

Elastic wave-equation-based reflection kernel analysis and traveltime inversion using wave mode decomposition

T.F. Wang¹, J.B Cheng^{1,2,*}, Q. Guo³ and C.L. Wang¹

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 Department of Physical Science and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

* Corresponding author: Jiubing Cheng, Shanghai, China. E-mail: cjb1206@tongji.edu.cn

SUMMARY

Elastic reflection waveform inversion (ERWI) utilizes 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 through elastic wave-equation reflections travelttime inversion (ERTI) would be effective and robust since travelttime information relates to the background model more linearly. The wave mode decomposition, both on the recording surface and in the underground space, are important for ERTI. On the one hand, P/S separation of multicomponent seismograms on the surface provides individual P or S wave data residuals. Thus, we implement the ERTI using the L_2 norm of the isolated P or S wave travelttime residuals extracted by the dynamic image warping (DIW) as objective function. On the other hand, the underground spatial wave mode decomposition provides separated wavefields to pre-

condition the kernels or gradients. However, the reflection kernels in elastic media are complicated and difficult to use, especially when calculating the gradient of S-wave velocity. The investigation of reflection kernels show that mode decomposition can suppress the artifacts in gradient calculation. Accordingly, 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 recovered V_p . Numerical example of Sigsbee2A model validates the effectiveness of the algorithms and strategies for ERTI.

Key words: Inverse theory, Waveform inversion, Elasticity, Seismic tomography

1 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 & 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 & 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). Waveform inversion provides high-resolution model estimation of the elastic properties, but it suffers 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 unavailable 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 & Alkhalifah 2016).

For 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(Stork 1992; Chavent et al. 1994; Clement et al. 2001; Symes 2008b; Xu et al. 2012). This process can be implemented in the

image domain or the data domain. Actually, the image-domain ray-based tomography (Stork 1992; 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. To overcome the limit of ray theory, people have made great efforts to develop the approaches of reflection inversion that employ waveform or travelttime information based on wave-equation (Xu et al. 2012; Ma & Hale 2013; Wu & 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 subsurface image gathers (Symes 2008a; Almomin & Biondi 2012; Sun & Symes 2012; Biondi & Almomin 2013). Raknes & Weibull (2016) developed an 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. Recently, Zhou et al. (2015) proposed a joint FWI method that combines the diving and reflected waves to utilize both RWI and the conventional FWI. Wu & Alkhalifah (2015) developed an RWI scheme to simultaneously invert the background velocity and the perturbation in acoustic media, and recently extended to elastic case by Guo & Alkhalifah (2016). 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 one.

Compared with the waveform information, travelttime is more sensitive and linearly related to the low-wavenumber components of the model. Therefore, travelttime inversion will be more robust and helpful to build good initial models for conventional FWI (Wang et al. 2014). Ma & Hale (2013) introduced a wave equation reflected travelttime inversion method based on 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, travelttime shifts of a particular wave mode 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 mul-

ticomponent seismograms directly. In addition, since the multi-parameter trade-offs will increase the nonlinearity of inversion, more hierarchical strategies should be considered to deal with this problem. Wang & Cheng (2017) preconditioned the gradients of EFWI through wave mode decomposition to mitigate parameter trade-offs and explained that this precondition employs the off-diagonal Hessian blocks when inverting V_s . As a natural tool to obtain the separated data subsets, the wave mode decomposition similarly has the potential to precondition the elastic wave-equation reflection traveltime inversion (ERTI) with more flexible hierarchical strategies.

In this paper, we will exploit 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 kernel 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 travelttime residuals of PP and PS reflections through DIW. Based on the analysis of elastic reflection kernels and the separated travelttime residuals, we design a two-stage workflow to implement the ERTI method, in which the travelttime of PP reflections are firstly used to recover the background V_p model and then the travelttime of PS reflections are used to recover the background V_s model. According to our investigation, it is not suitable to estimate the time shifts of PS reflections at near offset if we use the PS image to predict the corresponding reflection during demigration. Instead, we recommend to use PP image to do that. Moreover, we precondition the V_s gradient through the spatial wave mode decomposition. Finally, the numerical example of Sigsee2A model proves the robustness and validity of our ERTI method.

2 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 taken as the demigrated reflection data using the image perturbation c_{ijkl}^1 obtained from reverse time migration (RTM) or other imaging method. In ERTI, we aim to minimize the travelttime differences (or time shifts) between the observed data \mathbf{d}^o

and the calculated data \mathbf{d}^c , using the following objective function:

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

where the time differences $\tau(\mathbf{x}_r, t)$ can be extracted through DIW (The readers can refer to Hale (2013) for detail).

After the derivation in Appendix A using the adjoint state method, 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, representing the forward background wavefields and the perturbed wavefields, respectively. While ψ_i and $\hat{\psi}_i$ are the adjoint state variables, representing the back-propagated background wavefields and the perturbed wavefields, 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

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

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 of equation (4), the two terms indicate two cross-correlations that represent 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 get the gradients in terms of P- and S- wave velocities through the chain rule, namely:

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

3 ELASTIC BORN REFLECTION KERNELS

Reflection travelttime 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. 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).

In elastic case, due to the complex wave phenomena, the four wavefields in eq. (8) contain both P and S waves, then the cross-correlations between different wave mode conversions will generates cross-talks in the kernel (Wang et al. 2017b). Therefore, the wavepath of elastic reflections will be far more complicated than that in acoustic case. 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. For example, K^{PS} is the cross-correlation of P-mode forward wavefield and S-mode adjoint wavefields. So the M or N is the mode type of the wavefields involved in cross-correlation but not the mode type of shot data. 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) is consistent with the forward wavefields, the cross-correlation will produce the low-wavenumber wavepath. Otherwise, the cross-correlation should be a high-wavenumber migration impulse.

For instance, 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, we can investigate the reflection kernel using the waveform data residual as the adjoint source for back-propagation. In the first test, we place a single V_p reflector in the homogeneous background (Figs 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 Figs 2c and d, the reflection kernel consists of two “rabbit-ears”, 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.

In the second test, we use the V_s reflector (Figs 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 Fig. 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 (Fig. 3d), due to mode conversions, multi-wavepaths overlapping with each other make it much more difficult to calculate and exploit 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 Fig. 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. Fig. 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 the calculation. Note that, when PS data are back-propogated, 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 (Ravasi & Curtis 2013), another migration impulse can be seen at the receiver side. Accordingly, the non-physical conversions related to 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 Figs 6c and d. Then, the isolation of PP and PS data injection provides four different components, respectively (Figs 6e to h). Figs 6e and f show the decomposition of the kernel in Fig. 6c while Figs 6g and h corresponds to the decomposition of the kernel in Fig. 6d. These four components can be further decomposed with the help of mode decomposition. For simplicity, 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, P and S wavefields are generated simultaneously in ψ when PS data is injected. 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 Figs 7d and e 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 know that the cross-correlation between different wave modes cause cross-talks in gradient calculation. The injection of separated PP reflection data helps to reduce the artifacts in V_p gradients, so does the injection of separated PS reflection data for V_s gradient calculation. 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.

4 WORKFLOW OF ELASTIC WERTI

In elastic case, it is common to observe that different mode-conversions, e.g. 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 $\tau(\mathbf{x}_r, t)$ would be inaccurate due to these singularities. Moreover, 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, it is relatively easier to estimate the time shifts through DIW for PP and PS reflection, respectively. Accordingly, we implement the ERTI through a two-stage workflow, i.e. estimating V_p using PP reflections followed by estimating V_s using PS reflections.

4.1 Stage I: ERTI of PP reflection

In this stage, we use traveltimes of PP reflection to recover the low-to-intermediate wavenumbers of the 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 take the two-way traveltime of zero-offset as an invariant during reflection inversion, such as reflection traveltime 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 fail to ensure a correct zero offset traveltime in the demigration when velocity anomaly is complex. A better method would be the demigration with the extended image (Weibull & Arntsen 2014; Hou & Symes 2015; Guo et al. 2017), but it is out of the scope of this paper. Besides, since the traveltime rather than the amplitude is fitted in ERTI, ERTM (instead of

its least-squares counterpart) is sufficient to obtain the image perturbation for the demigration. Thus, the objective function can be written as:

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

where \mathbf{d}_{pp}^c and \mathbf{d}_{pp}^o are synthesized and observed PP reflection after 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 travelttime inversion for V_p in elastic media is similar to the acoustic case except that the isolation of PS component is required 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.

4.2 Stage II: ERTI of PS reflection

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

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

where \mathbf{d}_{ps}^c and \mathbf{d}_{ps}^o are synthesized and observed PS reflections, respectively. After the first 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:

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 as the virtual source to generate the PS reflections.

Unfortunately, the predicted PS reflection with method **I** are inconsistent with the error of V_s model at near offset. In Appendix C, we illustrate this inconsistency for inversion through a single reflector example with map migration and demigration under the high-frequency approximation. Therefore, we recommend to use method **II**. Nonetheless, the amplitude of the predicted PS will be inaccurate when using method **II**, even if δV_p is obtained by the least-squares ERTM. Fortunately,

the inaccurate amplitude of PS prediction has little effect on ERTI because ERTI only cares the kinematic information. Note that the traveltime residuals of the PS reflection may be large in the deeper part (Fig. A4). 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 while the deep part is inverted 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 , the wave mode decomposition is applied to calculate $K_{V_s}^{SS}$, namely:

$$\frac{\partial E_{ps}}{\partial V_s} = -2\rho V_s \int \left(\frac{\partial \hat{u}_i^S}{\partial x_j} \frac{\partial \psi_k^S}{\partial x_l} \right) (\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}). \quad (12)$$

This is similar to the gradient preconditioning for EFWI proposed by Wang & Cheng (2017). It can mitigate parameter trade-offs and suppress artifacts for the V_s inversion.

5 NUMERICAL EXAMPLE

We select a part of the Sigsbee2A model (Figs 8a and b) to test the inversion algorithm and strategy. The V_s model is generated through V_p model with a constant V_s -to- V_p ratio of 0.66. The initial model for ERTI are shown in Figs 8c 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. 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 are 16m and 1.2ms, respectively.

Fig. 9 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 and PS image are not well recovered in the ERTM. Especially, the diffractions collapse in PP image and faults almost can not be seen 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 are acceptable due to the exsistence of diving wave and low-frequency data. However, the inversion of deep part suffers 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. After

40 iterations for each stage, ERTI provides a good recovery of the low-to-intermediate components of V_p and V_s model. Figs 10 shows the inverted V_p and V_s model of ERTI. In addition, we apply the structure-oriented smooth filter to the gradients to regularize the model update based on the structure tensor (Hale 2009; Ma et al. 2010; Williamson et al. 2011). Note, the structure-oriented smooth filter is an appropriate regularization to stabilize and accelerate the convergence. In each iteration of ERTI, we constrain the gradients of ERTI with the seismic image obtained by ERTM. Fig. 11a 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 improper update during inversion. After the structure-oriented smooth, the gradient (Fig. 11b) 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 Figs 12c 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 diffractions are migrated as well. In addtion, Fig. 13 shows the comparison of the demigrated PP and PS reflections using the initial , inverted and 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 of EFWI will be mitigated greatly if using the starting model in Fig. 10. Fig. 14 shows the inverted results of conventional EFWI with the same four-stage hierarchical strategy from low to high frequency. Comparing with Fig. 9, EFWI with the new starting model improves quite a lot, especially in the deep part. On the bottom of model, the background components are not well recovered due to the lack of reflection coverage. That's why EFWI cannot provide good recovery of this part as well. The vertical profiles at 1.4km and 3km validate the effectiveness of our workflow.

Usually, the low-frequency components in the data and the low-to-intermediate wavenumber in the initial model are two key factors to mitigate the cycle-skipping problem in FWI. In field data, the low frequency components are difficult to obtain and also more easily contaminated by the noises. Therefore, good initial models are more important to reduce the nonlinearity caused by 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 thresholds 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 Fig. 16, even starting from 5Hz, EFWI still provides the acceptable inverted results like the no-low-frequency-cut test (Fig. 14). When the starting freqeucy 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.

6 CONCLUSIONS

We extend the wave equation reflection traveltimes inversion towards elastic case to build the low-wavenumber component of elastic model. ERTI can recover the low-wavenumber components effectively, because time misfits are more sensitive and linearly related to the low-wavenumber model perturbation. 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 components 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 cross-talks and artifacts. The Sigsbee2A model example shows that even starting with a bad initial model, the two-stage ERTI can provide reliable starting model for conventional EFWI.

7 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).

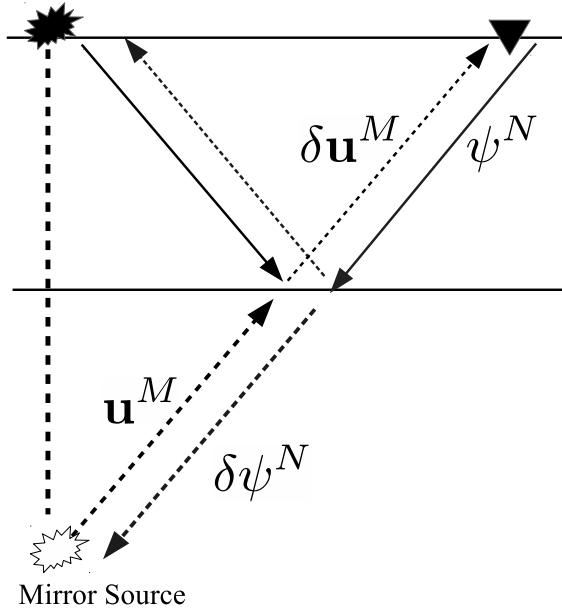


Figure 1. Schematic illustration of the reflection kernels K^{MN} . The cross-correlation can be regarded as the conventional FWI kernel between the mirror source and the receiver. The kernel can be a migration impulse or a transimission-type wavepath.

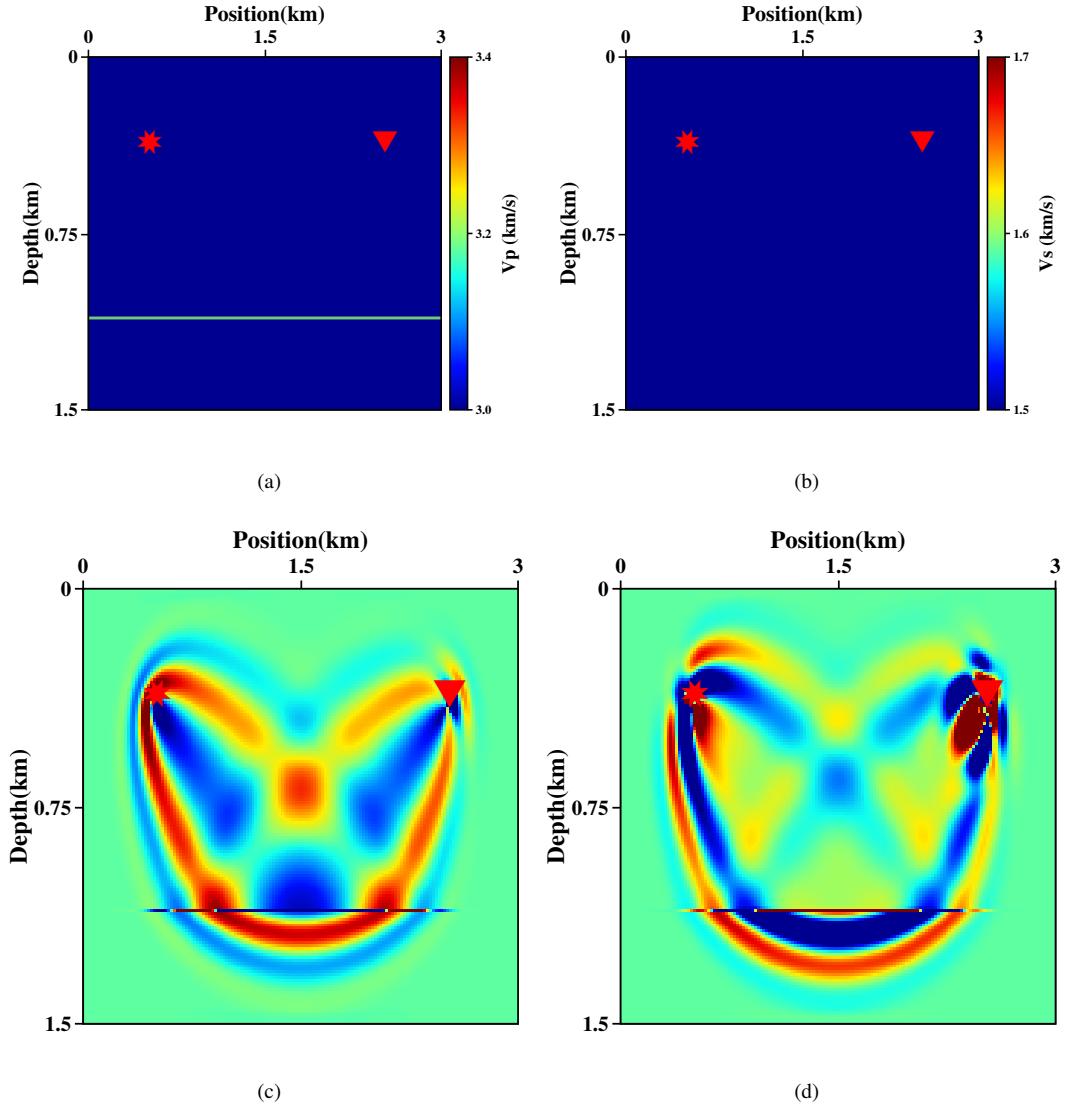


Figure 2. Kernels with single reflector in V_p model. (a) V_p model, (b) V_s model, (c) K_{V_p} , (d) K_{V_s} .

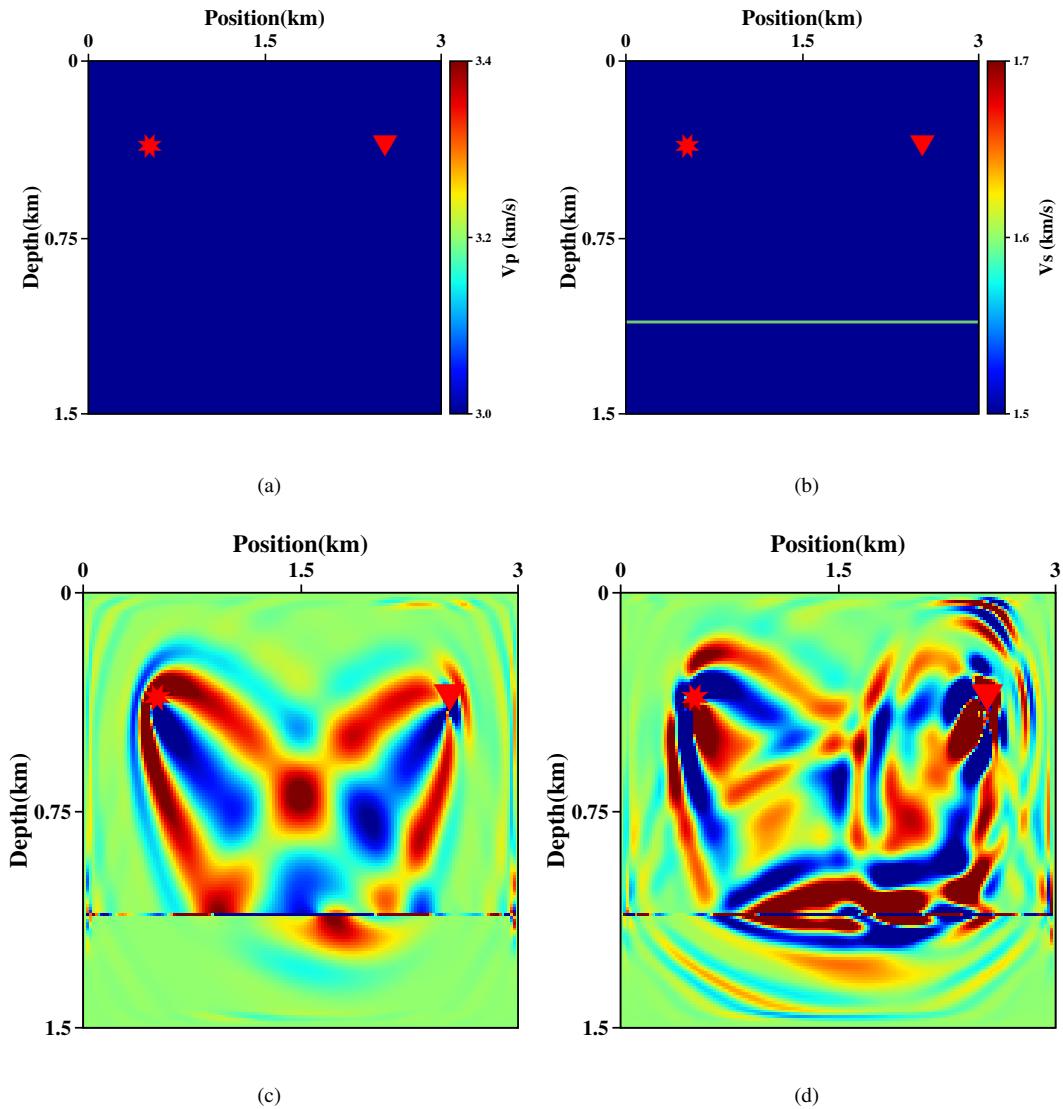


Figure 3. Kernels with single reflector in V_s model. (a) V_p model, (b) V_s model, (c) K_{V_p} , (d) K_{V_s} .

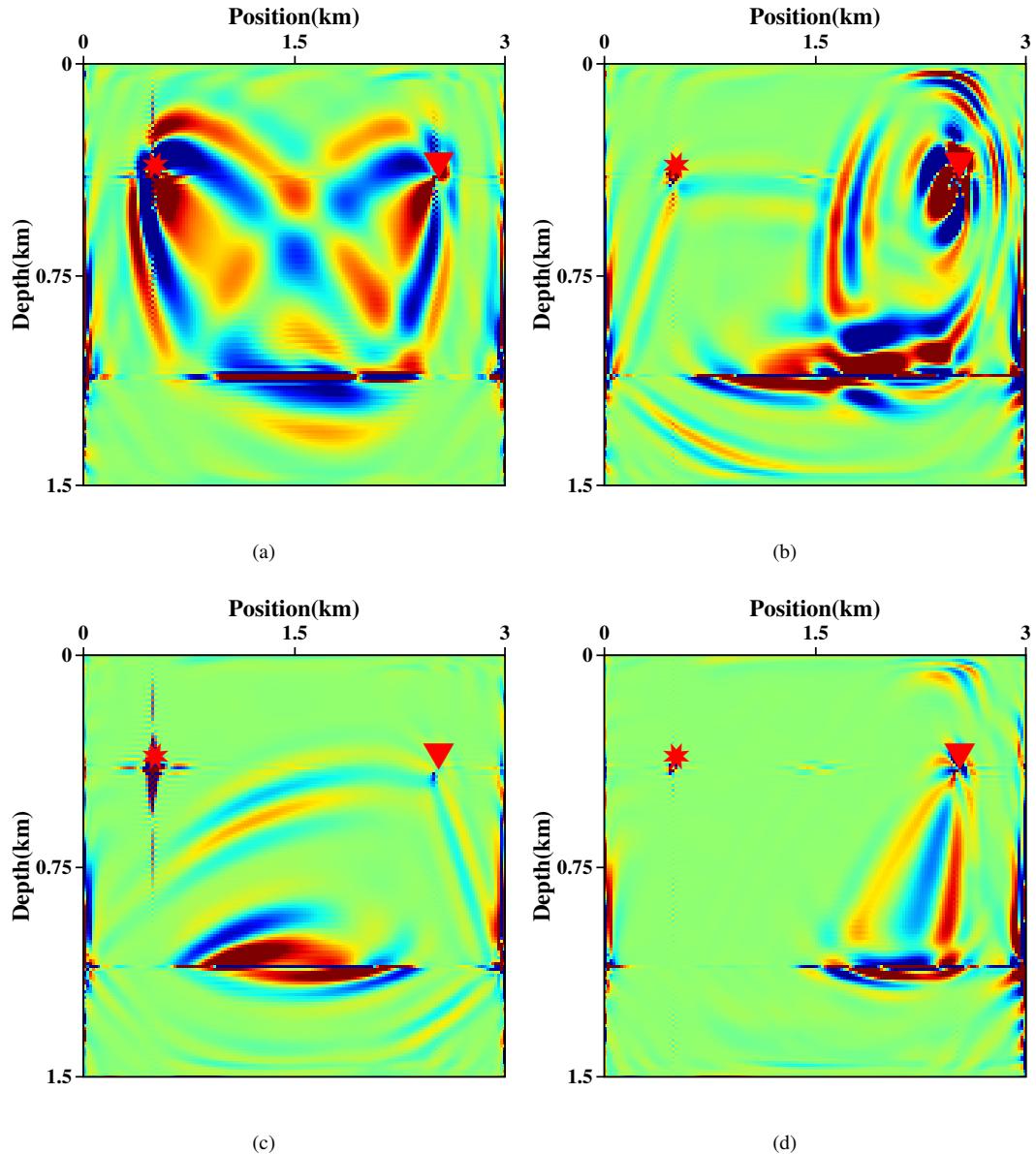


Figure 4. Four components of K_{V_s} . Here we only decompose the subsurface wavefields as in eq. (9) but do not separate the data for injection. (a) $K_{V_s}^{PP}$, (b) $K_{V_s}^{PS}$, (c) $K_{V_s}^{SP}$, (d) $K_{V_s}^{SS}$.

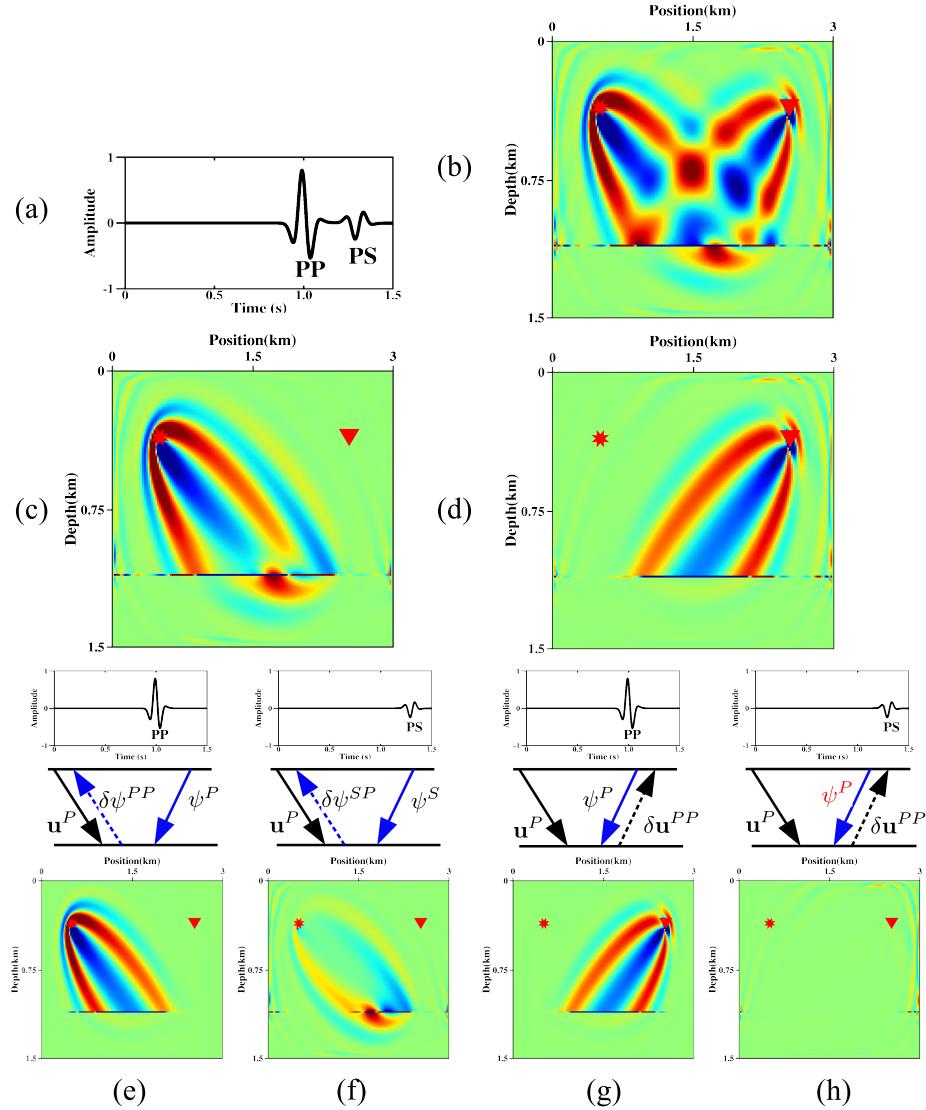


Figure 5. Further decomposition of K_{V_p} for the second test by injecting the PP and PS event separately as adjoint source and decomposing the wavefields with different manners. (a) Original adjoint source for back-propagation, (b) K_{V_p} , (c) Source side of K_{V_p} , (d) Receiver side of K_{V_p} . (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)

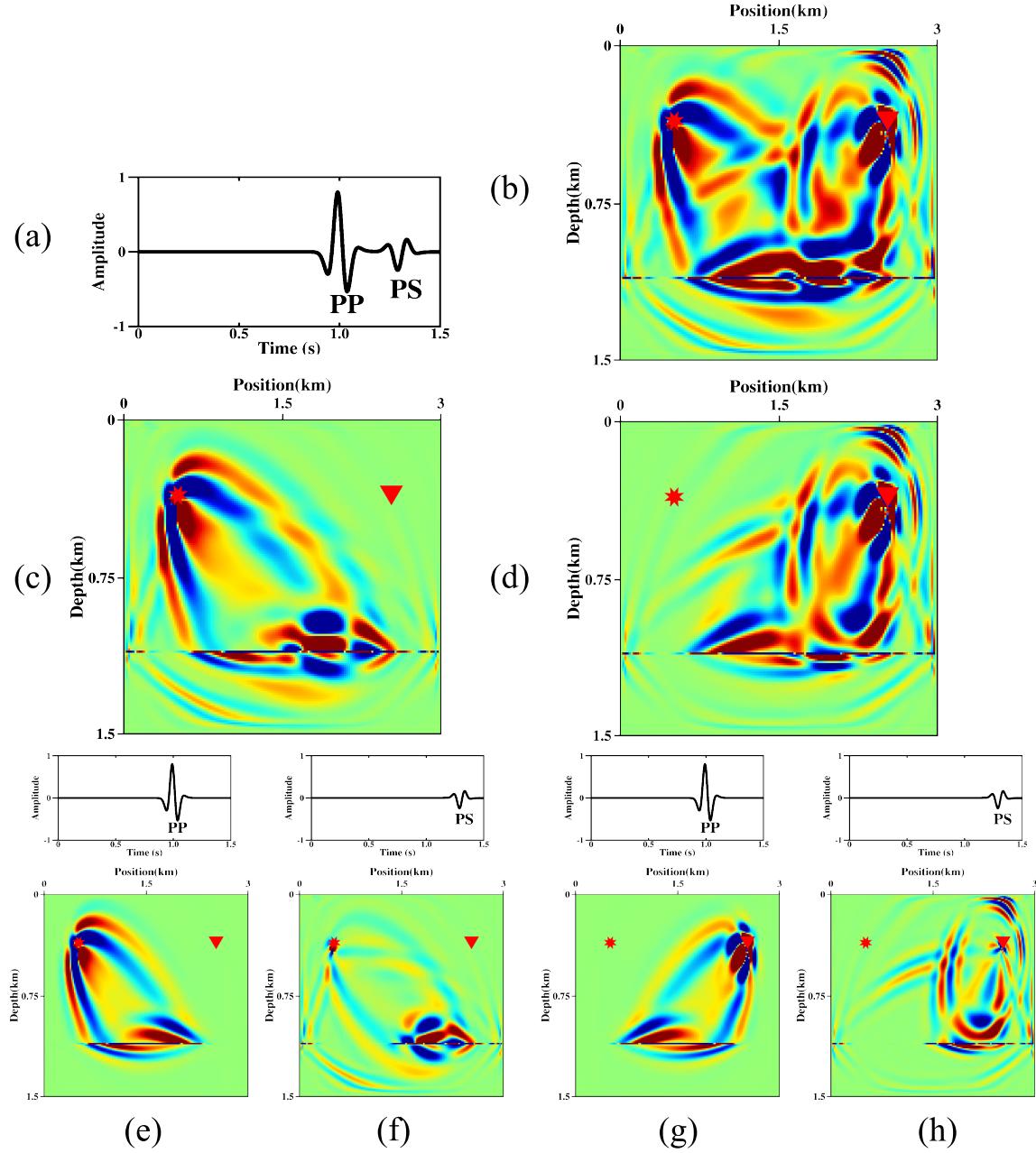


Figure 6. Further decomposition of K_{V_s} for the second test. Here we only separate PP and PS data for injection. (a): Original adjoint source , (b) K_{V_s} , (c) Source term of K_{V_s} , (d) Receiver term of K_{V_s} . (e)-(h): The first line denotes the adjoint source for injection. The second line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

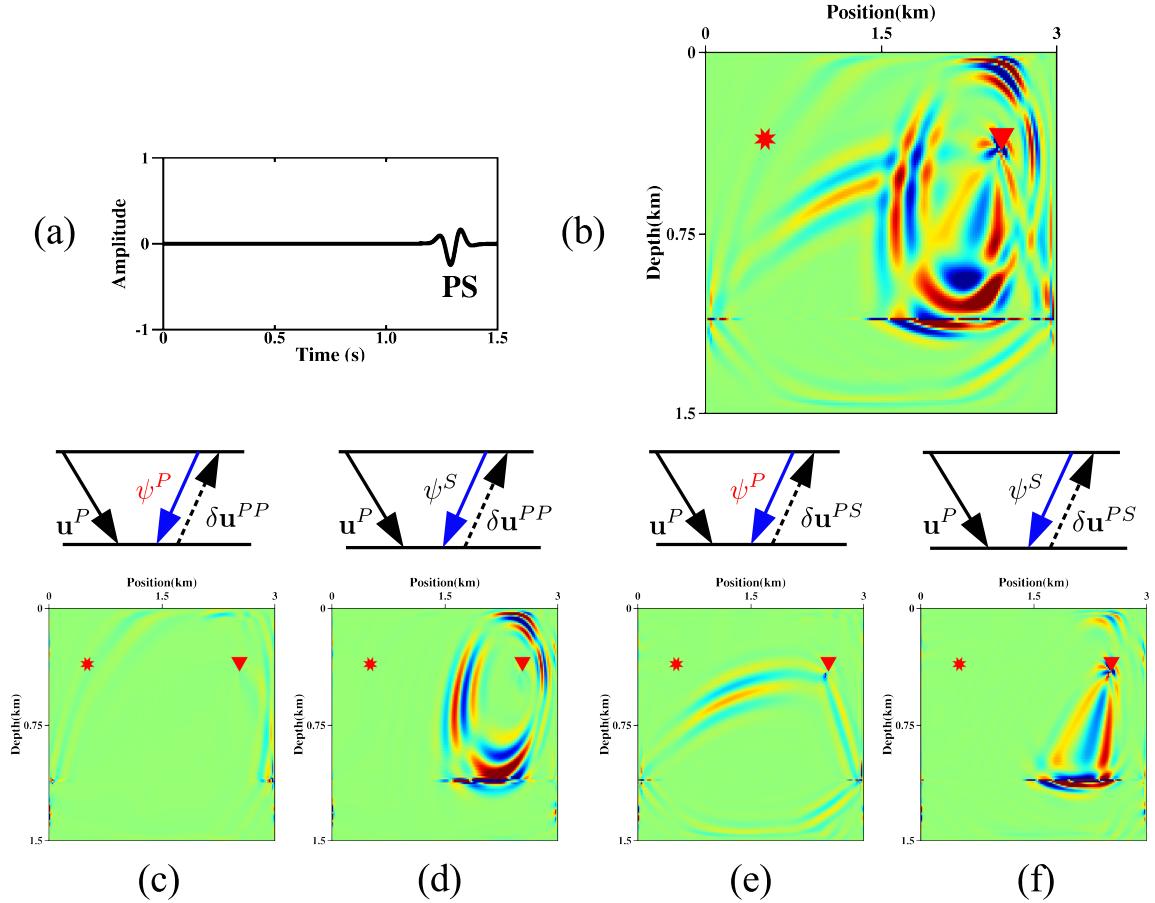


Figure 7. Further decomposed components of Fig 6h using spatial wave mode decomposition when only injecting PS data. (a) The adjoint source for injection, (b) Receiver term of K_{V_s} using (a) as adjoint source, (c)-(f) The first line denotes the manner and wavefield type for cross-correlation. The second line denotes the corresponding kernels. The ψ^P in red denotes the non-physical conversion when injecting PS data.

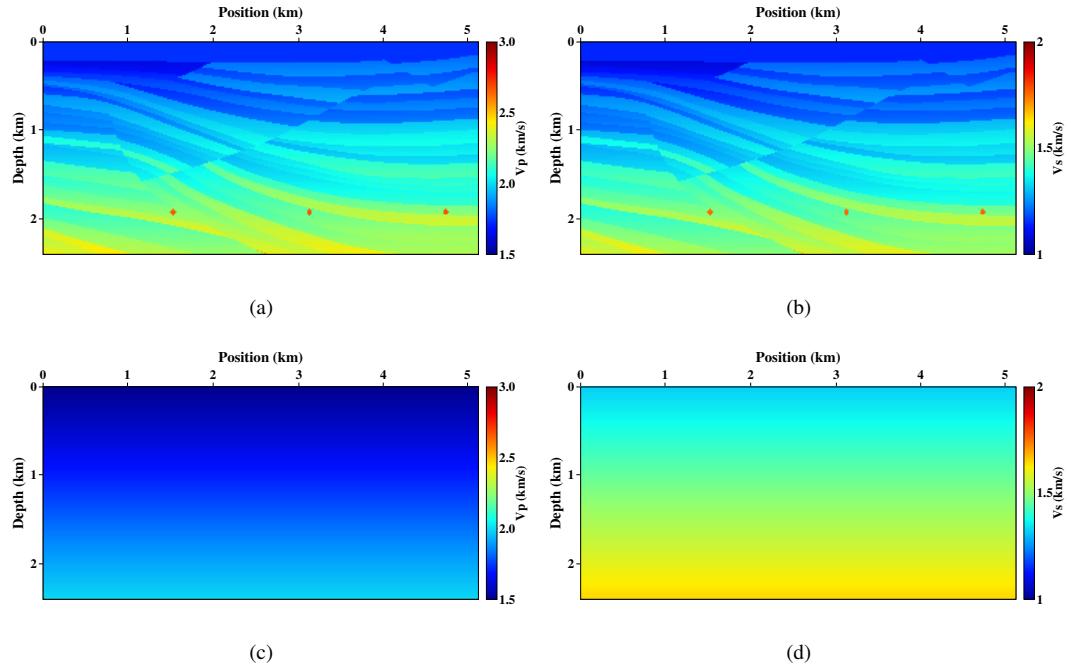


Figure 8. Sigbee2A model example. On the top are true models of V_p (a) and V_s (b). On the bottom are initial models of V_p (c) and V_s (d) linearly increasing with depth.

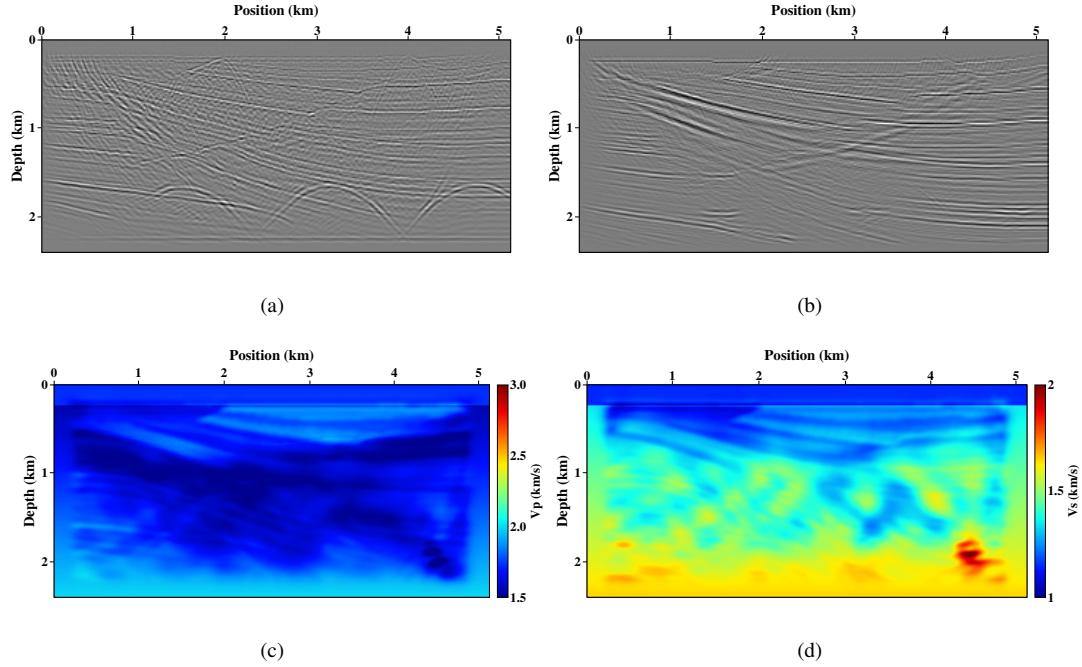


Figure 9. The results of ERTM and EFWI using initial model: (a) and (b) are PP and PS image of ERTM with near offset data, (c) and (d) are inverted V_p and V_s with EFWI.

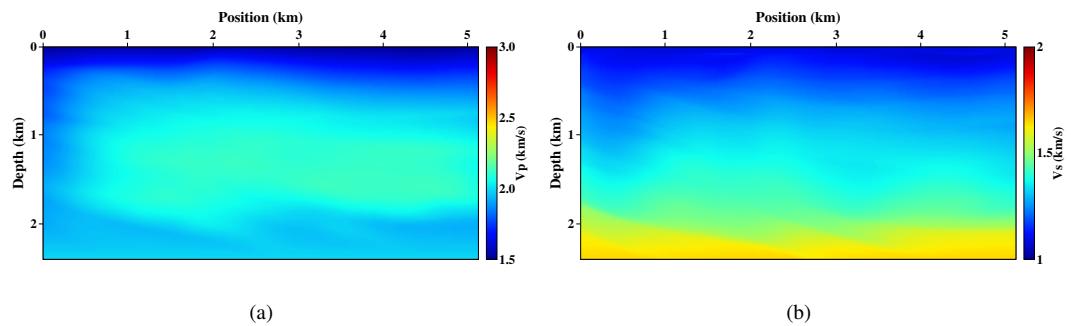


Figure 10. Inverted results of ERTI: (a) V_p , (b) V_s .

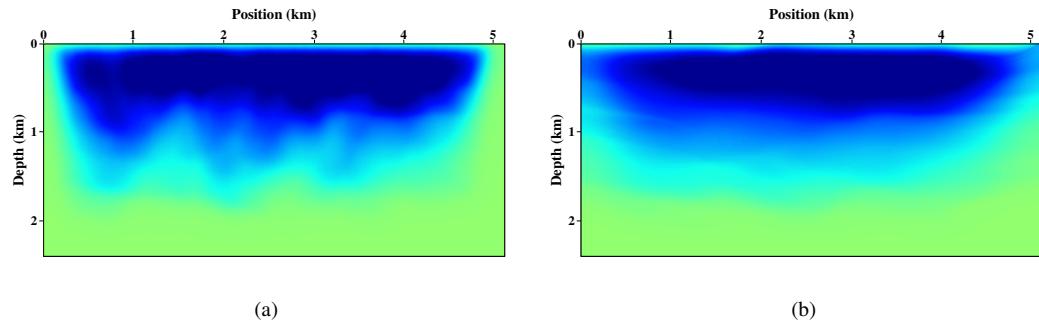


Figure 11. The comparison of gradients without (a) and with (b) the structure-oriented constrain.

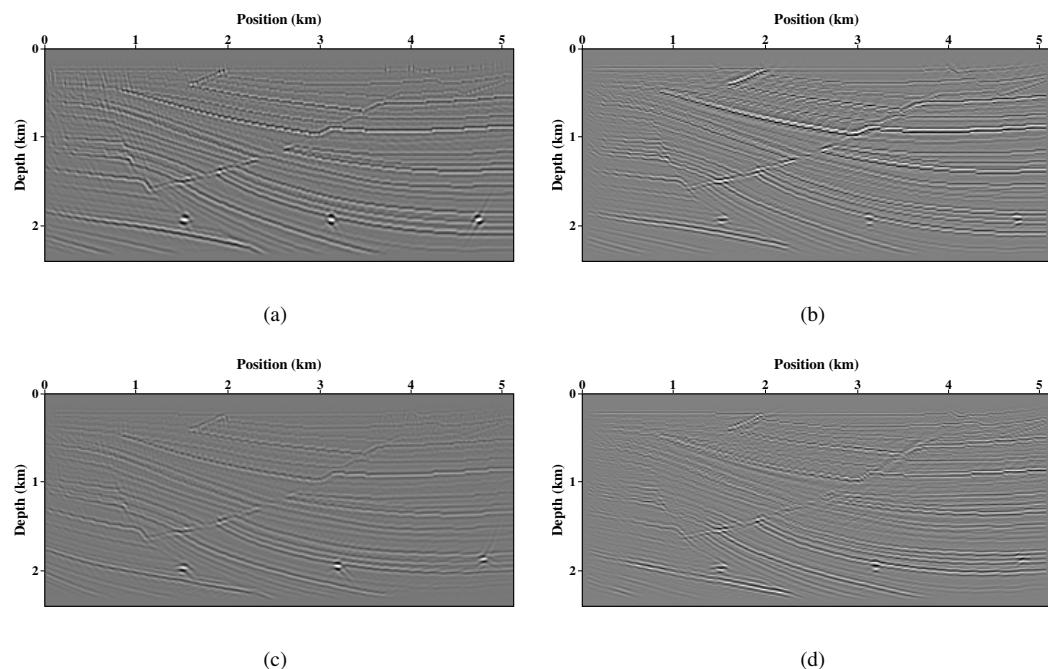


Figure 12. ERTM results using the true model (a, b) and inverted model (c, d). (a) and (c) are the PP image(b) and (d) are the PS image

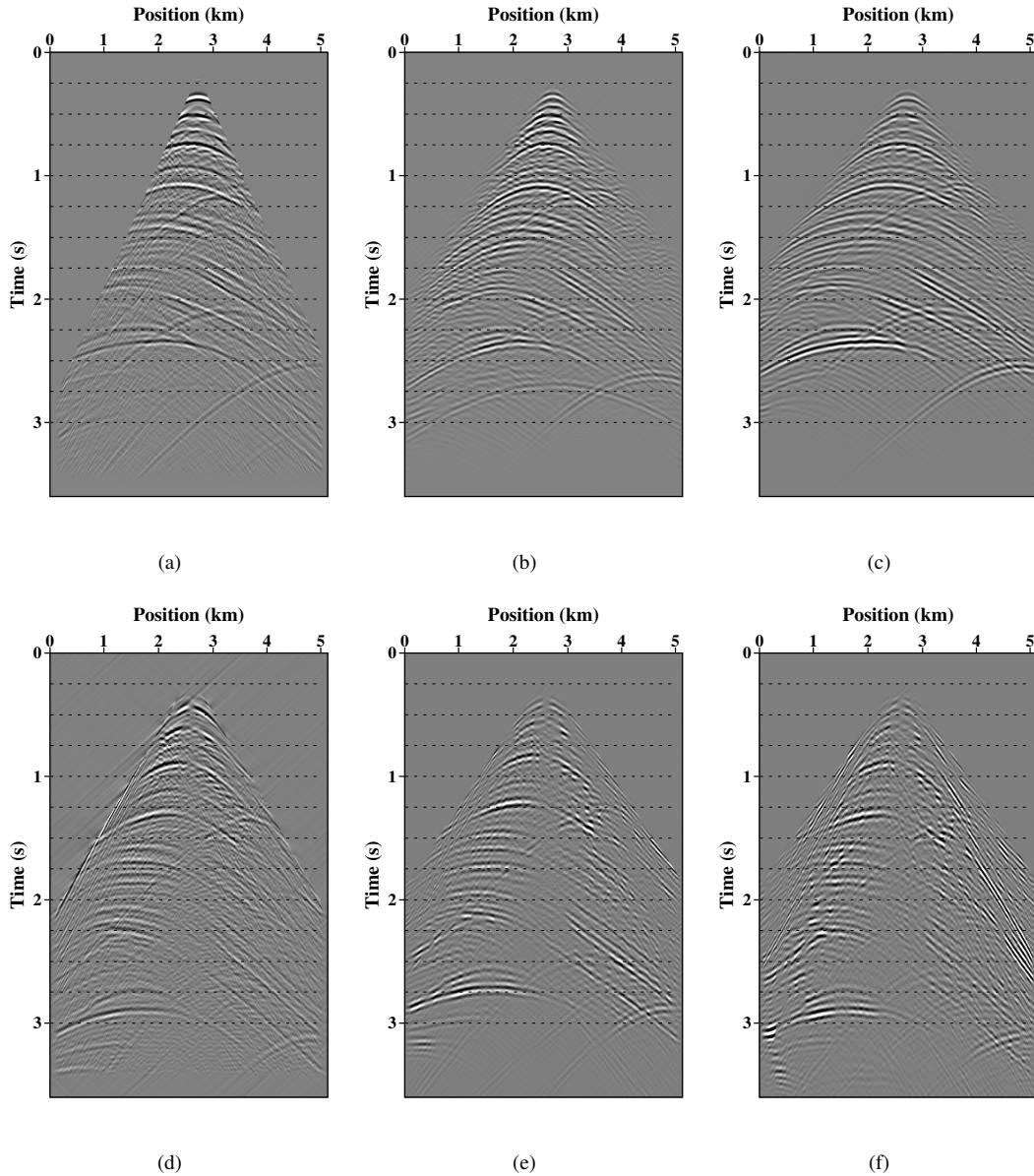


Figure 13. Comparison of the observed and the demigrated reflection data using initial model and the inverted model. The first row are the separated PP reflection, while the second row are the separated PS reflection. The left, middle and right column are the observed reflection data, the demigrated reflection data with initial model and the demigrated data with inverted model, respectively.

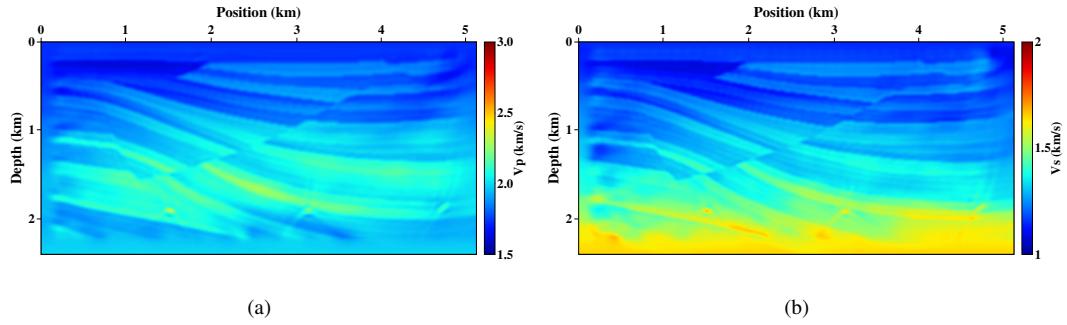


Figure 14. EFWI results using the ERTI model as starting model. (a) V_p , (b) V_s .

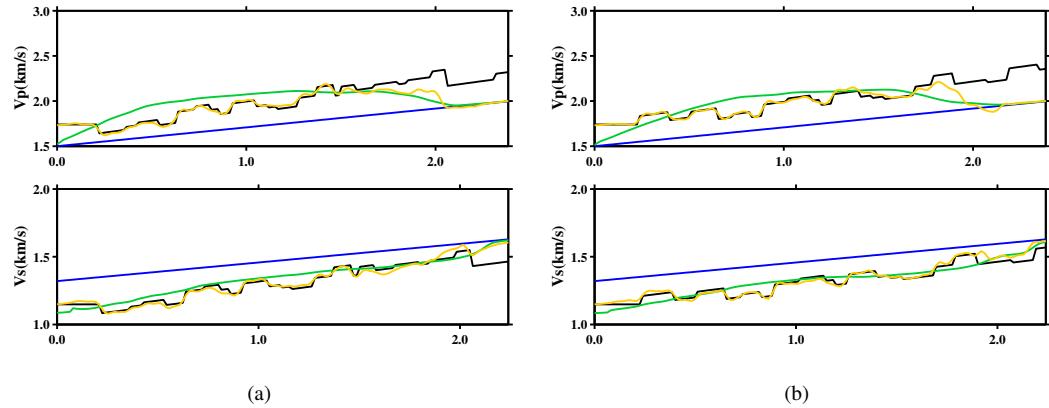


Figure 15. Vertical profile of ERTI and EFWI model at 1.4km (a) and 3km (b). The black and blue lines denote the true and initial model. The green and yellow denote the ERTI and EFWI model.

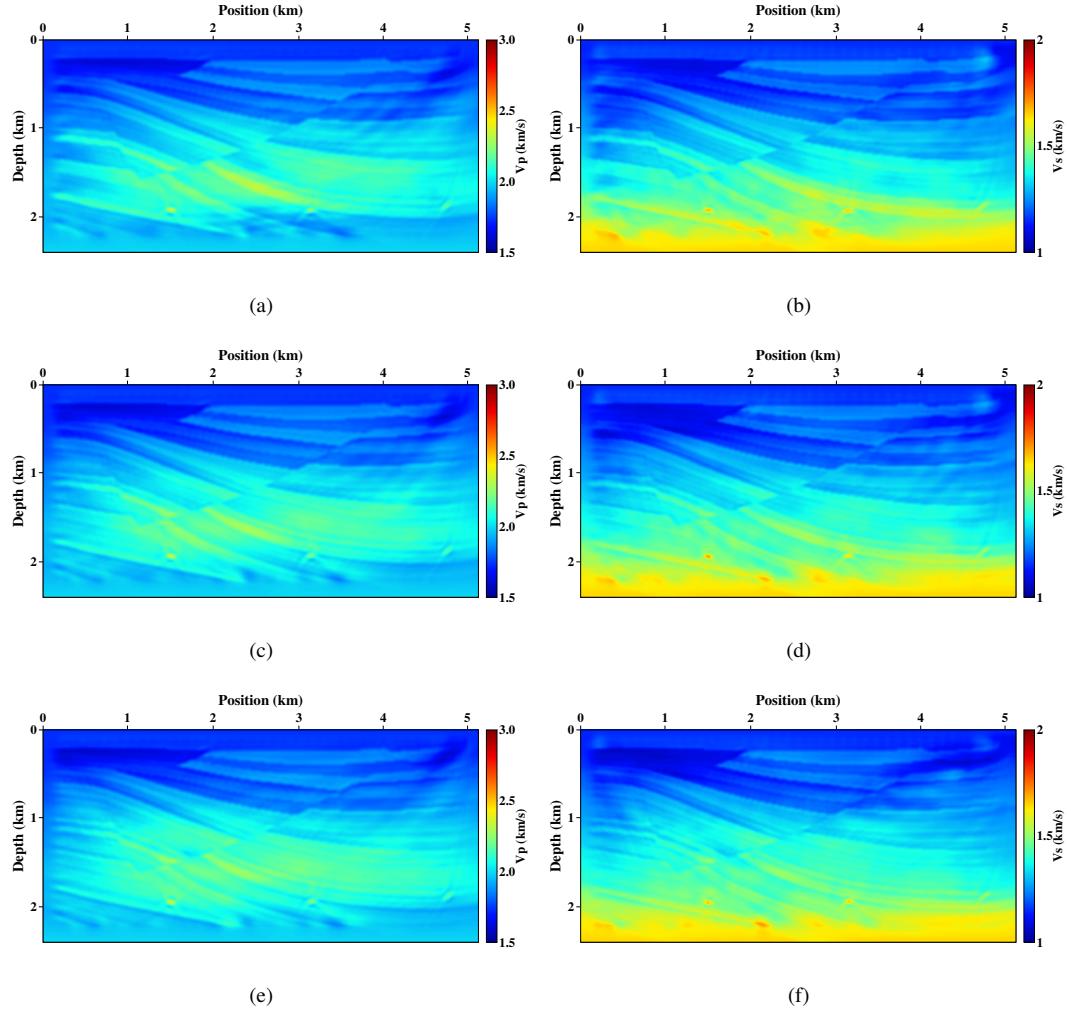


Figure 16. EFWI results with different starting frequency. (a), (c), (e) are V_p , (b), (d), (f) are V_s . The starting frequency are 3Hz, 5Hz and 7Hz from top to bottom row, respectively.

REFERENCES

- Almomin, A. & Biondi, B., 2012. Tomographic full waveform inversion: Practical and computationally feasible approach, in *SEG Technical Program Expanded Abstracts 2012*, Society of Exploration Geophysicists.
- Biondi, B. & Almomin, A., 2013. Tomographic full-waveform inversion (TFWI) by combining FWI and wave-equation migration velocity analysis, *The Leading Edge*, **32**(9), 1074–1080.
- 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.
- Chavent, G., Clément, F., & Gómez, S., 1994. Automatic determination of velocities via migration-based traveltimes waveform inversion: A synthetic data example, in *SEG Technical Program Expanded Abstracts 1994*, Society of Exploration Geophysicists.
- Chi, B., Dong, L., & Liu, Y., 2015. Correlation-based reflection full-waveform inversion, *Geophysics*, **80**(4), R189–R202.
- Clement, F., Chavent, G., & Gomez, S., 2001. Migration-based traveltimes waveform inversion of 2-d simple structures: A synthetic example, *Geophysics*, **66**, 845–860.
- Guo, Q. & Alkhalifah, T., 2016. *A nonlinear approach of elastic reflection waveform inversion*, pp. 1421–1425, Society of Exploration Geophysicists.
- Guo, Q., Alkhalifah, T., & Wu, Z., 2017. Velocity building by reflection waveform inversion without cycle-skipping, in *79th EAGE Conference and Exhibition 2017*, EAGE Publications.
- Hale, D., 2009. Structure-oriented smoothing and semblance, *CWP report*.
- Hale, D., 2013. Dynamic warping of seismic images, *GEOPHYSICS*, **78**(2), S105–S115.
- Hou, J. & Symes, W. W., 2015. An approximate inverse to the extended born modeling operator, *GEOPHYSICS*, **80**(6), R331–R349.
- Jones, I., 2010. Tutorial: Velocity estimation via ray-based tomography, *First Break*, **28**(1719).
- Li, Z., Ma, X., Fu, C., Gu, B., & Liang, G., 2016. Frequency-wavenumber implementation for p- and s-wave separation from multi-component seismic data, *Exploration Geophysics*, **47**(1), 32.
- Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bulletin of the Seismological Society of America*, **96**(6), 2383–2397.
- Ma, Y. & Hale, D., 2013. Wave-equation reflection traveltimes inversion with dynamic warping and full waveform inversion, *Geophysics*, **78**(6), R223–R233.
- Ma, Y., Hale, D., Meng, Z. J., & Gong, B., 2010. *Full waveform inversion with imageguided gradient*, pp. 1003–1007.
- Oh, J.-W. & Alkhalifah, T., 2016. The scattering potential of partial derivative wavefields in 3-D elastic orthorhombic media: an inversion prospective, *Geophysical Journal International*, **206**(3), 1740–1760.
- Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Metivier, L., & Virieux, J., 2013. A guided tour of multiparameter full-waveform inversion with multicomponent data: From theory to practice, *The Leading Edge*, **32**(9), 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**(2), 495–503.
- Pratt, R. G., Shin, C., & Hick, G., 1998. Gauss–newton and full newton methods in frequency–space seismic waveform inversion, *Geophysical Journal International*, **133**(2), 341–362.
- Prieux, V., Brossier, R., Operto, S., & Virieux, J., 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*, p. ggt178.
- Raknes, E. B. & Weibull, W., 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.
- Ravasi, M. & Curtis, A., 2013. Nonlinear scattering based imaging in elastic media: Theory, theorems, and imaging conditions, *Geophysics*, **78**(3), S137–S155.
- Sears, T., Singh, S., & Barton, P., 2008. Elastic full waveform inversion of multi-component OBC seismic data, *Geophysical Prospecting*, **56**, 843–862.
- Shipp, R. M. & Singh, S. C., 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophysical Journal International*, **151**(2), 325–344.
- Stork, C., 1992. Reflection tomography in the postmigrated domain, *Geophysics*, **57**(5), 680–692.
- Sun, D. & Symes, W. W., 2012. Waveform inversion via nonlinear differential semblance optimization, in SEG Technical Program Expanded Abstracts 2012, Society of Exploration Geophysicists.
- Symes, W. W., 2008a. Approximate linearized inversion by optimal scaling of prestack depth migration, *Geophysics*, **73**(2), R23–R35.
- Symes, W. W., 2008b. Migration velocity analysis and waveform inversion, *Geophysical Prospecting*, **56**(6), 765–790.
- Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics*, **49**(8), 1259–1266.
- Tarantola, A., 1986. A strategy for nonlinear elastic inversion of seismic reflection data, *Geophysics*, **51**(10), 1893–1903.
- Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Wang, C., Weibull, W., Cheng, J., & Arntsen, B., 2017a. Automatic shear-wave velocity analysis with elastic reverse time migration, *79th EAGE Conference and Exhibition 2017, Expanded Abstracts*.
- Wang, H., Singh, S. C., Audebert, F., & Calandra, H., 2015a. Inversion of seismic refraction and reflection data for building long-wavelength velocity models, *Geophysics*, **80**(2), R81–R93.
- Wang, T. & Cheng, J., 2017. Elastic full waveform inversion based on mode decomposition: the approach and mechanism, *Geophysical Journal International*, **209**(2), 606–622.
- Wang, T., Kang, W., & Cheng, J., 2014. Migration velocity model building using local angle domain nonlinear tomography, in Beijing 2014 International Geophysical Conference and Exposition, Beijing, China, 21–24 April 2014, Society of Exploration Geophysicists.

- Wang, T., Cheng, J., & Wang, C., 2015b. Elastic wave mode decoupling for full waveform inversion, *77th EAGE Conference and Exhibition 2015, Expanded Abstracts*.
- Wang, T., Cheng, J., Guo, Q., & Wang, C., 2017b. Elastic wave-equation reflection traveltime inversion using dynamic warping and wave mode decomposition, *79th EAGE Conference and Exhibition 2017, Expanded Abstracts*.
- Weibull, W. W. & Arntsen, B., 2014. Reverse-time demigration using the extended-imaging condition, *GEO-PHYSICS*, **79**(3), WA97–WA105.
- Williamson, P., Atle, A., Fei, W., & Hale, D., 2011. Regularization of wave-equation migration velocity analysis by structure-oriented smoothing, in *SEG Technical Program Expanded Abstracts 2011*, Society of Exploration Geophysicists.
- Woodward, M. J., 1992. Wave-equation tomography, *Geophysics*, **57**(1), 15–26.
- Woodward, M. J., Nichols, D., Zdraveva, O., Whitfield, P., & Johns, T., 2008. A decade of tomography, *Geophysics*, **73**(5), VE5–VE11.
- Wu, Z. & Alkhalifah, T., 2015. Simultaneous inversion of the background velocity and the perturbation in full-waveform inversion, *Geophysics*, **80**(6), R317–R329.
- Xu, S., Wang, D., Chen, F., Lambare, G., & Zhang, Y., 2012. Inversion on reflected seismic wave, *82nd Annual International Meeting, SEG, Expanded Abstracts*, pp. 1–7.
- Zhou, W., Brossier, R., Operto, S., & Virieux, J., 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.

APPENDIX A: GRADIENT DERIVATION FOR ERTI WITH LAGRANGE MULTIPLIER

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

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

The optimization of eq.(A.1) satisfies:

$$\frac{\partial D}{\partial \tau} = \int_0^T \sum_r \boldsymbol{\alpha}(\mathbf{x}_r, t) dt = 0. \quad (\text{A.2})$$

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

$$\begin{aligned} \rho \partial_t^2 \mathbf{u} - \nabla \cdot (\mathbf{c}^0 : \nabla \mathbf{u}) &= \mathbf{f} \\ \rho \partial_t^2 \hat{\mathbf{u}} - \nabla \cdot (\mathbf{c}^0 : \nabla \hat{\mathbf{u}}) &= \nabla \cdot (\mathbf{c}^1 : \nabla \mathbf{u}), \end{aligned} \quad (\text{A.3})$$

where ∂_t is the time derivative. Therefore, using eqs (A.3) and (A.2) we can define the Lagrangian \mathcal{L} :

$$\begin{aligned} \mathcal{L} = & \frac{1}{2} \sum_r \int_0^T \boldsymbol{\tau}^2(\mathbf{x}_r, t) dt - \sum_r \int_0^T \boldsymbol{\mu}(\mathbf{x}_r, t) \boldsymbol{\alpha}(\mathbf{x}_r, t) dt \\ & - \int_0^T \int_{\Omega} \hat{\psi} [\rho \partial_t^2 \mathbf{u} - \nabla \cdot (\mathbf{c}^0 : \nabla \mathbf{u}) - \mathbf{f}] d^3 \mathbf{x} dt \\ & - \int_0^T \int_{\Omega} \psi [\rho \partial_t^2 \hat{\mathbf{u}} - \nabla \cdot (\mathbf{c}^0 : \nabla \hat{\mathbf{u}}) - \nabla \cdot (\mathbf{c}^1 : \nabla \mathbf{u})] d^3 \mathbf{x} dt, \end{aligned} \quad (\text{A.4})$$

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. Note, the predicted reflection data are the wavefields observed at the receivers, i.e., $\mathbf{d}^c(\mathbf{x}_r, t) = \delta(\mathbf{x} - \mathbf{x}_r) \delta \hat{\mathbf{u}}$. Here, we only consider the variation of \mathbf{c}^0 , then the perturbation of eq. (A.4) is:

$$\begin{aligned} \delta \mathcal{L} = & \int_0^T \sum_r [\boldsymbol{\tau}(\mathbf{x}_r, t) - \boldsymbol{\mu}(\mathbf{x}_r, t) \mathbf{h}(\mathbf{x}_r, t)] \cdot \delta \boldsymbol{\tau} dt \\ & + \int_0^T \int_{\Omega} \boldsymbol{\mu}(\mathbf{x}_r, t) \dot{\mathbf{d}}^o(\mathbf{x}_r, t + \boldsymbol{\tau}) \delta(\mathbf{x} - \mathbf{x}_r) \cdot \delta \hat{\mathbf{u}} d^3 \mathbf{x} dt \\ & + \int_0^T \int_{\Omega} [\hat{\psi} \nabla \cdot (\delta \mathbf{c}^0 : \nabla \mathbf{u}) + \psi \nabla \cdot (\delta \mathbf{c}^0 : \nabla \hat{\mathbf{u}})] d^3 \mathbf{x} dt \\ & - \int_0^T \int_{\Omega} \hat{\psi} [\rho \partial_t^2 \delta \mathbf{u} - \nabla \cdot (\mathbf{c}^0 : \nabla \delta \mathbf{u})] d^3 \mathbf{x} dt \\ & - \int_0^T \int_{\Omega} \psi [\rho \partial_t^2 \delta \hat{\mathbf{u}} - \nabla \cdot (\mathbf{c}^0 : \nabla \delta \hat{\mathbf{u}}) - \nabla \cdot (\mathbf{c}^1 : \nabla \delta \mathbf{u})] d^3 \mathbf{x} dt, \end{aligned} \quad (\text{A.5})$$

where $\mathbf{h}(\mathbf{x}_r, t) = \dot{d}_i^o(\mathbf{x}_r, t + \tau)^2 - \ddot{d}_i^o(\mathbf{x}_r, t + \tau)(d_i^c(\mathbf{x}_r, t) - d_i^o(\mathbf{x}_r, t + \tau))$. 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:

$$\begin{aligned} \delta\mathcal{L} = & \int_0^T \sum_r [\boldsymbol{\tau}(\mathbf{x}_r, t) - \boldsymbol{\mu}(\mathbf{x}_r, t)\mathbf{h}(\mathbf{x}_r, t)] \cdot \delta\boldsymbol{\tau} dt \\ & + \int_0^T \int_{\Omega} \boldsymbol{\mu}(\mathbf{x}_r, t) \dot{\mathbf{d}}^o(\mathbf{x}_r, t + \tau) \delta(\mathbf{x} - \mathbf{x}_r) \cdot \delta\hat{\mathbf{u}} d^3\mathbf{x} dt \\ & - \int_0^T \int_{\Omega} [\nabla\hat{\psi} : \delta\mathbf{c}^0 : \nabla\mathbf{u}] + [\nabla\psi : \delta\mathbf{c}^0 : \nabla\hat{\mathbf{u}}] d^3\mathbf{x} dt \\ & - \int_0^T \int_{\Omega} [\rho\partial_t^2\hat{\psi} - \nabla \cdot (\mathbf{c}^0 : \nabla\hat{\psi}) - \nabla \cdot (\mathbf{c}^1 : \nabla\psi)] \delta\mathbf{u} d^3\mathbf{x} dt \\ & - \int_0^T \int_{\Omega} [\rho\partial_t^2\psi - \nabla \cdot (\mathbf{c}^0 : \nabla\psi)] \delta\hat{\mathbf{u}} d^3\mathbf{x} dt \\ & - \int_{\Omega} [\rho(\hat{\psi} \cdot \partial_t \delta\mathbf{u} - \partial_t \hat{\psi} \cdot \delta\mathbf{u} + \psi \cdot \partial_t \delta\hat{\mathbf{u}} - \partial_t \psi \cdot \delta\hat{\mathbf{u}})]_0^T d^3\mathbf{x} \\ & + \int_0^T \int_{\partial\Omega} \hat{\psi} \cdot [\mathbf{n} \cdot (\delta\mathbf{c}^0 : \nabla\mathbf{u} + \mathbf{c}^0 : \nabla\delta\mathbf{u})] - \mathbf{n} \cdot (\mathbf{c}^0 : \nabla\hat{\psi}) \cdot \delta\mathbf{u} d^2\mathbf{x} dt \\ & + \int_0^T \int_{\partial\Omega} \psi \cdot [\mathbf{n} \cdot (\delta\mathbf{c}^0 : \nabla\hat{\mathbf{u}} + \mathbf{c}^0 : \nabla\delta\hat{\mathbf{u}})] - \mathbf{n} \cdot (\mathbf{c}^0 : \nabla\psi) \cdot \delta\hat{\mathbf{u}} d^2\mathbf{x} dt \\ & + \int_0^T \int_{\partial\Omega} \psi \cdot [\mathbf{n} \cdot (\mathbf{c}^1 : \nabla\delta\mathbf{u})] - \mathbf{n} \cdot (\mathbf{c}^1 : \nabla\psi) \cdot \delta\mathbf{u} d^2\mathbf{x} dt \end{aligned} \quad (\text{A.6})$$

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:

$$\begin{aligned} \mathbf{u}(\mathbf{x}, 0) = 0, \partial_t \mathbf{u}(\mathbf{x}, 0) = 0, \mathbf{u}(\mathbf{x}, t)|_{x \rightarrow \infty} & \rightarrow 0 \\ \hat{\mathbf{u}}(\mathbf{x}, 0) = 0, \partial_t \hat{\mathbf{u}}(\mathbf{x}, 0) = 0, \hat{\mathbf{u}}(\mathbf{x}, t)|_{x \rightarrow \infty} & \rightarrow 0 \end{aligned} \quad (\text{A.7})$$

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

$$\begin{aligned} \psi(\mathbf{x}, T) = 0, \partial_t \psi(\mathbf{x}, T) = 0, \psi(\mathbf{x}, t)|_{x \rightarrow \infty} & \rightarrow 0 \\ \hat{\psi}(\mathbf{x}, T) = 0, \partial_t \hat{\psi}(\mathbf{x}, T) = 0, \hat{\psi}(\mathbf{x}, t)|_{x \rightarrow \infty} & \rightarrow 0 \end{aligned} \quad (\text{A.8})$$

Thus all the surface integrals in eq.(A.6) will disappear, then:

$$\begin{aligned} \delta\mathcal{L} = & \int_0^T \sum_r [\boldsymbol{\tau}(\mathbf{x}_r, t) - \boldsymbol{\mu}(\mathbf{x}_r, t)\mathbf{h}(\mathbf{x}_r, t)] \delta\boldsymbol{\tau} dt \\ & - \int_0^T \int_{\Omega} [\rho\partial_t^2\hat{\psi} - \nabla \cdot (\mathbf{c}^0 : \nabla\hat{\psi}) - \nabla \cdot (\mathbf{c}^1 : \nabla\psi)] \delta\mathbf{u} d^3\mathbf{x} dt \\ & - \int_0^T \int_{\Omega} [\rho\partial_t^2\psi - \nabla \cdot (\mathbf{c}^0 : \nabla\psi) - \boldsymbol{\mu}(\mathbf{x}_r, t) \dot{\mathbf{d}}^o(\mathbf{x}_r, t + \tau) \delta(\mathbf{x} - \mathbf{x}_r)] \delta\hat{\mathbf{u}} d^3\mathbf{x} dt \\ & - \int_0^T \int_{\Omega} [\nabla\hat{\psi} : \delta\mathbf{c}^0 : \nabla\mathbf{u}] + [\nabla\psi : \delta\mathbf{c}^0 : \nabla\hat{\mathbf{u}}] d^3\mathbf{x} dt. \end{aligned} \quad (\text{A.9})$$

To obtain the stationary point, setting the coefficients of $\delta\tau$ to zero yields:

$$\boldsymbol{\tau}(\mathbf{x}_r, t) - \boldsymbol{\mu}(\mathbf{x}_r, t)\mathbf{h}(\mathbf{x}_r, t) = 0, \quad (\text{A.10})$$

and setting $\delta\mathbf{u}$ and $\delta\hat{\mathbf{u}}$ to zero yields another two adjoint state equations:

$$\begin{aligned} \rho\partial_t^2\psi - \nabla\cdot(\mathbf{c}^0 : \nabla\psi) &= \boldsymbol{\mu}(\mathbf{x}_r, t)\dot{\mathbf{d}}^o(\mathbf{x}_r, t + \tau)\delta(\mathbf{x} - \mathbf{x}_r), \\ \rho\partial_t^2\hat{\psi} - \nabla\cdot(\mathbf{c}^0 : \nabla\hat{\psi}) &= \nabla\cdot(\mathbf{c}^1 : \nabla\psi). \end{aligned} \quad (\text{A.11})$$

The adjoint sources are different in the above equations. The first line represents that the adjoint background wavefields ψ is determined by a new adjoint source at the receiver locations. While the second line represents that the perturbed adjoint wavefields $\hat{\psi}$ is determined by the virtual source generated by the high-wavenumber image perturbations. The adjoint wavefields ψ and $\hat{\psi}$ can be obtained by solving eq.(A.11) in a time-reversed manner. Then, the gradient is:

$$\frac{\partial\mathcal{L}}{\partial\mathbf{c}^0} = - \int_0^T \int_{\Omega} [\nabla\hat{\psi}\nabla\mathbf{u} + \nabla\psi\nabla\hat{\mathbf{u}}] d^3\mathbf{x} dt, \quad (\text{A.12})$$

or in a more detailed manner:

$$\frac{\partial E}{\partial c_{ijkl}^0} = - \int \left(\frac{\partial u_i}{\partial x_j} \frac{\partial \hat{\psi}_k}{\partial x_l} + \frac{\partial \hat{u}_i}{\partial x_j} \frac{\partial \psi_k}{\partial x_l} \right), \quad (\text{A.13})$$

APPENDIX B: FURTHER DECOMPOSITION OF V_S REFLECTION KERNELS

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. Figs 6e-g. Fig. A1 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 background wavefields back-propagated in the media. Mode conversions will produce different type of migration impulse at the source side. It is easy to recognize the PP wavepath shown in Fig. A1e. While the cross-mode correlation produces high-wavenumber energy near the interface. When only injecting PS data, the ψ^{PS-P} will occur as well. Fig. A1g shows a similar case as the Fig. 5f. Note, in A1h 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.(6)).

Finally, we decompose the kernel in Fig. 6g. Only PP data is injected. The non-physical conver-

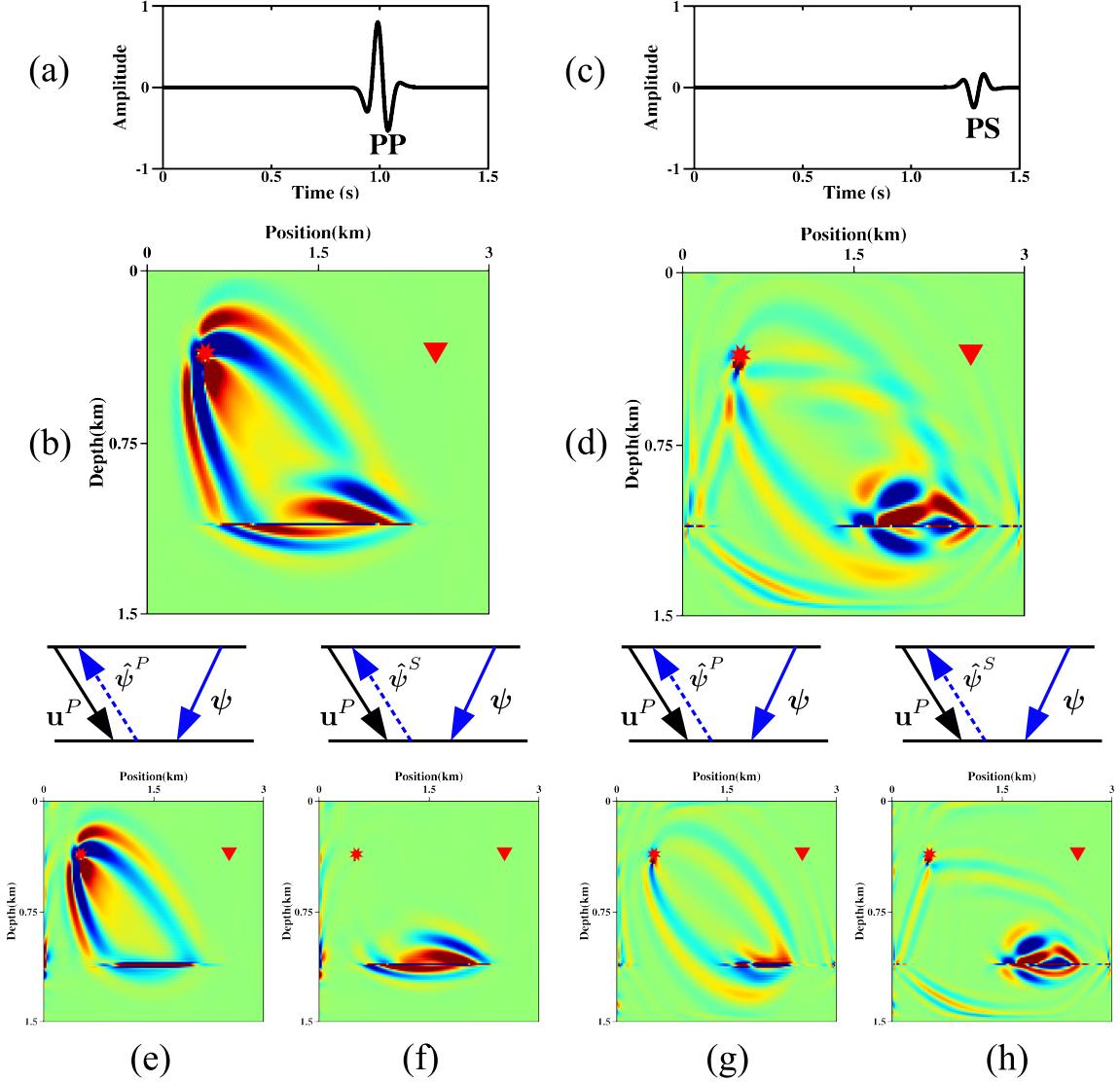


Figure A1. Further decomposed components of Fig 6e and 6f. (a) The PP data for adjoint source , (b) Source side of K_{V_s} using (a) as adjoint source, (c) The PS data for adjoint source, (d) Source side of K_{V_s} using (c) as adjoint source, (e)-(h): The first line denotes the manner and wavefield type for cross-correlation. The second line denotes the corresponding kernels.

sions at receiver are highlighted as red. Fig. A2c 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 (Figs A2d and e). In Fig. A2f, 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.

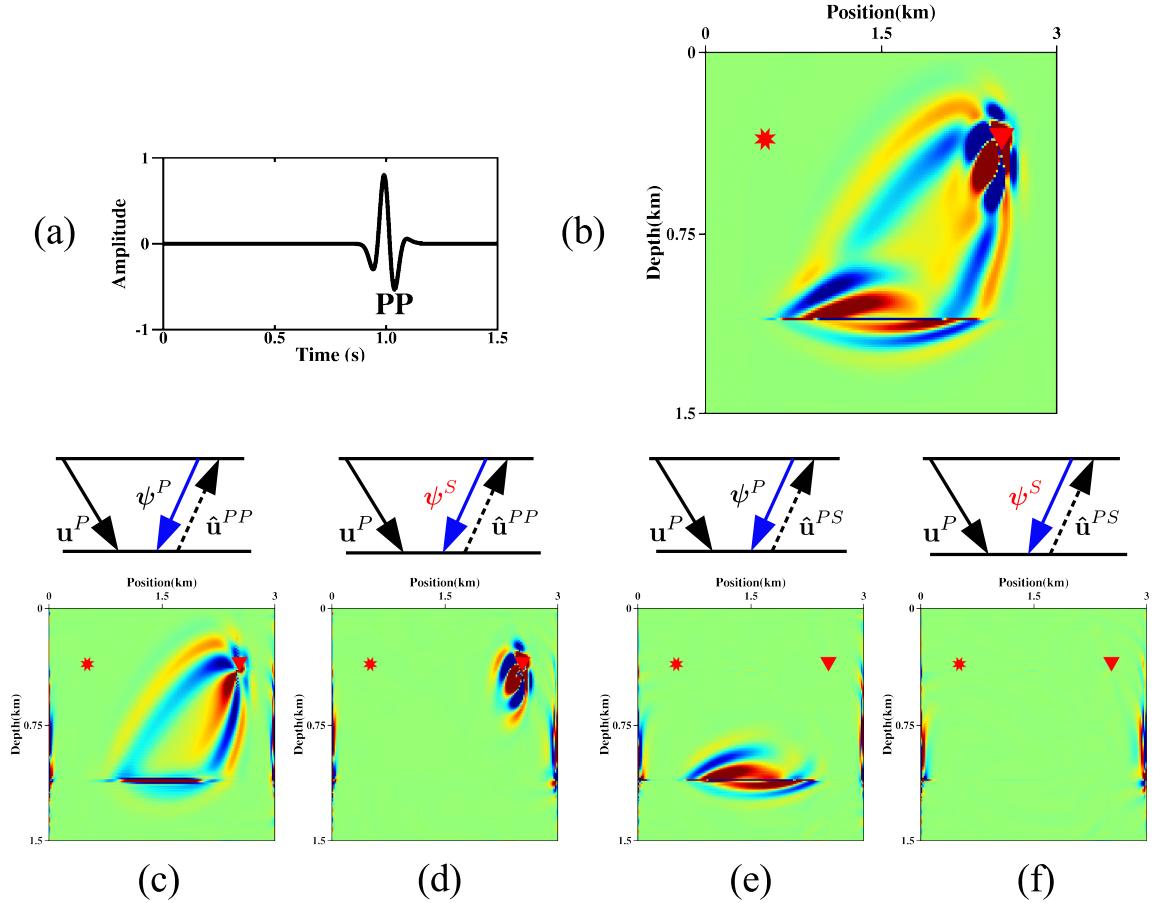


Figure A2. Further decomposed components of Fig 6g. (a) The adjoint source for injection, (b) Receiver term of K_{V_s} using (a) as adjoint source, (c)-(f) The first line denotes the manner and wavefield type for cross-correlation. The second line denotes the corresponding kernels.

APPENDIX C: PROBLEMS IN PS DEMIGRATION

Under the high-frequency approximation, we will illustrate the difference between method **I** and **II** when predicting the PS reflection. A single reflector example is adopted to show the inappropriate behavior of method **I** for inversion through map migration and demigration.

Fig. A3 shows the asymmetric ray path of PS reflection. According to the Snell's law and geometrical relationship, we have:

$$\frac{\sin\theta_1}{V_p} = \frac{\sin\theta_2}{V_s}, \quad (C.1)$$

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

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:

$$t = \frac{z}{\cos\theta_1 V_p} + \frac{z}{\cos\theta_2 V_s}. \quad (\text{C.2})$$

The difference between method **I** and **II** is the migrated depth when velocity model is inaccurate. However, the zero offset map migration for PS data is impossible because PS conversion at near offset is very weak. The energy of near offsets PS data imaging mainly consists of the reflection from certain angles. Therefore, we should use the finite offset map migration to determine the migrated depth.

In method **I**, the finite offset map migration is illustrated by Fig. A3, the migrated depth (z_1) satisfies:

$$\begin{aligned} \frac{z_1}{\cos\alpha_2 V'_p} + \frac{z_1}{\cos\alpha_2 V'_s} &= \frac{z}{\cos\theta_1 V_p} + \frac{z}{\cos\theta_2 V_s}, \\ z_1(\tan\alpha_1 + \tan\alpha_2) &= z(\tan\theta_1 + \tan\theta_2). \\ \frac{\sin\alpha_1}{V'_p} &= \frac{\sin\alpha_2}{V'_s}, \end{aligned} \quad (\text{C.3})$$

where V'_p and V'_s are the wrong velocity, α_1 and α_2 are the incident and reflected angles, respectively, when velocity is wrong. Given a certain incident angle θ_1 , we can calculate the migration depth combining eqs (C.1) and (C.3). While in method **II**, we can use the zero offset map migration. Easily, the depth after migration is:

$$z_2 = \frac{V'_p}{V_p} z. \quad (\text{C.4})$$

Then, we use this depth and the wrong velocity through equation (C.2) to predict the traveltimes of PS reflection.

To show the differences, we compare the predicted traveltimes of the two methods with the observed ones. Here, $V_p = 2500\text{m/s}$, $V_s = 1500\text{m/s}$, $V'_s = 1300\text{m/s}$ are fixed and the value of V'_p are assumed to be correct (2500m/s). We assume the depth of reflector is 1000m, the incident angle is $\theta_1 = 10^\circ$ (offset 281m) As shown in Figure A4a, the predicted PS traveltimes with method **I** (black) is below the true value (red) at large offset, which is consistent with the velocity error. However, at the near offset (Fig.A4b), the predicted traveltimes is above the true one. This will lead to an negative update of the S-wave velocity even the V'_s is lower than the true one. In this case, the PS demigration with method **I** would give a wrong estimation of traveltime residual at near offset. If the incident angle (offset) is smaller, the vertex of the black line will descend until to a “kissing point” when incident angle is zero (green line). Here, we use the ray theory to assume a zero-offset map migration of PS data, but the zero offset data of PS reflection is unavailable in real case. Therefore, the nonlinearity of inversion increases obviously due to this wrong estimation. 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.

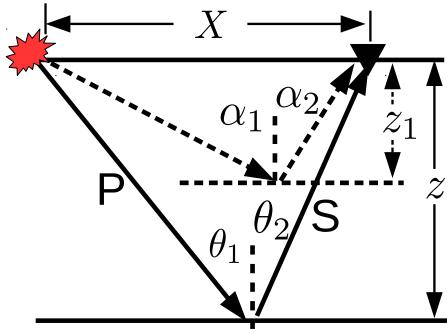


Figure A3. Schematic illustration of PS reflection ray path and its finite offset map migration. The solid line indicates the ray path of PS reflection with a single reflector. The dashed line indicates the finite offset map migration when velocity is wrong.

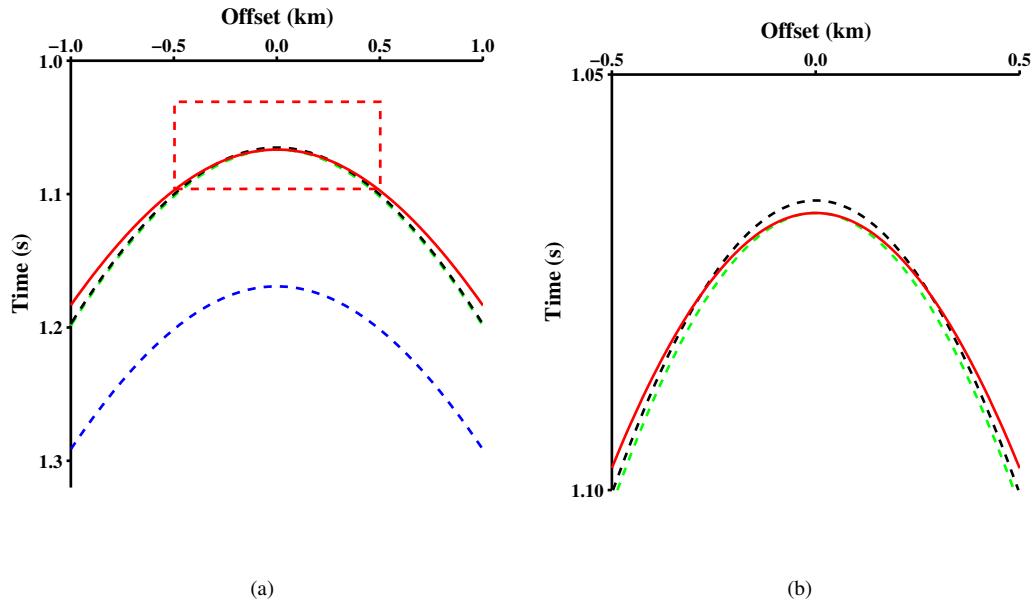


Figure A4. The comparison among demigrated PS traveltimes moveout of method I (black, green), method II (blue) and the real (red) one. The black and green lines represents the results of $\theta_1 = 10^\circ$ and $\theta_1 = 0^\circ$, respectively. (a) the original curves, (b) the curves in the dash box in (a).