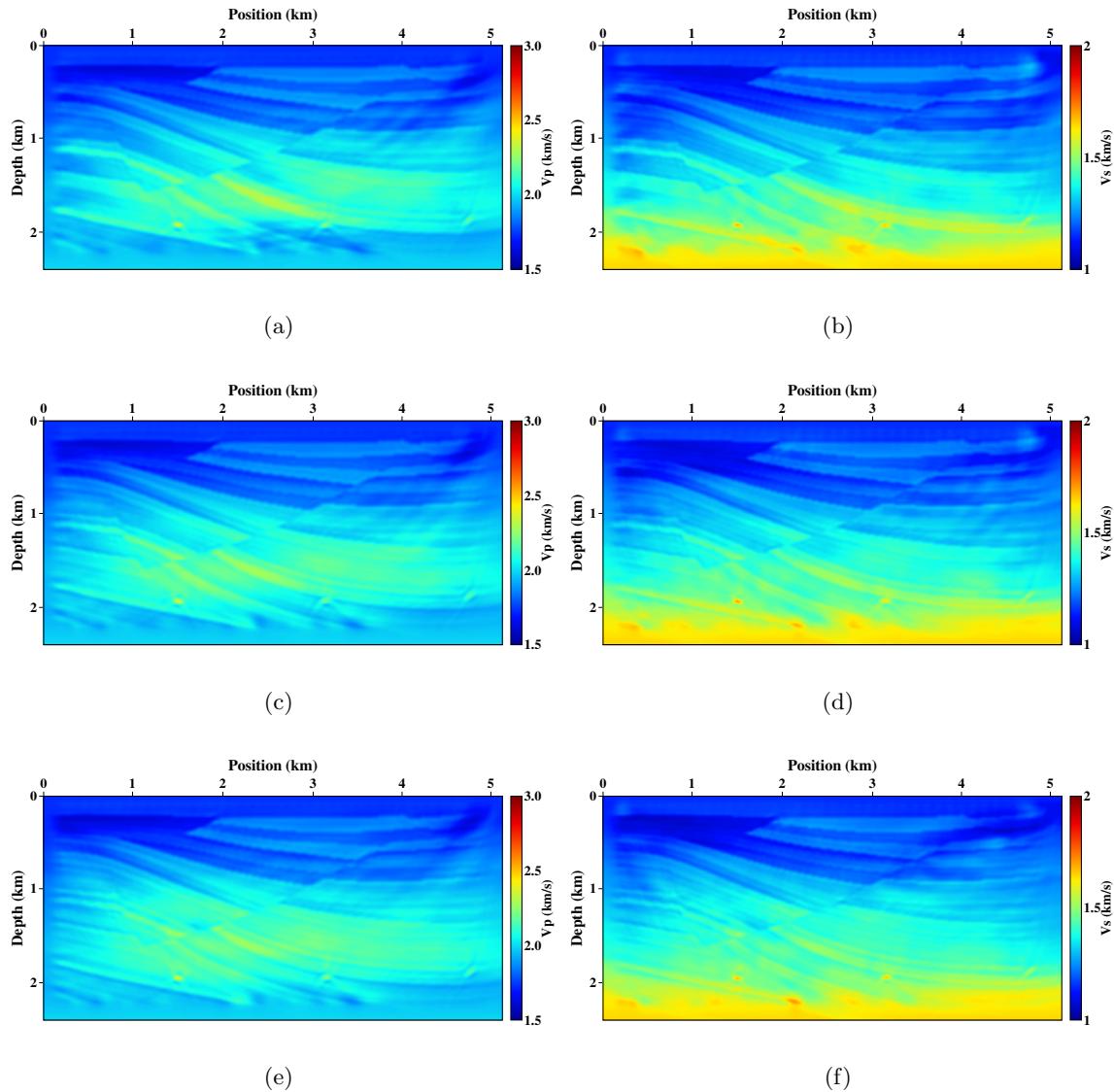


Figure 18: EFWI results with different starting frequency. (a), (c), (e) are V_p , (b), (d), (f) are V_s . The starting frequency are 3Hz, 5Hz and 7Hz from top to bottom row, respectively.

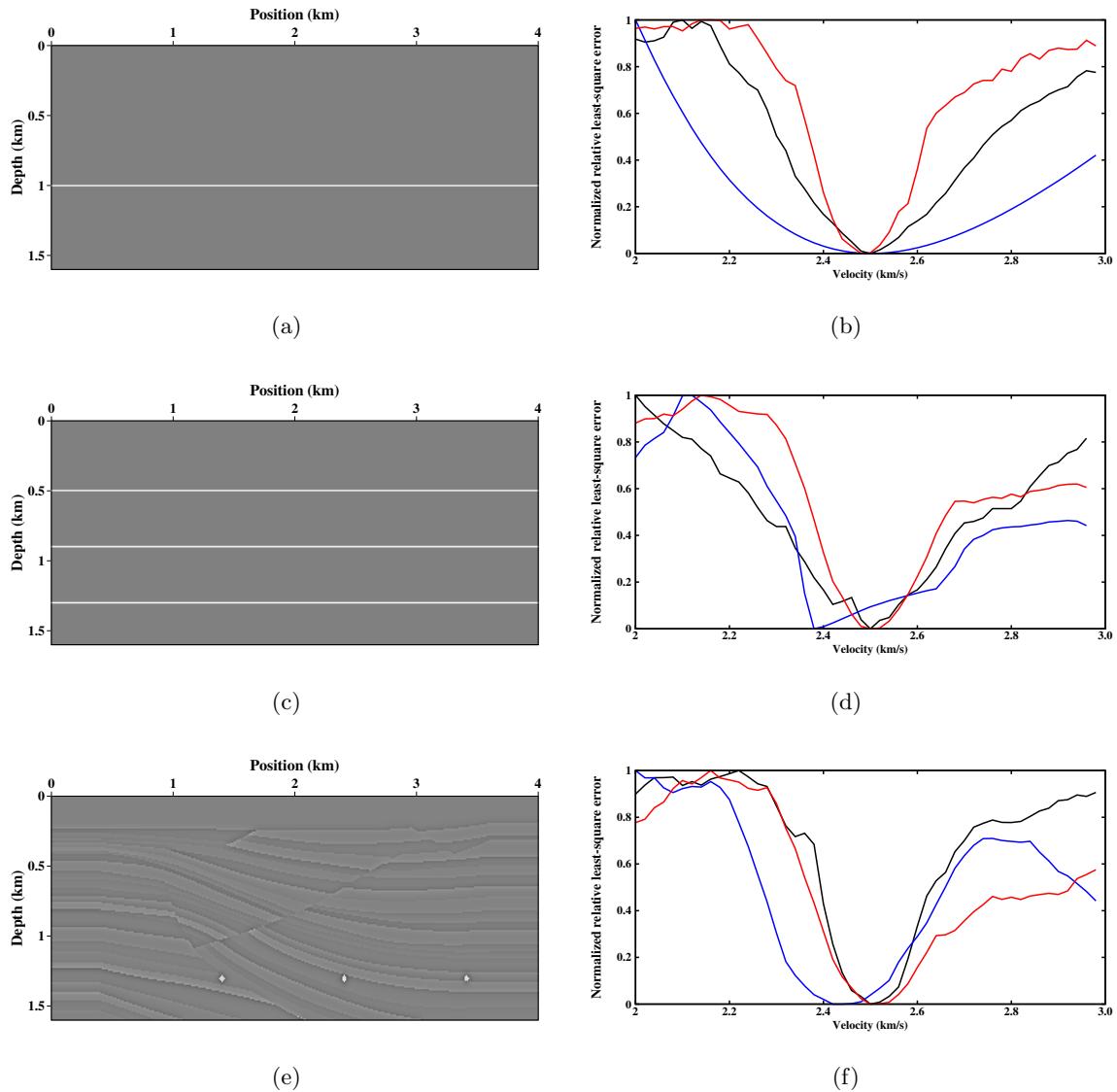


Figure 19: Normalized PP traveltimes misfits for different objective functions for different model;

APPENDIX A

In this appendix, we will derive the gradients of ERTI using adjoint state method Plessix (2006); Liu and Tromp (2006). According to Hale (2013), DIW aims to minimize the distance between observed and calculated data:

$$D(\tau) = \frac{1}{2} \int_0^T \sum_r [(\mathbf{x}_r, t) - d^o(\mathbf{x}_r, t + \tau(\mathbf{x}_r, t))]^2 dt. \quad (21)$$

The optimization of eq.(21) satisfies:

$$\frac{\partial D}{\partial \tau} = \int_0^T \sum_r \boldsymbol{\alpha}(\mathbf{x}_r, t) dt = 0. \quad (22)$$

where $\boldsymbol{\alpha}(\mathbf{x}_r, t) = \dot{d}^o(\mathbf{x}_r, t + \tau)(d^o(\mathbf{x}_r, t + \tau) - d^c(\mathbf{x}_r, t))$. For simplicity during derivation, we rewrite eq (1) and (2) as:

$$\begin{aligned} \rho \partial_t^2 \mathbf{u} - \nabla \cdot (\mathbf{c}^0 : \nabla \mathbf{u}) &= \mathbf{f} \\ \rho \partial_t^2 \hat{\mathbf{u}} - \nabla \cdot (\mathbf{c}^0 : \nabla \hat{\mathbf{u}}) &= \nabla \cdot (\mathbf{c}^1 : \nabla \mathbf{u}), \end{aligned} \quad (23)$$

where ∂_t is the time derivative. Therefore, using eq (23) and (22) we can define the Lagrangian \mathcal{L} :

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sum_r \int_0^T \boldsymbol{\tau}^2(\mathbf{x}_r, t) dt - \sum_r \int_0^T \boldsymbol{\mu}(\mathbf{x}_r, t) \boldsymbol{\alpha}(\mathbf{x}_r, t) dt \\ & - \int_0^T \int_{\Omega} \hat{\psi} (\rho \partial_t^2 \mathbf{u} - \nabla \cdot (\mathbf{c}^0 : \nabla \mathbf{u}) - \mathbf{f}) d^3 \mathbf{x} dt \\ & - \int_0^T \int_{\Omega} \psi (\rho \partial_t^2 \hat{\mathbf{u}} - \nabla \cdot (\mathbf{c}^0 : \nabla \hat{\mathbf{u}}) - \nabla \cdot (\mathbf{c}^1 : \nabla \mathbf{u})) d^3 \mathbf{x} dt, \end{aligned} \quad (24)$$

in which $\boldsymbol{\tau}$, \mathbf{u} and $\hat{\mathbf{u}}$ are state variables, Ω is the integration domain, and $\boldsymbol{\mu}$, ψ and $\hat{\psi}$ are Lagrange multipliers that need to be determined. Here, we only consider the variation of

\mathbf{c}^0 , then the change of eq. (24) is:

$$\begin{aligned}
\delta \mathcal{L} = & \int_0^T \sum_r [\boldsymbol{\tau}(\mathbf{x}_r, t) - \boldsymbol{\mu}(\mathbf{x}_r, t)\mathbf{h}(\mathbf{x}_r, t)] \delta(\mathbf{x} - \mathbf{x}) \cdot \delta \boldsymbol{\tau} dt \\
& - \int_0^T \int_{\Omega} \boldsymbol{\mu}(\mathbf{x}_r, t) \dot{\mathbf{d}}^o(\mathbf{x}_r, t + \boldsymbol{\tau}) \delta(\mathbf{x} - \mathbf{x}) \cdot \delta \hat{\mathbf{u}} d^3 \mathbf{x} dt \\
& + \int_0^T \int_{\Omega} [\hat{\psi} \nabla \cdot (\delta \mathbf{c}^0 : \nabla \mathbf{u}) + \psi \nabla \cdot (\delta \mathbf{c}^0 : \nabla \hat{\mathbf{u}})] d^3 \mathbf{x} dt \\
& - \int_0^T \int_{\Omega} \hat{\psi} (\rho \partial_t^2 \delta \mathbf{u} - \nabla \cdot (\mathbf{c}^0 : \nabla \delta \mathbf{u})) d^3 \mathbf{x} dt \\
& - \int_0^T \int_{\Omega} \psi (\rho \partial_t^2 \delta \hat{\mathbf{u}} - \nabla \cdot (\mathbf{c}^0 : \nabla \delta \hat{\mathbf{u}}) - \nabla \cdot (\mathbf{c}^1 : \nabla \delta \mathbf{u})) d^3 \mathbf{x} dt,
\end{aligned} \tag{25}$$

Upon integrating the terms involving spatial and temporal derivatives of \mathbf{u} , $\delta \mathbf{u}$, $\hat{\mathbf{u}}$ and $\delta \hat{\mathbf{u}}$

by parts, we have:

$$\begin{aligned}
\delta \mathcal{L} = & \int_0^T \sum_r [\boldsymbol{\tau}(\mathbf{x}_r, t) - \boldsymbol{\mu}(\mathbf{x}_r, t)\mathbf{h}(\mathbf{x}_r, t)] \delta(\mathbf{x} - \mathbf{x}) \cdot \delta \boldsymbol{\tau} dt \\
& - \int_0^T \int_{\Omega} \boldsymbol{\mu}(\mathbf{x}_r, t) \dot{\mathbf{d}}^o(\mathbf{x}_r, t + \boldsymbol{\tau}) \delta(\mathbf{x} - \mathbf{x}) \cdot \delta \hat{\mathbf{u}} d^3 \mathbf{x} dt \\
& - \int_0^T \int_{\Omega} [\nabla \hat{\psi} : \delta \mathbf{c}^0 : \nabla \mathbf{u} + \nabla \psi : \delta \mathbf{c}^0 : \nabla \hat{\mathbf{u}}] d^3 \mathbf{x} dt \\
& - \int_0^T \int_{\Omega} (\rho \partial_t^2 \hat{\psi} - \nabla \cdot (\mathbf{c}^0 : \nabla \hat{\psi}) - \nabla \cdot (\mathbf{c}^1 : \nabla \psi)) \delta \mathbf{u} d^3 \mathbf{x} dt \\
& - \int_0^T \int_{\Omega} (\rho \partial_t^2 \psi - \nabla \cdot (\mathbf{c}^0 : \nabla \psi)) \delta \hat{\mathbf{u}} d^3 \mathbf{x} dt \\
& - \int_{\Omega} [\rho (\hat{\psi} \cdot \partial_t \delta \mathbf{u} - \partial_t \hat{\psi} \cdot \delta \mathbf{u} + \psi \cdot \partial_t \delta \hat{\mathbf{u}} - \partial_t \psi \cdot \delta \hat{\mathbf{u}})]_0^T d^3 \mathbf{x} \\
& + \int_0^T \int_{\partial \Omega} \hat{\psi} \cdot [\mathbf{n} \cdot (\delta \mathbf{c}^0 : \nabla \mathbf{u} + \mathbf{c}^0 : \nabla \delta \mathbf{u})] - \mathbf{n} \cdot (\mathbf{c}^0 : \nabla \hat{\psi}) \cdot \delta \mathbf{u} d^2 \mathbf{x} dt \\
& + \int_0^T \int_{\partial \Omega} \psi \cdot [\mathbf{n} \cdot (\delta \mathbf{c}^0 : \nabla \hat{\mathbf{u}} + \mathbf{c}^0 : \nabla \delta \hat{\mathbf{u}})] - \mathbf{n} \cdot (\mathbf{c}^0 : \nabla \psi) \cdot \delta \hat{\mathbf{u}} d^2 \mathbf{x} dt \\
& + \int_0^T \int_{\partial \Omega} \psi \cdot [\mathbf{n} \cdot (\mathbf{c}^1 : \nabla \delta \mathbf{u})] - \mathbf{n} \cdot (\mathbf{c}^1 : \nabla \psi) \cdot \delta \mathbf{u} d^2 \mathbf{x} dt
\end{aligned} \tag{26}$$

where \mathbf{n} is the unit outward vector normal on the surface $\partial \Omega$. The displacement wavefields (state variables) are subject to the initial and boundary condition:

$$\begin{aligned}
\mathbf{u}(\mathbf{x}, 0) = 0, \partial_t \mathbf{u}(\mathbf{x}, 0) = 0, \mathbf{u}(\mathbf{x}, t)|_{x \rightarrow \infty} \rightarrow 0 \\
\hat{\mathbf{u}}(\mathbf{x}, 0) = 0, \partial_t \hat{\mathbf{u}}(\mathbf{x}, 0) = 0, \hat{\mathbf{u}}(\mathbf{x}, t)|_{x \rightarrow \infty} \rightarrow 0
\end{aligned} \tag{27}$$

while the adjoint state variables satisfy the “final” (at Time T) and boundary condition:

$$\begin{aligned}\psi(\mathbf{x}, T) &= 0, \partial_t \psi(\mathbf{x}, T) = 0, \psi(\mathbf{x}, t)|_{x \rightarrow \infty} \rightarrow 0 \\ \hat{\psi}(\mathbf{x}, T) &= 0, \partial_t \hat{\psi}(\mathbf{x}, T) = 0, \hat{\psi}(\mathbf{x}, t)|_{x \rightarrow \infty} \rightarrow 0\end{aligned}\tag{28}$$

Thus all the surface integrals in eq.(26) will disappear. To obtain the stationary point, setting the coefficients of $\delta\boldsymbol{\tau}$, $\delta\mathbf{u}$ and $\delta\hat{\mathbf{u}}$ to zero yields three adjoint state equations:

$$\begin{aligned}\boldsymbol{\tau}(\mathbf{x}_r, t) - \boldsymbol{\mu}(\mathbf{x}_r, t)\mathbf{h}(\mathbf{x}_r, t) &= 0, \\ \rho\partial_t^2 \hat{\psi} - \nabla \cdot (\mathbf{c}^0 : \nabla \hat{\psi}) &= \nabla \cdot (\mathbf{c}^1 : \nabla \psi) \\ \rho\partial_t^2 \psi - \nabla \cdot (\mathbf{c}^0 : \nabla \psi) &= \boldsymbol{\mu}(\mathbf{x}_r, t)\dot{\mathbf{d}}^o(\mathbf{x}_r, t + \tau)\end{aligned}\tag{29}$$

where $\mathbf{h}(\mathbf{x}_r, t) = (\dot{d}_i^o(\mathbf{x}_r, t + \tau))^2 - \ddot{d}_i^o(\mathbf{x}_r, t + \tau)(d_i^c(\mathbf{x}_r, t) - d_i^o(\mathbf{x}_r, t + \tau))$. The adjoint state variables can be obtained by solving eq.(29). Then, the gradient is:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{c}^0} = - \int_0^T \int_{\Omega} [\nabla \hat{\psi} \nabla \mathbf{u} + \nabla \psi \nabla \hat{\mathbf{u}}] d^3 \mathbf{x} dt,\tag{30}$$

or in a more detailed manner:

$$\frac{\partial E}{\partial c_{ijkl}^0} = - \int \left(\frac{\partial u_i}{\partial x_j} \frac{\partial \hat{\psi}_k}{\partial x_l} + \frac{\partial \hat{u}_i}{\partial x_j} \frac{\partial \psi_k}{\partial x_l} \right),\tag{31}$$

APPENDIX B: FURTHER DECOMPOSITION OF V_S KERNEL

In the previously section, we give the further decomposition of Fig 6h. Here, we will investigate the decomposition of the rest three parts, i.e. Fig 6e-6g. Fig. 20 show the decomposition of two source terms. When only injecting PP data, the wavefield ψ^{PP-S} will occur due to the non-physical mode conversions. In ψ^{PP-S} , PP indicates the injection data while S indicates the conversion type at receiver. The same definition is used in the following context. For simplicity, we only label the figure with ψ to represent the wavefields

back-propagated in the background media. It is easy to recognize the PP wavepath shown in Fig. 20c. While the cross-mode correlation produces high-wavenumber energy near the interface. mode conversions will produce different type of migration impulse at the source side. When only injecting PS data, the ψ^{PS-P} will occur as well. Fig. 20g shows a similar case as the Fig. 5f. Note, in 20h there are two migration impulse because two type of S-wave conversions exist in $\hat{\psi}$. First, the normal ψ^{PS} wavefield produce a S-wave reflection ψ^{PS-S} when incidenting at the interface. Second, the injection-induced non-physical conversion at receiver ψ^{PS-P} also generates a S-wave conversion ψ^{PS-P-S} . Separating the above two migration impulses is possible but requires more efforts and calculations in which one should separate the wavefields ψ into ψ^{PS} and ψ^{PS-P} in demigration (solving eq.(??)).

Finally, we decompose the kernel in Fig 6g. Only PP data is injected. The non-physical conversions at receiver are highlighted as red. Figure 21c shows the component of PP reflection wavepath. We can see the cross-mode correlations produce high-wave number energy near the interface or around the receiver location. In Fig.21f, there is no energy in the SS-mode component. One plausible reason is that the two S-wavefields do not meet with each other to give a migration impulse due to the inconsistency of traveltime.

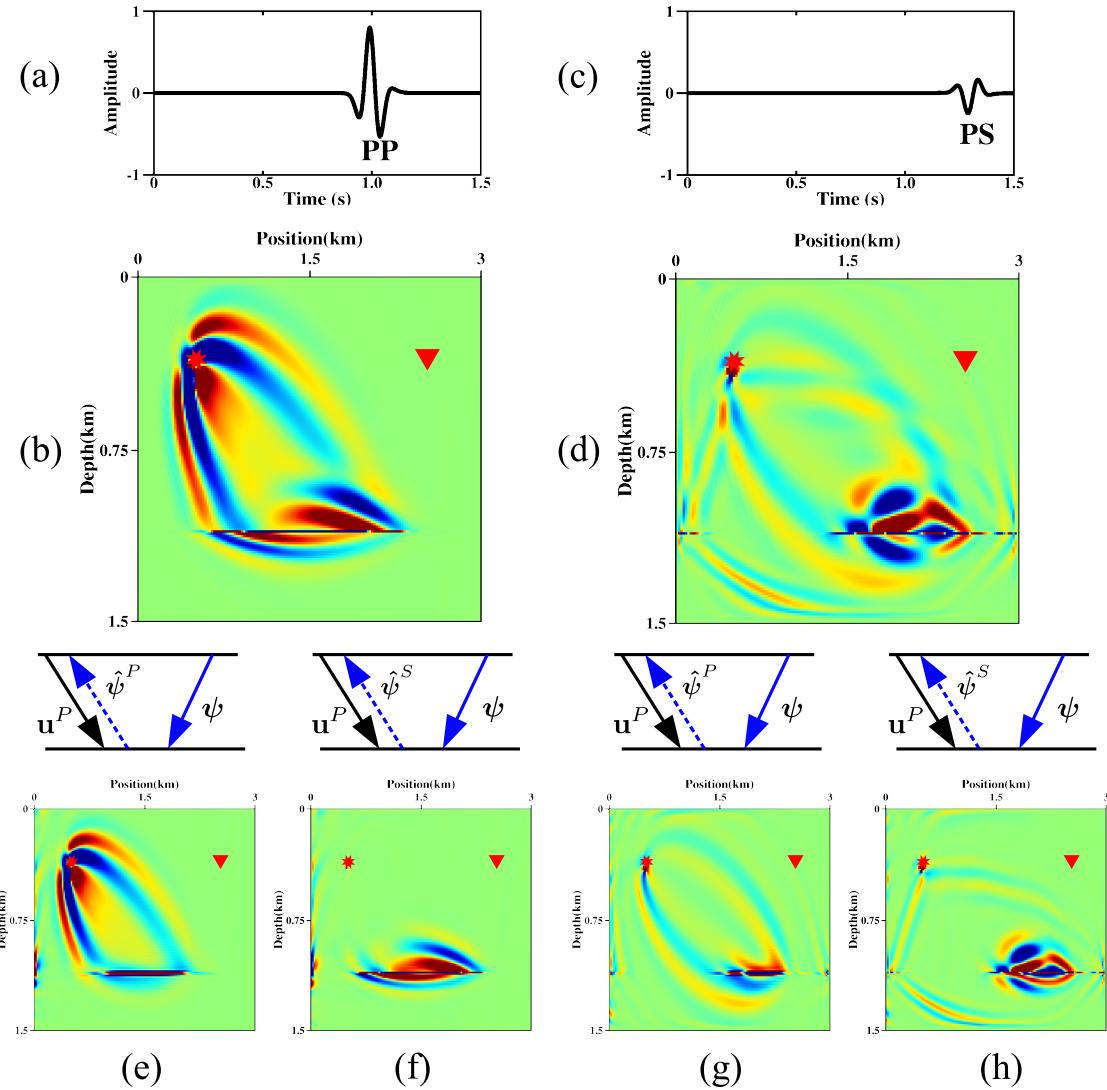


Figure 20: Further decomposed components of Fig 6e and 6f. (a) The PP data adjoint source , (b) Source side of K_{V_s} using (a) as adjoint source, (c) The PS data adjoint source, (d) Source side of K_{V_s} using (c) as adjoint source, (e)-(h): The first line denotes the manner and wavefield type for cross-correlation. The second line denotes the corresponding kernels.

REFERENCES

Almomin, A., and B. Biondi, 2012, Tomographic full waveform inversion: Practical and computationally feasible approach: Presented at the SEG Technical Program Expanded
43

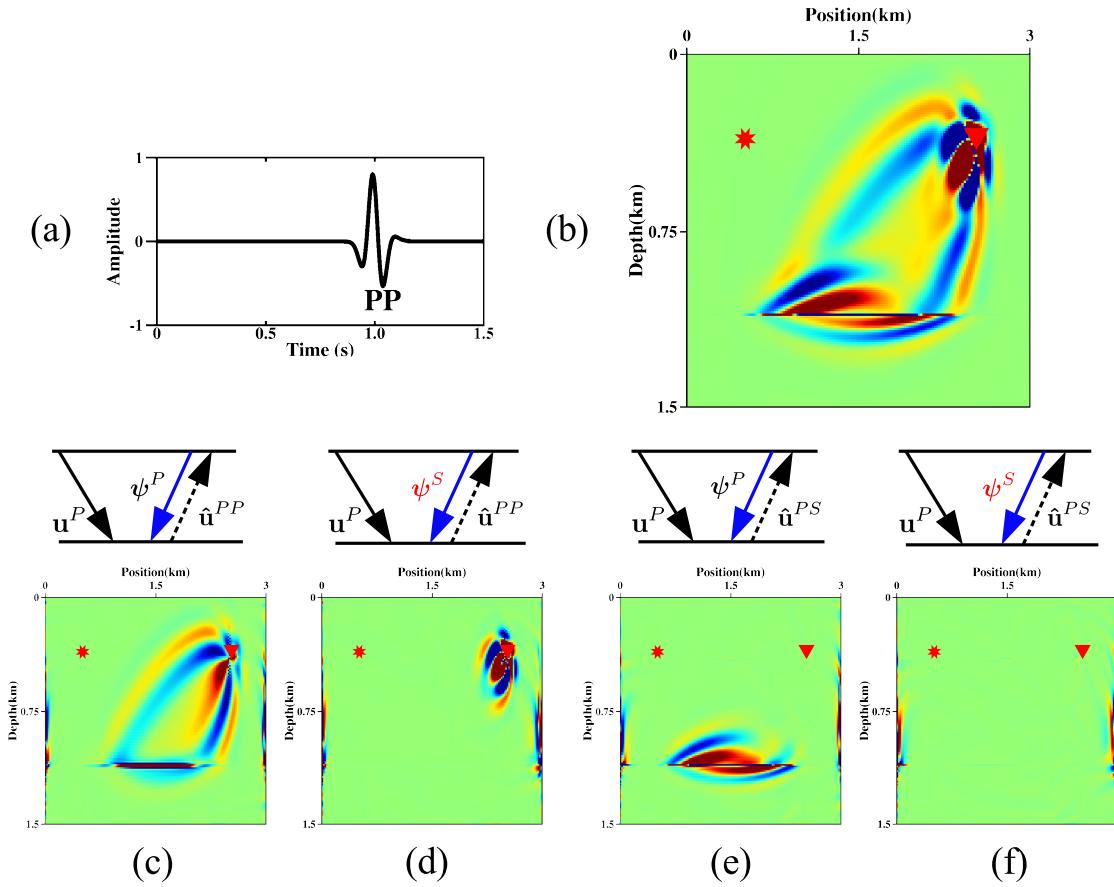


Figure 21: Further decomposed components of Fig 6g. (a) The adjoint source for injection, (b) Receiver term of K_{V_s} using (a) as adjoint source, (c)-(f) The first line denotes the manner and wavefield type for cross-correlation. The second line denotes the corresponding kernels.

Abstracts 2012, Society of Exploration Geophysicists.

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

Brossier, R., S. Operto, and J. Virieux, 2009, Seismic imaging of complex onshore struc-

tures by 2d elastic frequency-domain full-waveform inversion: *Geophysics*, **74**, WCC105–WCC118.

Chavent, G., F. Clément, and S. Gómez, 1994, Automatic determination of velocities via migration-based traveltime waveform inversion: A synthetic data example: Presented at the SEG Technical Program Expanded Abstracts 1994, Society of Exploration Geophysicists.

Chi, B., L. Dong, and Y. Liu, 2015, Correlation-based reflection full-waveform inversion: *Geophysics*, **80**, R189–R202.

Clement, F., G. Chavent, and S. Gomez, 2001, Migration-based traveltime waveform inversion of 2-d simple structures: A synthetic example: *Geophysics*, **66**, 845–860.

Guo, Q., and T. Alkhalifah, 2016, *in* A nonlinear approach of elastic reflection waveform inversion: Society of Exploration Geophysicists, 1421–1425.

Hale, D., 2013, Dynamic warping of seismic images: *GEOPHYSICS*, **78**, S105–S115.

Jones, I., 2010, Tutorial: Velocity estimation via ray-based tomography: *First Break*, **28**.

Li, Z., X. Ma, C. Fu, B. Gu, and G. Liang, 2016, Frequency-wavenumber implementation for p- and s-wave separation from multi-component seismic data: *Exploration Geophysics*, **47**, 32.

Liu, Q., and J. Tromp, 2006, Finite-frequency kernels based on adjoint methods: *Bulletin of the Seismological Society of America*, **96**, 2383–2397.

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

Oh, J.-W., and T. Alkhalifah, 2016, The scattering potential of partial derivative wavefields in 3-D elastic orthorhombic media: an inversion prospective: *Geophysical Journal International*, **206**, 1740–1760.

- Operto, S., Y. Gholami, V. Prieux, A. Ribodetti, R. Brossier, L. Metivier, and J. Virieux, 2013, A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice: *The Leading Edge*, **32**, 1040–1054.
- Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: *Geophysical Journal International*, **167**, 495–503.
- Pratt, R. G., C. Shin, and G. Hick, 1998, Gauss–newton and full newton methods in frequency–space seismic waveform inversion: *Geophysical Journal International*, **133**, 341–362.
- Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013, Multiparameter full waveform inversion of multicomponent ocean-bottom-cable data from the valhall field. part 2: imaging compressive-wave and shear-wave velocities: *Geophysical Journal International*, ggt178.
- Raknes, E. B., and W. Weibull, 2016, Combining wave-equation migration velocity analysis and full-waveform inversion for improved 3D elastic parameter estimation, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists.
- Sears, T., S. Singh, and P. Barton, 2008, Elastic full waveform inversion of multi-component OBC seismic data: *Geophysical Prospecting*, **56**, 843–862.
- Shipp, R. M., and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data: *Geophysical Journal International*, **151**, 325–344.
- Sun, D., and W. W. Symes, 2012, Waveform inversion via nonlinear differential semblance optimization: Presented at the SEG Technical Program Expanded Abstracts 2012, Society of Exploration Geophysicists.
- Symes, W. W., 2008, Approximate linearized inversion by optimal scaling of prestack depth migration: *Geophysics*, **73**, R23–R35.

- Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: *Geophysics*, **49**, 1259–1266.
- , 1986, A strategy for nonlinear elastic inversion of seismic reflection data: *Geophysics*, **51**, 1893–1903.
- van Leeuwen, T., and W. A. Mulder, 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: *Geophysical Journal International*, **182**, 1383–1394.
- Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: *Geophysics*, **74**, WCC1–WCC26.
- Wang, C., W. Weibull, J. Cheng, and B. Arntsen, 2017a, Automatic shear-wave velocity analysis with elastic reverse time migration: 79th EAGE Conference and Exhibition 2017, Expanded Abstracts.
- Wang, H., S. C. Singh, F. Audebert, and H. Calandra, 2015a, Inversion of seismic refraction and reflection data for building long-wavelength velocity models: *Geophysics*, **80**, R81–R93.
- Wang, T., and J. Cheng, 2017, Elastic full waveform inversion based on mode decomposition: the approach and mechanism: *Geophysical Journal International*, **209**, 606–622.
- Wang, T., J. Cheng, Q. Guo, and C. Wang, 2017b, Elastic wave-equation reflection traveltime inversion using dynamic warping and wave mode decomposition: 79th EAGE Conference and Exhibition 2017, Expanded Abstracts.
- Wang, T., J. Cheng, and C. Wang, 2015b, Elastic wave mode decoupling for full waveform inversion: 77th EAGE Conference and Exhibition 2015, Expanded Abstracts.
- Wang, T., W. Kang, and J. Cheng, 2014, Migration velocity model building using local angle domain nonlinear tomography: Presented at the Beijing 2014 International Geophysical Conference and Exposition, Beijing, China, 21-24 April 2014, Society of Exploration

Geophysicists.

Woodward, M. J., 1992, Wave-equation tomography: *Geophysics*, **57**, 15–26.

Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: *Geophysics*, **73**, VE5–VE11.

Wu, Z., and T. Alkhalifah, 2015, Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion: *Geophysics*, **80**, R317–R329.

Xu, S., D. Wang, F. Chen, G. Lambare, and Y. Zhang, 2012, Inversion on reflected seismic wave: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–7.

Zhou, W., R. Brossier, S. Operto, and J. Virieux, 2015, Full waveform inversion of diving & reflected waves for velocity model building with impedance inversion based on scale separation: *Geophysical Journal International*, **202**, 1535–1554.