



Joint PP and PS Pre-stack Seismic Inversion for Stratified Models Based on the Propagator Matrix Forward Engine

Cong Luo¹ · Jing Ba¹  · José M. Carcione^{1,2} · Guangtan Huang³ · Qiang Guo⁴

Received: 20 May 2020 / Accepted: 22 July 2020 / Published online: 5 August 2020
© Springer Nature B.V. 2020

Abstract

Pre-stack seismic inversion of the P- and S-wave velocities and bulk density is important in seismic exploration for evaluating lithological units and fluid properties. Generally, this inversion is based on ray-tracing modeling, which introduces errors and requires substantial pre-processing for stratified models due to its oversimplified single-interface assumption. To overcome those problems, we propose a pre-stack inversion method, using wave-equation-based modeling as a forward engine. Most wave-equation-based pre-stack inversions are based on the reflectivity method and adopt nonlinear optimization algorithms, although accurate, but computationally expensive. Hence, we use a fast propagator matrix (PM) method valid for layered media. To improve the stability and accuracy, the PP data inversion is extended to joint PP and PS PM-based inversion (JPPI). A linear inversion scheme is adopted to reduce the computational cost, and the Fréchet derivatives are computed accordingly. Moreover, to obtain an optimal model solution, the L-BFGS (Limited-memory Broyden–Fletcher–Goldfarb–Shanno) optimization algorithm and L-curve criterion, an adaptive regularization parameter acquisition method, are implemented. A posterior probability analysis shows that the method has a higher parameter sensitivity than the joint exact Zoeppritz-based inversion and gives better estimations than the single-data inversion. We discuss the effects of dataset weight, internal multi-reflections, time window setting, noise level and initial model by using model tests. Synthetic and real-data examples demonstrate that the algorithm is better than the single PP inversion in terms of stability and accuracy, especially for S-wave velocity and density estimations.

Keywords Propagator matrix · Pre-stack waveform inversion · Joint inversion · Seismic exploration

✉ Jing Ba
jba@hhu.edu.cn

¹ School of Earth Sciences and Engineering, Hohai University, Nanjing 211100, China

² Istituto Nazionale di Oceanografia e di Geofisica Sperimentale (OGS), Borgo Grotta Gigante 42c, 34010 Sgonico, Trieste, Italy

³ School of Earth Sciences, Zhejiang University, Hangzhou 310027, China

⁴ College of Information Engineering, China Jiliang University, Hangzhou 310018, China

1 Introduction

The seismic attributes of a seismic signal (amplitude and phase) are highly affected by the incident angles dictated by the recording system (Ostrander 1984). Amplitude-versus-offset (AVO) or amplitude-versus-angle (AVA) studies use this fact, and pre-stack seismic inversions based on these techniques are widely used to extract information of the elastic properties from seismic data, which are particularly valuable for reservoir prediction, lithological identification and structural analysis. In hydrocarbon exploration, this technique is playing an increasingly significant role (Chopra and Castagna 2007).

Forward engines of inversion process can be useful to model the observed data. According to Simmons and Backus (1994), the AVO/AVA modeling can be classified into two categories, ray-tracing-based and wave-equation-based methods. The first has been used in seismic inversion for several decades, including the Zoeppritz equation (Zoeppritz and Erdbebnenwellen 1919) and its simplification (Shuey 1985; Fatti 1994) for isotropic media, and Graebner equation (Graebner 1992), Rüger approximation (Rüger 1996) and various rock-physics-based equations (Pan et al. 2018b; Pan and Zhang 2018; Pan et al. 2018a; Zong et al. 2015) for anisotropic media. They specify the reflection–transmission coefficients in terms of displacements and medium properties (velocity, mass density, stiffness and anisotropy coefficients). The ray-tracing-based AVO/AVA modeling has two steps. First, rays are traced to compute the two-way travel time and incident angles of the PP or PS reflections from each interface, which is generally replaced by a depth-time matching in the real inversion process. Second, a suitable equation (Zoeppritz or related approximations) is selected to compute the corresponding reflection coefficients versus offset or angle. For isotropic media, the Zoeppritz equation and its simplified forms are frequently used as a forward operator in pre-stack inversion (Zong et al. 2013; Li et al. 2017; Cheng et al. 2018; Huang et al. 2017a, b, 2018; Guo et al. 2018; Luo et al. 2018a; Zhang et al. 2015). The ray-tracing-based pre-stack inversion is quite successful, but has several shortcomings. Besides the approximations of small contrasts and large angle errors, the Zoeppritz equation is derived for a single interface, and this limitation leads to pre-processing requirements of the input seismic data. For an accurate inversion, these data should consist of pure PP reflections, without other unwanted events (transmission losses, surface-related multiples, internal multiples and converted wave modes). Mallick (2007) pointed out that complex wave-propagation effects from large offsets affect the seismic records and the AVO behavior.

Generally, the subsurface consists of many layers, which complex environments, such as sand-mudstone thin interbeds, whose events contaminate the primary reflections. Then, ray-tracing modeling is no longer applicable and a second class of AVO modeling is needed. Derived from an analytical solution to the wave equation, this class improves the single-interface assumption and adopts a recursive mode to compute an accurate wavefield, including primary reflections, primary converted waves, surface-related multiples, internal multiples, internal converted modes and transmissions. It is described as wave-equation-based or recursive matrix-based AVO modeling and has been derived from the classical matrix method (Thomson 1950; Haskell 1953). The concept of “propagator matrix” represents a transfer operator for stress and displacement between two layers in a stratified model (Backus 1962). Using the Thomson–Haskell matrix method, Fuchs and Müller (1971) proposed the reflectivity method to simulate seismograms of a stratified model. This method was developed successively by Kennett and Kerry (1979), Fryer (1980), Kennett (1983a) and extended to the anisotropic media by Booth and Crampin (1983), Fryer

and Frazer (1984) and Mallick and Frazer (1990). Kennett simplified the complex expression of the RM and developed the Kennett recursive matrix (Kennett 1983b; Liu et al. 2016). Besides the reflectivity method, the second class AVO modeling also includes the propagator matrix (PM) method. Carcione (1988, 1995, 1996, 1997, 2014) obtained the reflection coefficients of PP- and PS-converted waves in elastic-anisotropic, viscoelastic-isotropic and viscoelastic-anisotropic media, for a single interface. Similar to the reflectivity method, Carcione (2001, 2014) combined a recursive form with his previous equations to derive a generalized PM method for wavefield simulation in multi-layer viscoelastic and anisotropic media.

Wave-equation-based modeling has been used in pre-stack seismic inversion (Sen and Stoffa 1991; Gisolf et al. 2014; Mallick 1999; Padhi and Mallick 2013; Li and Mallick 2013; Padhi and Mallick 2014; Sen and Roy 2003; Liu et al. 2016). Most of these approaches adopted a nonlinear optimization scheme, which can be found in Sen and Stoffa (1991), Mallick and Adhikari (2015), and Li and Mallick (2015). Although the nonlinear inversion scheme is helpful to obtain an optimum solution and it is derivative-free (Guo et al. 2018, 2020a, b), it is computationally expensive. Therefore, Sen and Roy (2003) and Liu et al. (2016) used the Kennett recursive matrix as forward engine and adopted a Gauss–Newton linear optimization. Most commonly used linear methods are gradient-based algorithms, such as the gradient descent, conjugate gradient, Gauss–Newton and Newton methods. The gradient descent is not applicable when using the first-order derivative as updating direction. The Newton method computes the optimal model solution along the second-order derivative, which has the fast convergence rate. However, it is difficult to obtain the Hessian matrix. Therefore, quasi-Newton optimization algorithms are preferred to approximate the Newton method. The Gauss–Newton method, one of the quasi-Newton methods, uses the square of the first derivative to replace the second derivatives, but it has slow or no convergence. Speeding up the convergence rates and reducing the computational costs of wave-equation-based inversion is important for 2D/3D multi-parameter inversions.

In isotropic media, we have to invert for three properties, namely, P- and S-wave velocities and density of each layer. It is a multi-parameter inversion problem, leading to instabilities, in part due to the insufficient input data. To improve the stability and accuracy, one way is to reduce the inversion dimension. Another approach is the joint PP and PS wave inversion, whose effectiveness was verified in seismic pre-stack inversion (Lu et al. 2015b, 2018; Padhi and Mallick 2014; Li and Mallick 2015). Density is an important parameter for identifying lithology and detecting hydrocarbon. According to Luo et al. (2018b), the S-wave velocity and density cannot be obtained accurately by using P-wave inversion. Even in the case of long-offset data, the inaccuracy in determining density is still a problem due to its insensitivity to the primary (PP) reflection amplitudes (Debski and Tarantola 1995; Luo et al. 2018b). The converted wave mode (PS data) contains more information on S-wave velocity and density than PP reflections. Therefore, a joint PP and PS inversion can provide better results. However, a problem occurs in ray-tracing-based modeling for joint inversion, i.e., the PS time should be aligned with the PP time, increasing the calculations and introducing processing errors due to the manual picking process and PS waveform distortion by stretching and compressing. Unlike the ray-tracing-based inversion, the wave-equation-based inversion method uses a depth-domain initial model directly, which means that the algorithm does not require the PS-to-PP time alignment.

All previous work based on wave-equation-based inversion uses the reflectivity method (Kennett 1983b; Fryer and Frazer 1984) as forward operator (Sen and Stoffa 1991; Sen and Roy 2003; Mallick 1999; Li and Mallick 2015; Pafeng, Mallick, and Sharma 2016; Liu et al.

2016). Here, we implement the PM (Carcione 2001, 2014) simplified for the isotropic case. Moreover, use of joint PP and PS data completes the approach, the joint PM-based inversion (JPMI). To reduce computational cost, a linear inversion scheme is adopted to obtain the model solution. The Fréchet derivatives, the first derivatives of the PM modeling with respect to target parameters, are derived. Speeding up the convergence of the optimization algorithm is crucial, so that the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm, adopting an iterative method to approximate the Newton method, is used to solve the objective function and compute an optimal solution. The effectiveness of the L-BFGS optimization was verified with full-waveform inversion of real data (Yang and Engquist 2018; Yang et al. 2018) and using the Zoeppritz-based inversion (Huang et al. 2018). By introducing the L-curve algorithm into the linear inversion scheme, we adaptively acquire an optimal regularization weight of the JPMI. The effects of dataset weight, internal multi-reflections, time window setting, noise level and initial model on the JPMI are discussed by using model tests. Synthetic and real-data examples are given, and the method is compared to the PP inversion.

2 Methodology

2.1 Propagator Matrix Forward Modeling

The PM method (Carcione 2001, 2014) is used as the forward modeling engine for the inversion. Although it assumes a locally 1D subsurface model, it uses an exact solution to the full-wave equation.

In seismology, a layered system can be a good representation of the stratified earth model. We consider the stratified model consisting of three parts: a half-space top layer, an intermediate layered system (including N horizontal layers) and a half-space bottom layer. According to Carcione (2014), the particle-velocity/stress array, inside the j th layer at depth z , can be written as:

$$\mathbf{t}(z) = \begin{pmatrix} v_1 \\ v_3 \\ \delta_{33} \\ \delta_{13} \end{pmatrix} = \mathbf{T}(z) \cdot \begin{pmatrix} U_P^- \\ U_S^- \\ U_P^+ \\ U_S^+ \end{pmatrix}, \tag{1}$$

where v_1 and v_3 are the particle-velocity components along the x - and z -axis, δ_{33} and δ_{13} are the normal and tangential stresses, and U^- and U^+ are upgoing- and downgoing-wave amplitudes, respectively. The subscripts P and S denote the compressional and shear waves. $\mathbf{T}(z)$ has the form as

$$\mathbf{T}_j(z) = i\omega \begin{pmatrix} \xi_{jP} e^{i\omega s_P z} & \xi_{jS} e^{i\omega s_S z} & \xi_{jP} e^{-i\omega s_P z} & \xi_{jS} e^{-i\omega s_S z} \\ -\gamma_{jP} e^{i\omega s_P z} & -\gamma_{jS} e^{i\omega s_S z} & \gamma_{jP} e^{-i\omega s_P z} & \gamma_{jS} e^{-i\omega s_S z} \\ -X_{jP} e^{i\omega s_P z} & -X_{jS} e^{i\omega s_S z} & -X_{jP} e^{-i\omega s_P z} & -X_{jS} e^{-i\omega s_S z} \\ W_{jP} e^{i\omega s_P z} & W_{jS} e^{i\omega s_S z} & -W_{jP} e^{-i\omega s_P z} & -W_{jS} e^{-i\omega s_S z} \end{pmatrix} \tag{2}$$

with

$$X = (\alpha^2 - 2\beta^2) \rho \xi_P + \alpha^2 \rho \gamma_S, \quad W = \beta^2 \rho (\gamma_P + \xi_S), \tag{3}$$

where α , β and ρ are the P- and S-wave velocities, and bulk density, respectively. Moreover, ω is the angular frequency, t is the time variable. The plane-wave eigenvectors

(polarization) belonging to a particular eigenvalue can be obtained from the Kelvin–Christoffel equation. Then, we get the horizontal and vertical polarization ξ and γ (Carcione 1997a)

$$\xi = \sqrt{\frac{\alpha^2 s^2 + \beta^2 p^2 - 1}{(\alpha^2 + \beta^2)(p^2 + s^2) - 2}}, \quad \gamma = \sqrt{\frac{\alpha^2 p^2 + \beta^2 s^2 - 1}{(\alpha^2 + \beta^2)(p^2 + s^2) - 2}}, \tag{4}$$

where p and s are the horizontal and vertical slownesses, and the slowness relation (Carcione 1997a) is solved for s , given the horizontal slowness p . It yields

$$s = \pm \frac{1}{\sqrt{2}} \sqrt{K_1 \mp \sqrt{K_1^2 - 4K_2K_3}}, \tag{5a}$$

where

$$K_1 = \frac{1}{\alpha^2} + \frac{1}{\beta^2} - 2p^2, \quad K_2 = p^2 - \frac{1}{\alpha^2}, \quad K_3 = p^2 - \frac{1}{\beta^2}. \tag{5b}$$

The signs in s correspond to: (+, -) downward propagating P-wave, (+, +) downward propagating S-wave, (-, -) upward propagating P-wave and (-, +) upward propagating S-wave. By assuming that the thickness of the j th layer is h_j , the fields at $z = 0$ (the top interface of the j th layer) and $z = h_j$ (the bottom interface of the j th layer) are related by

$$\mathbf{t}_j(0) = \mathbf{B}_j \cdot \mathbf{t}_j(h_j). \tag{6a}$$

From Eq. (1), we have

$$\mathbf{B}_j = \mathbf{T}_j(0) \cdot [\mathbf{T}_j(h_j)]^{-1}. \tag{6b}$$

The particle-velocity/stress vectors \mathbf{t} have the convenient property that they are continuous across any horizontal plane. By matching boundary conditions at the interfaces between layers, the vectors at $z = 0$ and $z = h$ have the following relationship as

$$\mathbf{t}(0) = \prod_{j=1}^N \mathbf{B}_j \cdot \mathbf{t}(h), \tag{7}$$

where $h = \sum_{j=1}^N h_j$ is the total thickness of the intermediate N -layers layered system. There are incident, reflected and transmitted waves. Therefore, the particle-velocity/stress field at $z = 0$ can be expressed as

$$\mathbf{t}(0) = \mathbf{A}_1 \cdot \mathbf{R} + \mathbf{b}, \tag{8}$$

where $\mathbf{R} = (R_{PP} \ R_{PS} \ T_{PP} \ T_{PS})^T$ is the reflection and transmission coefficients of the whole model in the frequency-slowness domain, $\mathbf{b} = i\omega (\xi_P \ \gamma_P \ -X_P \ -W_P)^T$ is the incident P-wave array and

$$\mathbf{A}_1 = i\omega \begin{pmatrix} -\xi_{1P} & -\xi_{1S} & 0 & 0 \\ -\gamma_{1P} & -\gamma_{1S} & 0 & 0 \\ -X_{1P} & -X_{1S} & 0 & 0 \\ -W_{1P} & -W_{1S} & 0 & 0 \end{pmatrix}. \tag{9}$$

The bottom half-space, containing only outgoing or evanescent waves below $z = h$, has the particle-velocity/stress field

$$\mathbf{t}(h) = \mathbf{A}_2 \cdot \mathbf{R}, \tag{10}$$

where

$$\mathbf{A}_2 = i\omega \begin{pmatrix} 0 & 0 & \xi_{2P} e^{-i\omega s_p h} & \xi_{2S} e^{-i\omega s_s h} \\ 0 & 0 & \gamma_{2P} e^{-i\omega s_p h} & \gamma_{2S} e^{-i\omega s_s h} \\ 0 & 0 & -Z_{2P} e^{-i\omega s_p h} & -Z_{2S} e^{-i\omega s_s h} \\ 0 & 0 & -W_{2P} e^{-i\omega s_p h} & -W_{2S} e^{-i\omega s_s h} \end{pmatrix}. \tag{11}$$

By using Eqs. (7), (8) and (10), the total reflection and transmission coefficients of the whole layered system in the frequency-slowness domain can be calculated as

$$\mathbf{G} \cdot \mathbf{R} = \mathbf{b}, \tag{12}$$

with

$$\mathbf{G} = \left[\left(\prod_{j=1}^N \mathbf{B}_j \right) \cdot \mathbf{A}_2 - \mathbf{A}_1 \right]. \tag{13}$$

From Eq. (12), we obtain the total reflection coefficients of the PP and PS waves in the frequency-slowness domain, $R(s, \omega)$. Using the PM to obtain synthetics needs to integrate twice, slowness and frequency. To reduce the computation time and improve the stability, the PM modeling is implemented in the frequency-angle domain to synthesize the angle gathers by integrating over frequency only.

A time–space domain seismogram can be generated by integrating first over slowness and then over frequency (Mallick and Frazer 1987; Liu et al. 2016). Integrating over frequency, we get

$$r(s, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(s, \omega) e^{i\omega t} d\omega. \tag{14}$$

By introducing the relation between the horizontal slowness p and the incident angle θ , $p = \sin\theta / \alpha$ into Eq. (14), one can transform the reflection coefficients from the frequency-slowness domain to the frequency-angle domain:

$$r(\theta, t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} R(\theta, \omega) e^{i\omega t} d\omega. \tag{15}$$

2.2 Objective function of joint inversion

Let \mathbf{d} be the observed-data vector, and \mathbf{F} the synthetic-data vector for the unknown model \mathbf{m} . Here, \mathbf{F} represents a nonlinear forward operator, such as the PM described in detail above,

$$\mathbf{d}_{pp} = \mathbf{F}_{pp}(\mathbf{m}), \quad \mathbf{d}_{ps} = \mathbf{F}_{ps}(\mathbf{m}), \tag{16}$$

where \mathbf{m} is the matrix of the model parameters, \mathbf{d} is the physical realization, and F , a non-linear function of the model parameters. The subscripts PP and PS refer to the nonlinear forward engines based on the PM for the PP and PS records. According to the perturbation theory and a Taylor series expansion, the nonlinear modeling in Eq. (16) can be linearized as

$$\Delta \mathbf{d}_{PP} \approx \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta \mathbf{m}, \quad \Delta \mathbf{d}_{PS} \approx \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta \mathbf{m}. \tag{17}$$

The model perturbation, $\Delta \mathbf{m}$, is linear with the physical realization perturbation (data residual vector, difference between observed and computed data), $\Delta \mathbf{d}$. The quantities $\Delta \mathbf{d}_{PP}$ and $\Delta \mathbf{d}_{PS}$ represent the perturbations as

$$\Delta \mathbf{d}_{PP} = \mathbf{d}_{PP}^{obs} - \mathbf{d}_{PP}^{syn}, \quad \Delta \mathbf{d}_{PS} = \mathbf{d}_{PS}^{obs} - \mathbf{d}_{PS}^{syn}, \tag{18}$$

where \mathbf{d}_{PP}^{obs} and \mathbf{d}_{PS}^{obs} are the real PP and PS datasets, \mathbf{d}_{PP}^{syn} and \mathbf{d}_{PS}^{syn} are the synthetic PP and PS gathers. The quantities $\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}}$ and $\frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}}$ are the Fréchet derivate matrices, the first-order derivate matrices of the nonlinear forward engines with respect to the model parameters and also key intermediate computations of the linear inversion problem. Based on the convolution theory, the Fréchet derivative can be solved by using the derivative of the time-domain scattering coefficients \mathbf{r} (including the reflection and transmission coefficients) with respect to the target model parameters \mathbf{m} as

$$\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} = \mathbf{W}_{PP} \cdot \frac{\partial \mathbf{r}_{PP}}{\partial \mathbf{m}}, \quad \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} = \mathbf{W}_{PS} \cdot \frac{\partial \mathbf{r}_{PS}}{\partial \mathbf{m}}. \tag{19}$$

Let us assume that in the usual three-parameter pre-stack inversion, \mathbf{m} is a $3N$ -sample-length vector consisting of P-wave velocity, S-wave velocity and density sample points. Then,

$$\mathbf{m} = \left(m_1^\alpha \cdots m_N^\alpha, m_{N+1}^\beta \cdots m_{2N}^\beta, m_{2N+1}^\rho \cdots m_{3N}^\rho \right)^T, \tag{20}$$

where the superscripts α , β and ρ denote the P-wave velocity, S-wave velocity and density parameters, respectively. Based on Eq. (15), the partial derivatives of the reflectivity with respect to the model parameters can be calculated analytically by using the partial derivation of the frequency-angle-domain reflection coefficients,

$$\frac{\partial \mathbf{r}_{PP}(\theta_k, t_j)}{\partial m_i^*} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial \mathbf{R}_{PP}(\theta_k, \omega_l)}{\partial m_i^*} e^{i\omega t} d\omega, \tag{21a}$$

$$\frac{\partial \mathbf{r}_{PS}(\theta_k, t_j)}{\partial m_i^*} = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{\partial \mathbf{R}_{PS}(\theta_k, \omega_l)}{\partial m_i^*} e^{i\omega t} d\omega. \tag{21b}$$

The partial derivation matrix in the frequency-angle domain has the following expression as

$$\frac{\partial \mathbf{R}}{\partial \mathbf{m}} = \left[\frac{\partial \mathbf{R}}{\partial m^\alpha} \quad \frac{\partial \mathbf{R}}{\partial m^\beta} \quad \frac{\partial \mathbf{R}}{\partial m^\rho} \right], \tag{22}$$

where $\frac{\partial \mathbf{R}}{\partial m^{\alpha=\beta,\rho}}$ denotes a $KM \times N$ dimension matrix. Taking α as an example, we have

$$\frac{\partial \mathbf{R}}{\partial \boldsymbol{\alpha}} = \begin{bmatrix} \frac{\partial R(\theta_1, \omega_1)}{\partial \alpha_1} & \dots & \frac{\partial R(\theta_1, \omega_M)}{\partial \alpha_1} & \dots & \frac{\partial R(\theta_K, \omega_1)}{\partial \alpha_1} & \dots & \frac{\partial R(\theta_K, \omega_M)}{\partial \alpha_1} \\ \vdots & \ddots & \vdots & \dots & \vdots & \ddots & \vdots \\ \frac{\partial R(\theta_1, \omega_1)}{\partial \alpha_N} & \dots & \frac{\partial R(\theta_1, \omega_M)}{\partial \alpha_N} & \dots & \frac{\partial R(\theta_K, \omega_1)}{\partial \alpha_N} & \dots & \frac{\partial R(\theta_K, \omega_M)}{\partial \alpha_N} \end{bmatrix}^T. \tag{23}$$

where N , M and K denote the sample length of the model, the number of frequency samples and the number of incidence angles. By differentiating Eqs. (12) and (13) with respect to the model parameter $m_j^{*=\alpha, \beta, \rho}$, the Fréchet derivation can be calculated as

$$\frac{\partial \mathbf{R}}{\partial m_j^*} = -\mathbf{G}^{-1} \cdot \frac{\partial \mathbf{G}}{\partial m_j^*} \cdot \mathbf{R}, \tag{24}$$

where

$$\frac{\partial \mathbf{G}}{\partial m_j^*} = \left(\mathbf{B}_1 \mathbf{B}_2 \dots \frac{\partial \mathbf{B}_j}{\partial m_j^*} \dots \mathbf{B}_n \right) \cdot \mathbf{A}_2 \tag{25}$$

with

$$\frac{\partial \mathbf{B}_j}{\partial m_j^*} = \frac{\partial \mathbf{T}_j(0)}{\partial m_j^*} \cdot [\mathbf{T}_j(h_j)]^{-1} - \mathbf{B}_j \cdot \frac{\partial \mathbf{T}_j(h_j)}{\partial m_j^*} \cdot [\mathbf{T}_j(h_j)]^{-1}. \tag{26}$$

The partial derivatives $\frac{\partial \mathbf{T}_j(0)}{\partial m_j^*}$ and $\frac{\partial \mathbf{T}_j(h_j)}{\partial m_j^*}$ are given in Appendix. Based on the Fréchet derivatives in the frequency domain, according to Eq. (21), the derivatives in the time domain can be obtained by a frequency integration, which is commonly replaced by a fast Fourier transform to reduce computer time.

To solve the model vector, an objective function is defined. Solving the inverse problem means to seek the minimum of this objective function. Focusing on Bayesian inversion, we give the posterior distribution for model perturbation $\Delta \mathbf{m}$, and the physical realization perturbation (data residual vector, the difference between observed and computed data), $\Delta \mathbf{d}$, as

$$\begin{aligned} P(\Delta \mathbf{m} | \Delta \mathbf{d}) \propto & \left\{ -\frac{1}{2} \left[\phi \left(\Delta \mathbf{d}_{PP} - \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta \mathbf{m} \right)^T \left(\Delta \mathbf{d}_{PP} - \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta \mathbf{m} \right) \right. \right. \\ & + (1 - \phi) \cdot \left(\Delta \mathbf{d}_{PS} - \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta \mathbf{m} \right)^T \left(\Delta \mathbf{d}_{PS} - \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta \mathbf{m} \right) \left. \right] \\ & + \lambda (\Delta \mathbf{m} - \mathbf{u})^T \mathbf{C}_{\Delta m}^{-1} (\Delta \mathbf{m} - \mathbf{u}) \left. \right\} \end{aligned} \tag{27}$$

Here, we assume that the model perturbation $\Delta \mathbf{m}$ satisfies the Gaussian distribution. In Eq. (27), \mathbf{u} is the expectation of model parameters, and \mathbf{C} is the covariance matrix of the parameter vector to improve the stability of the multi-parameter inversion, which can be expressed as

$$\mathbf{C} = \begin{bmatrix} \boldsymbol{\sigma}_{\alpha\alpha} & \boldsymbol{\sigma}_{\alpha\beta} & \boldsymbol{\sigma}_{\alpha\rho} \\ \boldsymbol{\sigma}_{\alpha\beta} & \boldsymbol{\sigma}_{\beta\beta} & \boldsymbol{\sigma}_{\beta\rho} \\ \boldsymbol{\sigma}_{\alpha\rho} & \boldsymbol{\sigma}_{\beta\rho} & \boldsymbol{\sigma}_{\rho\rho} \end{bmatrix}, \tag{28}$$

where $\boldsymbol{\sigma}$ is an $N \times N$ dimension matrix. We take

$$\sigma_{\alpha\alpha} = \begin{bmatrix} \sigma_{\alpha_1\alpha_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_{\alpha_N\alpha_N} \end{bmatrix} \tag{29}$$

as an example. Maximizing the above posterior probability gives a model solution, which is equivalent to minimizing the following objective function

$$S = \frac{1}{2} \left[\phi \left(\Delta \mathbf{d}_{PP} - \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta \mathbf{m} \right)^T \left(\Delta \mathbf{d}_{PP} - \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta \mathbf{m} \right) + (1 - \phi) \cdot \left(\Delta \mathbf{d}_{PS} - \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta \mathbf{m} \right)^T \left(\Delta \mathbf{d}_{PS} - \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta \mathbf{m} \right) \right], \tag{30}$$

$$+ \lambda (\Delta \mathbf{m} - \mathbf{u})^T \mathbf{C}_{\Delta m}^{-1} (\Delta \mathbf{m} - \mathbf{u})$$

where ϕ is the dataset weight which is introduced to balance the contributions of the PP and PS data, which varies from 0 to 1. λ is the regularization weight which determines the relative importance between the data and model residuals. The nonlinear forward engines, such as the Zoeppritz equation and the PM used in this paper, are more complex than the linear methods, where a local linearization is employed to obtain model updates for each iteration, and therefore a constant regularization weight may not seem appropriate (Sen and Roy 2003). The PM-based inversion, which employs more computational time, needs an adaptive-acquisition method for the regularization parameter, more than the widely used Zoeppritz-based inversion. Detailed explanations are given below.

2.3 Optimization Algorithm

A linear inversion scheme is used to reduce calculation costs. Based on an initial value, one can update the model results by several iterations. Gradient-based optimization algorithms follow the iteration

$$\Delta \mathbf{m}_{k+1} = \Delta \mathbf{m}_k - a_k \mathbf{g}_k, \tag{31}$$

where $\Delta \mathbf{m}_k$ is the updated model and a_k denotes the step size of the k th iteration, obtained by the strong Wolfe line search algorithm (Zhi et al. 2016). $-\mathbf{g}_k$ is the updating direction of the model parameters. The L-BFGS optimization method makes use of an iterative process to approximate the Hessian matrix by using the Jacobian matrix. It has a good convergence rate and acceptable storage requirement. The iteration equation of the L-BFGS method is

$$\Delta \mathbf{m}_{k+1} = \Delta \mathbf{m}_k - a_k \mathbf{H}_k^{-1} \mathbf{J}_k, \tag{32}$$

where \mathbf{J} is the Jacobian matrix, the first-order derivative of objective function Eq. (30) and a function of the model perturbations $\Delta \mathbf{m}$,

$$\begin{aligned} \mathbf{J} = \nabla_{\Delta\mathbf{m}} S(\Delta\mathbf{m}) &= \phi \left(\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \right)^T \left(\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta\mathbf{m} - \Delta \mathbf{d}_{PP} \right) \\ &+ (1 - \phi) \cdot \left(\frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \right)^T \left(\frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta\mathbf{m} - \Delta \mathbf{d}_{PS} \right) \quad (33) \\ &+ \lambda \mathbf{C}_{\Delta\mathbf{m}}^{-1} (\Delta\mathbf{m} - \mathbf{u}) \end{aligned}$$

and \mathbf{H} is the quasi-Hessian matrix, computed as

$$\mathbf{H}_{k+1}^{-1} = \mathbf{V}_k^T \mathbf{H}_k^{-1} \mathbf{V}_k + \rho_k (\Delta\mathbf{m}_k - \Delta\mathbf{m}_{k-1}) \cdot (\Delta\mathbf{m}_k - \Delta\mathbf{m}_{k-1})^T, \quad (34a)$$

where

$$\rho_k = \frac{1}{(\mathbf{J}_k - \mathbf{J}_{k-1})^T (\Delta\mathbf{m}_k - \Delta\mathbf{m}_{k-1})}, \quad (34b)$$

$$\mathbf{V}_k = \mathbf{I} - \rho_k (\mathbf{J}_k - \mathbf{J}_{k-1}) \cdot (\Delta\mathbf{m}_k - \Delta\mathbf{m}_{k-1})^T. \quad (34c)$$

For the first iteration, we use the updating format of the Gauss–Newton method and

$$\mathbf{H}_1 = \frac{\partial \mathbf{J}_1}{\partial \Delta\mathbf{m}} \approx \phi \left(\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \right)^T \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} + (1 - \phi) \left(\frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \right)^T \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} + \lambda \mathbf{C}_{\Delta\mathbf{m}}^{-1}. \quad (35)$$

2.4 Determination of the Regularization Weight

The regularization parameter adaptive-acquisition methods can be classified into two categories: (1) priori selection methods that require knowledge of the level of the data error or noise; and (2) posteriori selection methods that do not require knowledge of this information (Sen and Roy 2003; Huang et al. 2017a, b). Since the level of noise is usually unknown, the posteriori selection (L-curve) (Hansen and O’Leary 1993; Hansen 2000; Sen and Roy 2003) and generalized cross-validation (GCV) (Craven and Wahba 1978; Golub and Von Matt 1997; Wahba 1990; Huang et al. 2017a) are often used to select the weight.

Huang et al. (2017b) compared the L-curve with the GCV criterion and concluded that first shows better stability and noise immunity. In this work, the L-curve method is considered to obtain the optimal weight adaptively.

The L-curve is a graphical representation of model misfit and data misfit by using a suitable plotting scale definition for the varying values of regularization weight. The typical L-curve method is derived from Tikhonov regularization (Sen and Roy 2003) and improved to the regularization with arbitrary norm constraints (Huang et al. 2017a). When using a norm constraint $q(\mathbf{m})$ in a diagonal matrix form, objective function (30) becomes

$$\begin{aligned} \Delta\mathbf{m}_\lambda = \arg \min \left\{ \right. &\left[\phi \left(\Delta \mathbf{d}_{PP} - \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta\mathbf{m} \right)^T \left(\Delta \mathbf{d}_{PP} - \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta\mathbf{m} \right) \right. \\ &\left. + (1 - \phi) \cdot \left(\Delta \mathbf{d}_{PS} - \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta\mathbf{m} \right)^T \left(\Delta \mathbf{d}_{PS} - \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta\mathbf{m} \right) + \lambda^2 q(\Delta\mathbf{m}) \right. \left. \right\}, \quad (36) \end{aligned}$$

where $q(\Delta\mathbf{m})$ is an arbitrary norm constraint in a diagonal matrix form. Then $\Delta\mathbf{m}_\lambda$ can be obtained at a minimum of the objective function,

$$\Delta\mathbf{m}_\lambda = \left[\phi \left(\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \right)^T \left(\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \right) + (1 - \phi) \cdot \left(\frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \right)^T \left(\frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \right) + \lambda^2 \mathbf{Q} \right]^{-1} \cdot \left[\phi \left(\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \right)^T \Delta \mathbf{d}_{PP} + (1 - \phi) \cdot \left(\frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \right)^T \Delta \mathbf{d}_{PS} \right] \tag{37}$$

where $\mathbf{Q} = \frac{1}{\Delta\mathbf{m}} \cdot \frac{\partial q}{\partial \Delta\mathbf{m}}$. Using the singular-value decomposition of $\frac{\partial \mathbf{F}}{\partial \mathbf{m}}$, $\Delta\mathbf{m}_\lambda$ can be expressed as

$$\begin{aligned} \Delta\mathbf{m}_\lambda &= [\phi \Sigma_{PP}^2 + (1 - \phi) \Sigma_{PS}^2 + \lambda^2 \mathbf{Q}]^{-1} \cdot \left[\phi \left(\frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \right)^T \Delta \mathbf{d}_{PP} + (1 - \phi) \left(\frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \right)^T \Delta \mathbf{d}_{PS} \right] \\ &= \sum_{i=1}^n \frac{\left[\phi v_i^{PP} \delta_i^{PP} (u_i^{PP})^T \Delta d_{PP} + (1 - \phi) v_i^{PS} \delta_i^{PS} (u_i^{PS})^T \Delta d_{PS} \right]}{\phi (\delta_i^{PP})^2 + (1 - \phi) \cdot (\delta_i^{PS})^2 + \lambda^2 q_i} \end{aligned} \tag{38}$$

Substituting $\Delta\mathbf{m}_\lambda$ into the two parts (misfit and regularization terms) of objective function (36) gives

$$\begin{aligned} \text{misfit} &= \phi \left\| \Delta \mathbf{d}_{PP} - \frac{\partial \mathbf{F}_{PP}}{\partial \mathbf{m}} \Delta \mathbf{m}_\lambda \right\|_2^2 + (1 - \phi) \left\| \Delta \mathbf{d}_{PS} - \frac{\partial \mathbf{F}_{PS}}{\partial \mathbf{m}} \Delta \mathbf{m}_\lambda \right\|_2^2 \\ &= \sum_{i=1}^n \left[\phi \left(\Delta d_{PP} - u_i^{PP} \delta_i^{PP} v_i^{PP} \Delta m_\lambda \right)^2 + (1 - \phi) \left(\Delta d_{PS} - u_i^{PS} \delta_i^{PS} v_i^{PS} \Delta m_\lambda \right)^2 \right] \end{aligned} \tag{39a}$$

$$\text{regular.} = q(\Delta\mathbf{m}_\lambda). \tag{39b}$$

Different regularization weights lead to different results of the misfit and regularization terms, which can be used to plot the L-curve. The corner of the curve (point of minimum curvature) corresponds to the optimal value of the weight.

3 Forward Modeling Analysis

3.1 Full-Wavefield Responses

Ray-tracing methods (exact Zoeppritz equation and its approximations) are often used as the forward operator of pre-stack inversion. A main difference between ray-tracing and wave-equation-based methods is that not all wave modes are accounted for by the first method. A three-layer model (model 1) is considered to test the wave modes modeled by PM and the exact Zoeppritz method. To compute the two-way travel time easily and identify the different wave modes clearly, the velocities are set at the same value, as shown in Table 1. Figure 1 displays the PP and PS synthetic seismograms modeled by the PM (Fig. 1a–d) and the Zoeppritz (Fig. 1e–f) method. Some multi-reflection modes have the same travel time in a homogeneous medium and overlapping events in the seismograms, for example, the PPSP and PSPP (triple-leg P- and single-leg S-waves), the PPPSPP and

PPSPPP (five-leg P- and single-leg S-waves), etc. We use the symbols $P(i)$ and $S(j)$ for the multi-reflections with i -leg P-wave and j -leg S-waves, respectively. The zero offset travel times of different wave modes of the full wavefield are computed and shown in Tables 2 and 3, and the corresponding events are marked in Fig. 1a, b. The PM synthetics without normal moveout (NMO) correction (Fig. 1a, b) show the full wavefield with multi-reflections (multiples and conversion waves), and the reflections of PP and PS seismograms from the second interface are $P(4)S(0)$ and $P(2)S(2)$, respectively. However, the Zoeppritz results contain two-leg primary reflections $P(2)S(0)$ and $P(1)S(1)$ only, see Fig. 1e, f. We extract the peak amplitude of events 1 and 2 from PP (Fig. 1c, e) and PS (Fig. 1d, f) synthetics calculated by the Zoeppritz method (Fig. 1e, f) and the PM (Fig. 1c, d), as shown in Fig. 2. The peak amplitudes of the first event show no difference (see Fig. 2a, b). However, the peak amplitudes of the second event calculated with the PM decrease compared to those of the Zoeppritz method (see Fig. 2c, d). Therefore, the PM algorithm effectively simulates the transmission losses and the Zoeppritz modeling only contains two-leg primary reflections.

3.2 Parameter Sensitivity

The amplitude variation with offset (AVO) effect in seismic data contains information of the elastic parameters of the geological layers. One can extract this information by using inversion methods. As the general solution of an inverse problem consists of a probability over all possible models, we calculate the probabilities by using PP, PS and PP-PS (joint PP and PS waves) data to test the parameter sensitivity. A test-stratified model with nine layers is assumed. We vary the properties of two layers to obtain several possible models, so that α , β and ρ are in the ranges -480 to 480 m/s, -415 to 415 m/s and -0.11 to 0.1 g/cm³, respectively. The corresponding perturbation percentages are $-19\% \sim 19\%$, $-23.75\% \sim 23.75\%$ and $-6\% \sim 5.5\%$. The “observed data” A_i^{OBS} is the “true model” and the synthetic data $A_i(\mathbf{m})$ from possible models \mathbf{m} . The details of posterior probability density are given in the paper of Debski and Tarantola (1995). Figure 3 shows the 2D marginal probability density by using the PP (a, d, g, j), PS (b, e, h, k) and PP-PS (c, f, i, l) angle gathers ranging from 0° to 35° . Figure 3a–f shows the 2D probability of α and β with the true ρ , and Fig. 3g–l shows the results of β and ρ with the true α .

The 2D posterior information of PP-wave data shows that the probabilities of α and β of the PM are more focused than those of the Zoeppritz method (Fig. 3a, d). It means that the PP-wave data modeled by the PM is more sensitive to α and β than the ray-tracing method. There are obvious differences between the probabilities of the PS-wave data by the two algorithms. The interpretation of such differences is important, because it represents the sensitivity information we have on the target parameters. Due to the PS insensitivity to α variations, as shown in Fig. 3e, the 2D probabilities of the Zoeppritz

Table 1 Model parameters of the sand and shale rock layers of model 1

	P-wave velocity α (m/s)	S-wave velocity β (m/s)	Density ρ (g/cm ³)
Shale rock	2000	816	3.5
Sand rock	2000	816	2.0
Shale rock	2000	816	3.5

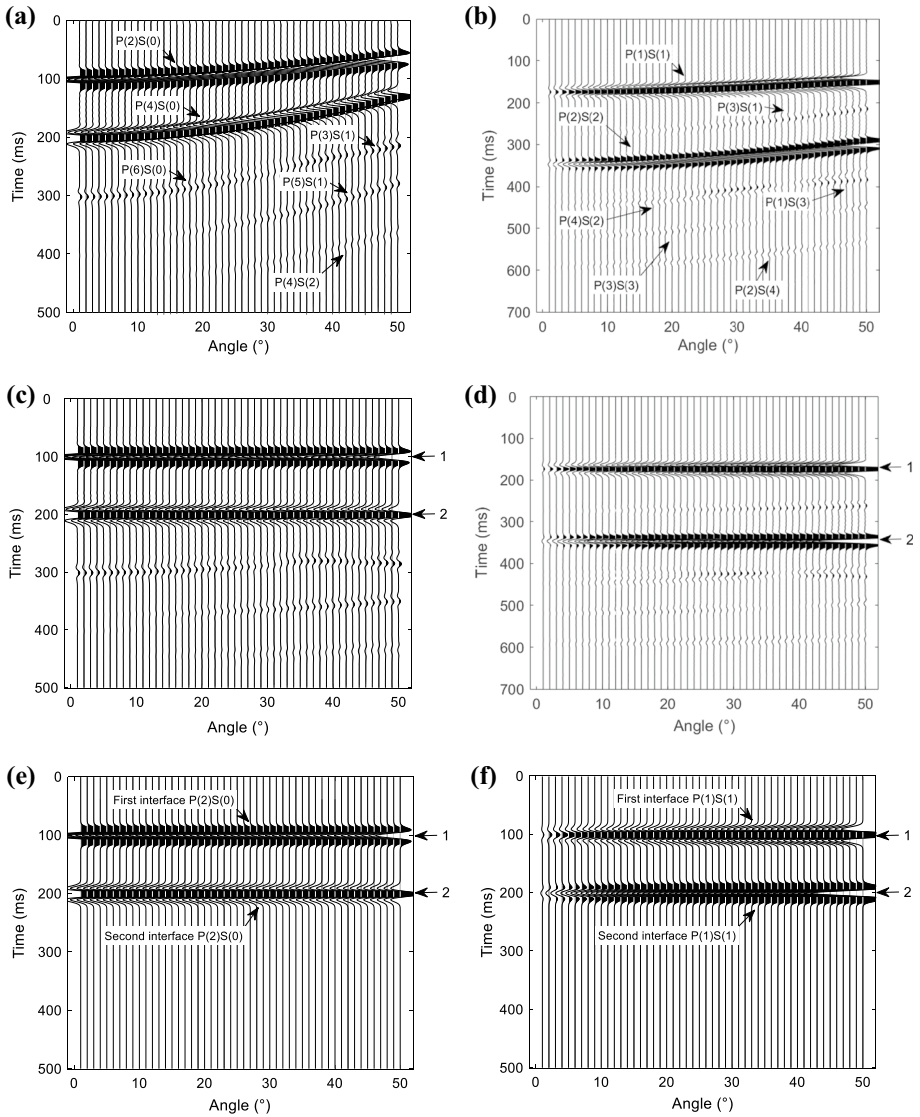


Fig. 1 PP wave (a, c, e) and PS wave (b, d, f) seismograms computed with the PM (a–d) and Zoeppritz’s equation (e, f). a, b The synthetic slant stack and c, d NMO-corrected angle gathers

Table 2 Wave modes of the PP seismogram of model 1 and estimated two-way travel time at zero offset

Wave mode	Two-way time (ms)
P(2)S(0)	100
P(4)S(0)	200
P(6)S(0)	300
P(3)S(1)	272
P(5)S(1)	372
P(4)S(2)	444

Table 3 Wave modes of the PS seismogram of model 1 and the estimated two-way travel time at zero offset

Wave mode	Two-way time (ms)
P(1)S(1)	172
P(2)S(2)	344
P(3)S(1)	272
P(1)S(3)	416
P(4)S(2)	444
P(3)S(3)	516
P(2)S(4)	588

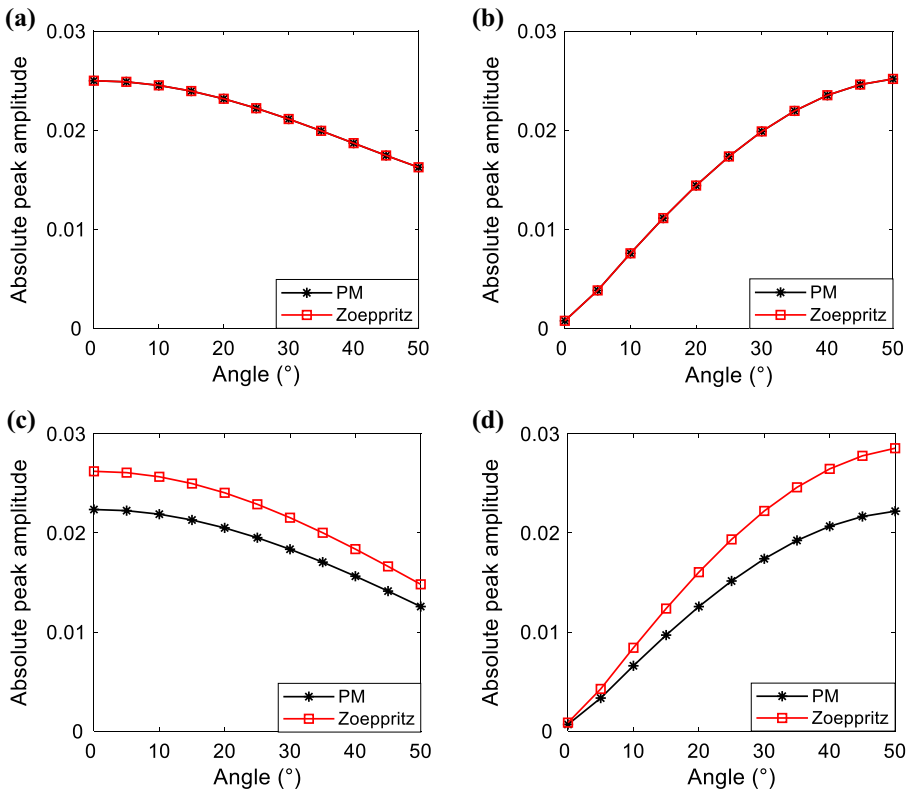


Fig. 2 Absolute peak amplitudes as a function of the incidence angle corresponding to the events marked by arrows in Fig. 1. Events 1 of the PP (1c, 1e) (a) and PS (1d, 1f) (b) data. Events 2 of the PP (1c, 1e) (c) and PS (1d, 1f) (d) data

method present a monotonic function along the α axis [the prior function is monotonic within the range, while the likelihood function, affected by the parameter variation, is an ellipsoid-shaped function]. Similarly, the PS data modeled by the Zoeppritz method is less sensitive to ρ , which is evidenced by the monotonic function along the ρ direction in Fig. 3k. The 2D posterior probabilities by using the PM (see Fig. 3b, h) are ellipsoid-shaped functions. It is shown that the PS-wave data modeled by the PM are more

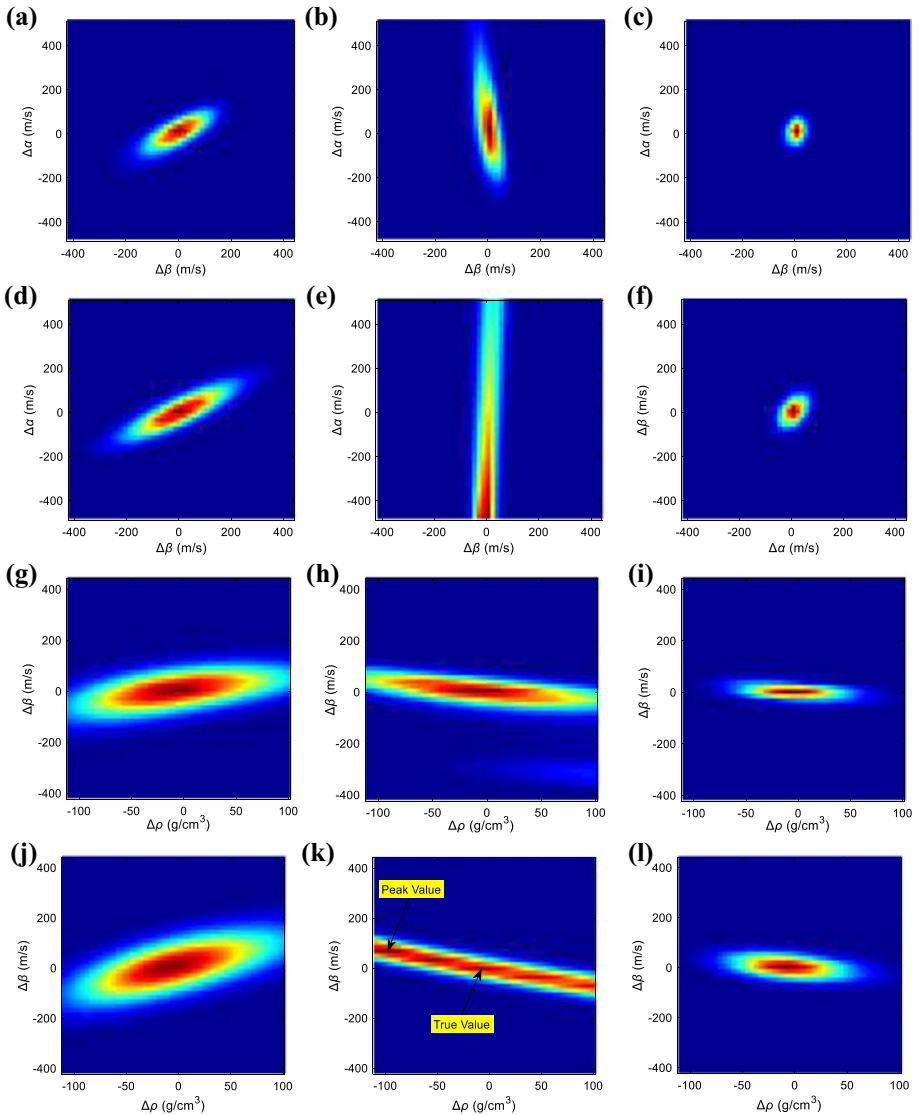


Fig. 3 2D posterior probability density of PP- (a, d, g, j), PS- (b, e, h, k) and joint PP- and PS-wave data (c, f, i, l) by the PM (a–c and g–i) and the Zoeppritz (d–f and j–l). a–f Probability density of α and β , with a fixed ρ (true value). g–l Probability density of β and ρ , with a fixed α (true value)

sensitive to β and ρ than that of the Zoeppritz modeling. Compared to the inversion with pure PP- or PS-wave data, joint inversion with PP and PS data provides better estimates of the parameters, which is shown in the 2D probability of PP-PS data (Fig. 3c, f, i, l), with a more focused function than that of PP- (Fig. 3a, d, g, j) and PS-wave data (Fig. 3b, e, h, k). The probabilities of the joint data also indicate that the PM has a more convex function along the α and β directions than the Zoeppritz method.

Figure 4 represents the 1D marginal probability densities of the PP (Fig. 4a, d, g), PS (Fig. 4b, e, h) and PP-PS data (Fig. 4c, f, i), obtained from 1D integrations of the 2D probabilities of Fig. 3. The two algorithms use the same a priori distribution; therefore, the higher sensitivity to parameter leads to the more convex function of the 1D probability. The probability of the PP-wave data in Fig. 4 shows that the PP reflections of the PM are more sensitive to the three parameters (slightly more sensitive to ρ) than that the Zoeppritz method. Compared to PP-wave data, the PS wave of the PM has higher parameter sensitivity than that the Zoeppritz method, especially α and ρ (Fig. 4b, h). The results of PP-PS show that the probabilities of the three parameters computed with the PM are more convex,

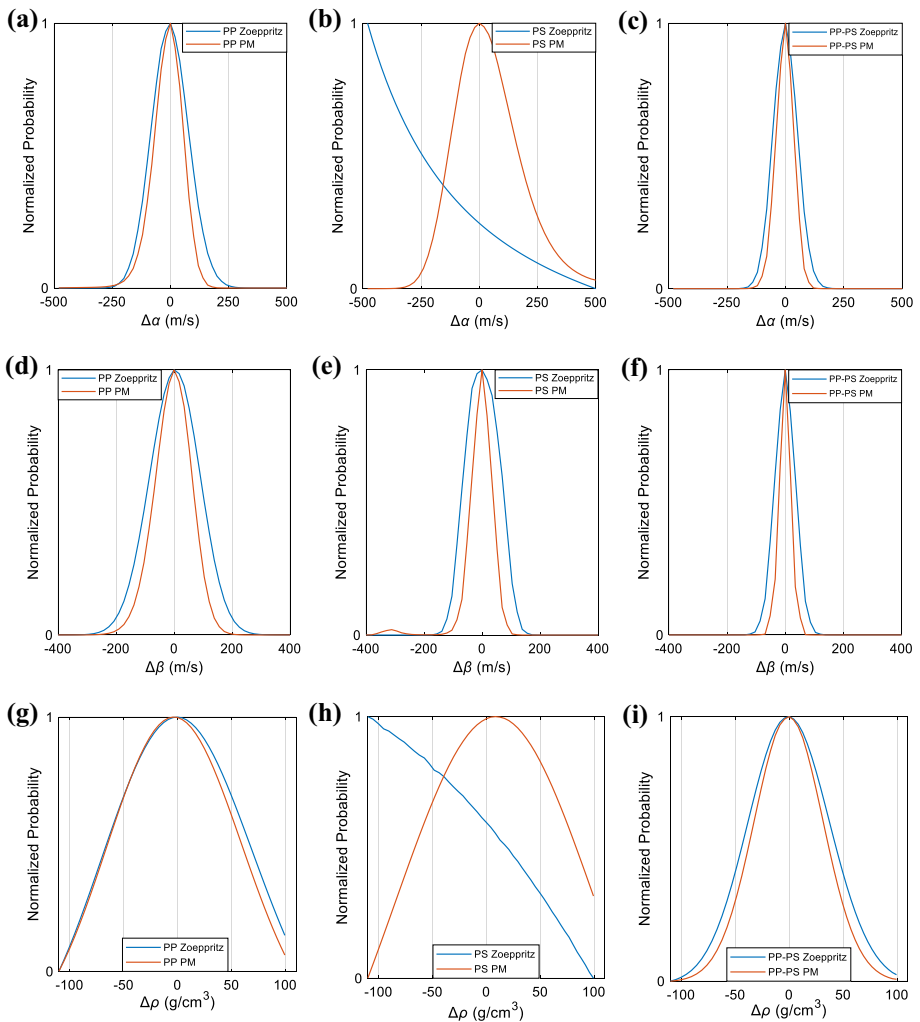


Fig. 4 1D posterior probability density of PP (a, d, g), PS (b, e, h) and joint PP and PS data (c, f, i) by the PM (red lines) and the Zoeppritz (blue lines). a–c 1D probabilities of α obtained by β integration of the 2D probabilities of Fig. 3a–f. d–f 1D probabilities of β obtained by ρ integration of the 2D probabilities of Fig. 3g–i. g–i 1D probabilities of ρ obtained by β integration of the 2D probabilities of Fig. 3g–i

i.e., more sensitive. The joint PP and PS inversion combined with the PM provides a better estimation.

Mallick (2007) showed that mode-converted reflections induce significant differences between the full-wave-mode and the Zoeppritz P-wave primary responses. The multi-reflections contain information on the elastic parameters, resulting in a big difference between the two algorithms for the multi-layer case.

3.3 Dataset Weight Factor

The weight factor ϕ is an important parameter to be used in dealing with PP- and PS-wave datasets. For noise-free or the same noise-level conditions, the weight ϕ is commonly set as 0.5 (Zhi et al. 2016; Lu et al. 2015a). However, PP- and PS-wave data have different sensitivities to α , β and ρ , as discussed in the last section. Then, it is necessary to determine suitable weight factors for a high-quality inversion. Posterior probability density is computed by using the model in Sect. 3.2 to test the weight factor. The variations of α , β and ρ are the same as those of the last section. Both “observed data” A_i^{OBS} and synthetic data $A_i(\mathbf{m})$ are obtained with the PM. Figure 5 shows the 1D probability densities with different weight factors by using 0° – 35° angle PP-PS data. When the weight factor ranges from 0 to 1, a clear difference of 1D probability density is observed. A more focused probability density curve means that it is easier to obtain a better inversion result with the same optimization method.

The steepness is introduced to represent the degree of focus of the probability density curves. It is defined as the ratio of the half-height of curve ε and the corresponding horizontal distance δ (Fig. 5c). A more focused curve leads to a higher steepness. The steepness of the target parameter varying with ϕ is shown in Fig. 6. When ϕ is 0, only PS-wave data are used in the estimation; when ϕ is 1, only PP-wave data are used. The weight factors $\phi = 0.35$ and $\phi = 0.6$ correspond to the highest steepness of α , β and ρ and then can be used as optimal factors for noise-free conditions. The PP-wave data have higher sensitivities to P-wave velocity and slightly higher sensitivity to density; therefore, α and ρ estimations require weight factors larger than 0.5. The appropriate factor for the S-wave velocity is less than 0.5, because the PS wave is more sensitive to β than the PP wave.

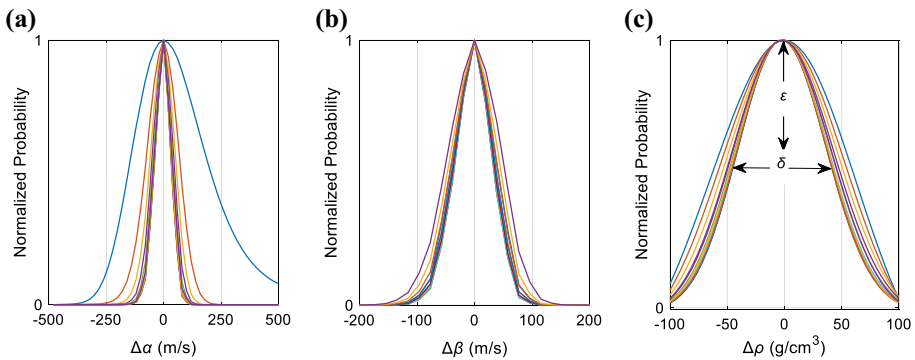


Fig. 5 1D posterior probability with different dataset weight by using the PM. **a** 1D probabilities of α . **b** 1D probabilities of β . **c** 1D probabilities of ρ

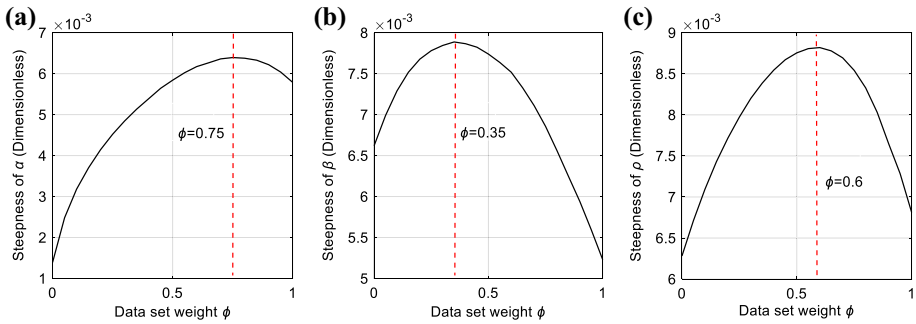


Fig. 6 Steepness as a function of the dataset weight factor ϕ . The steepness values are computed by using the 1D probability densities shown in Fig. 5. **a** P-wave velocity α . **b** S-wave velocity β . **c** Density ρ

Since the model test demonstrates the three elastic parameters have different sensitivities to the seismic dataset, a parallel strategy is proposed to improve the inversion accuracy and mitigate the effect of the coupling. We setup three objective functions with three dataset weights and three regularization parameters which correspond to the three target parameters. For noise-free conditions, the optimal dataset weights are set as 0.75, 0.35 and 0.6 (selected in Sect. 3.3) which are fixed through all the iterations, and the functions are solved separately to obtain three solution vectors. In each iteration, the adaptive-acquisition method (L-curve) is adopted to estimate three regularization parameters, each of which plays an important role to obtain a relatively “optimal solution” of one of the parameters. Then, the three optimal solutions are extracted from the model solution vectors and rebuilt into a new one, which is set as the initial model vector of the next iteration.

4 Inversion Analysis

4.1 Regularization Weight Test

The model illustrated in Fig. 7b (blue lines) is used to test the L-curve, a regularization weight adaptive-acquisition method. The reflectivity series obtained with the PM method is convolved with a Ricker wavelet of 30 Hz dominant frequency to generate the seismograms. Noiseless synthetic gathers and noisy gathers with signal-to-noise ratio of 3 (SNR; the ratio of root-mean-square amplitude of the signal to that of the noise) (Gaussian random noises are added to the gathers) are used to test inversion results. We use different regularization weights for the noise-free and the SNR = 3 cases and calculate the total RMS relative errors of the three parameters, as shown in Fig. 7a, c.

Different regularization weights parameters lead to different errors. It is shown that there is an optimal weight leading to a minimum error, and the L-curve criterion can be used to obtain this weight for both noise-free and high-noise cases. Figure 7b, d shows the results for the noise-free and SNR = 3 cases, respectively. When the noise level is increased (SNR = 3), the match between the inverted and true curves is still acceptable.

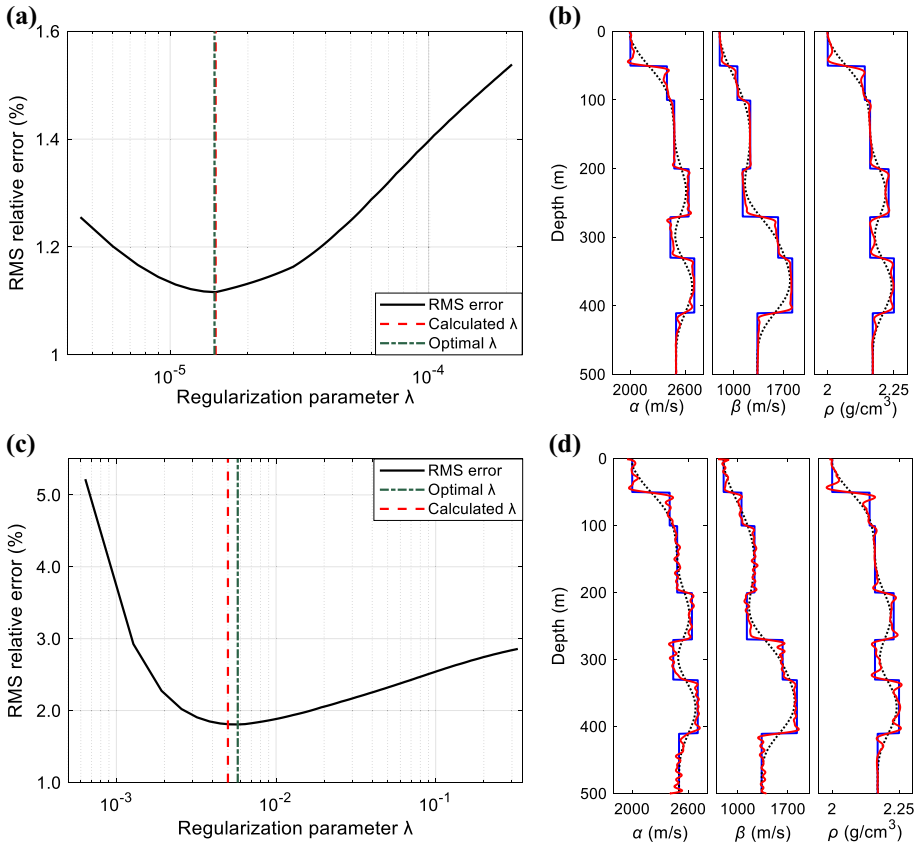


Fig. 7 Regularization weight test for the noise-free (a, b) and noisy case (SNR=3) (c, d). a, c RMS relative error of the inversion as a function of the regularization weight. b, d Inversion curves by using the weight computed with the L-curve criterion

4.2 Effect of Internal Multi-reflections

Actual seismic records are inevitably affected by transmission loss and contaminated by P-wave multiples and conversion modes, referred to as multi-reflections. Since the PM is an exact simulation method for obtaining full-wave responses of stratified models, the corresponding inversion can directly utilize the effects of transmission loss and multiple reflections. However, current processing flow of pre-stack seismic data is based mainly on ray-tracing-based inversion methods, such as Zoeppritz-based inversion (EZI). It means that, in pre-stack data currently available, transmission loss has been compensated and the multiple reflections have been suppressed to some extent. Therefore, we set up models 2 and 3 and perform the inversion test by using the angle gathers with transmission loss compensation and surface multiple removal. Model 2 is a simple block model used to simulate thick sedimentary reservoirs, whose properties are shown in Fig. 8a. Figure 9a, b displays the PP and PS slant angle gathers of model 2 by PM with surface multiples removal (synthetics of PM do not include surface multiples due to the assumption of a half-space top layer), but without transmission compensation, normal moveout (NMO)

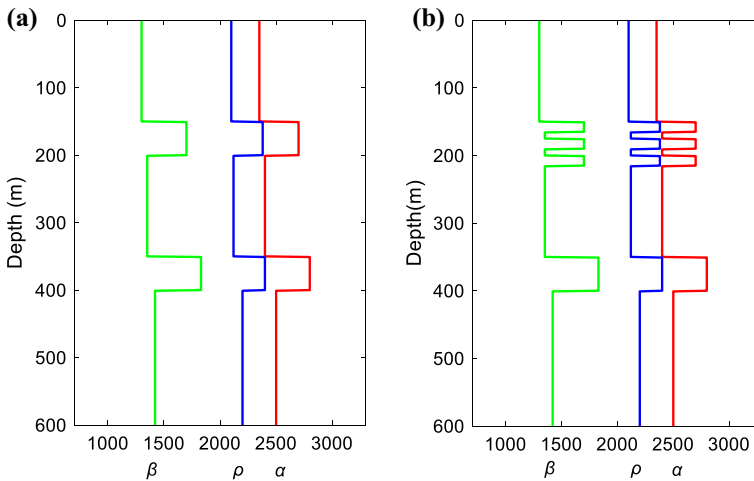


Fig. 8 P- (red) and S-wave velocities (green) and bulk density (blue) of model 2 (a) and 3 (b)

correction and internal multi-reflection suppression. After NMO correction and transmission compensation, we obtain the processed PP and PS gathers in Fig. 9c, d. Relatively strong energy, periodic features and apparently residual moveout make surface multiples easy to be identified and removed. However, compared with surface multiples, in practical processing, internal multiple reflections are difficult to distinguish from primary ones because of their weak energy, aperiodic features and small residual moveout after NMO correction, as marked by arrows in Fig. 9c, d. Moreover, they are often aliased with primary reflections and sometimes change their AVO responses (Mallick 2007). Therefore, the final gathers used for inversion undergo surface multiple removal, NMO correction and transmission loss compensation and contain primary and internal multi-reflections, as shown in Fig. 9c, d. We compute the P- α and S-wave velocity β and density ρ by joint PP and PS exact Zoeppritz-based inversion (JEZI) and JPMI. The inversion result in the time domain by JEZI and the results in the depth domain by JPMI are shown in Fig. 16a, b, respectively. The JEZI obtains good inverted parameters for layer 1, 2 and 4, which match the actual parameters well. However, multiples of input data being regarded as primary reflections cause the wrong parameter estimations in JEZI (marked by arrows in Fig. 10a). The correlation coefficients (CCs) of P- and S-wave velocities and density from time-domain JEZI are 0.8891, 0.8843 and 0.8689, respectively, shown in Table 4. JPMI can make use of the various internal multiple reflections. The inversion results in the depth domain are displayed in Fig. 16b. Eliminating the influence of multi-reflections brings higher consistency with actual properties. Although affected by transmission loss compensation, the estimations in highlighted areas are slightly higher than the true values, the results by JPMI are still acceptable with CCs of 0.9818 (P-wave velocity), 0.9806 (S-wave velocity) and 0.9786 (density), shown in Table 4.

Replacing the second layer of model 2 with a set of interbeds, model 3 is obtained, to test effects of these interbeds (Fig. 8b). Figure 11a, b is the synthetic PP and PS slant gathers containing no surface multiples, and Fig. 11c, d is the processed gathers including primary and internal multiple reflections after NMO correction and transmission loss compensation. Compared with model 2, the interbeds in model 3 generate more complex multi-reflection responses which pollute to some extent the effective

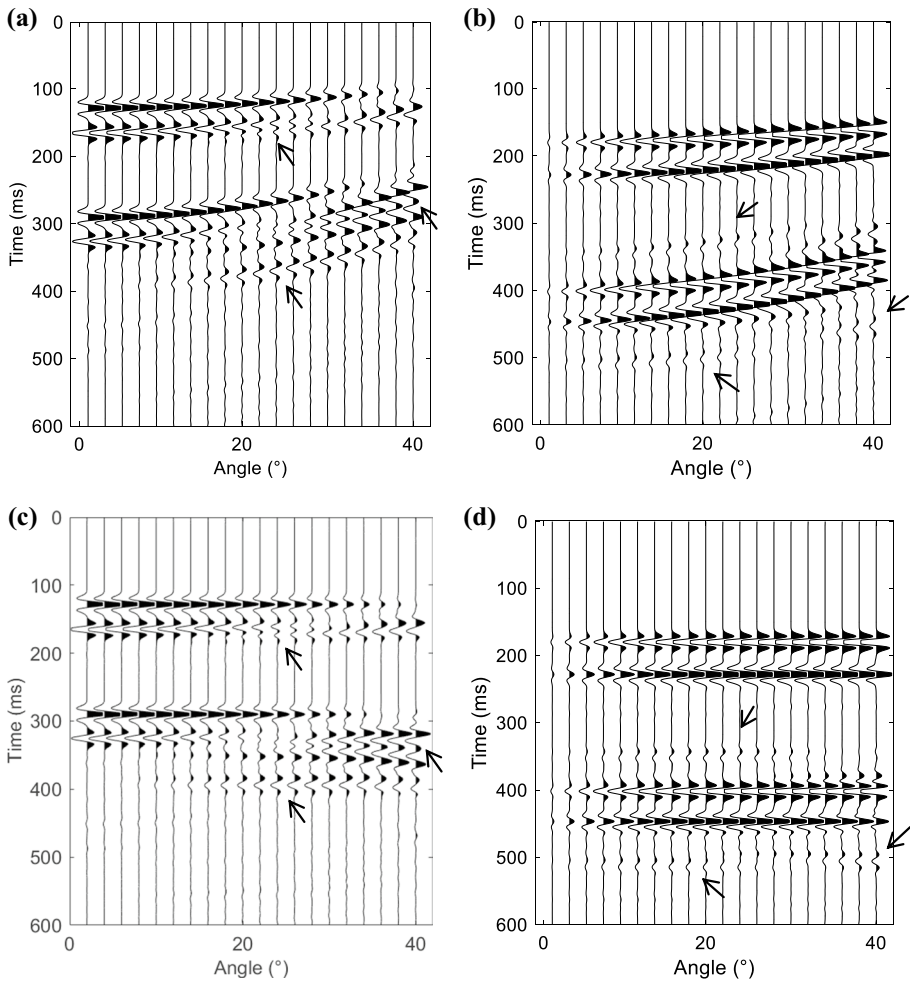


Fig. 9 Synthetic seismicograms obtained with the PM for the model 2 of Fig. 8. (a, c) PP seismicograms in PP time domain **a** without and **c** with normal moveout (NMO) correction and transmission loss compensation. **b, d** PS seismicograms in PS time domain **b** without and **d** with NMO correction and transmission loss compensation

primary reflections from the underlying interfaces (see the marked areas in Fig. 11c, d). Using the input data in Fig. 11c, d, two inversion methods are adopted and the corresponding results are displayed in Fig. 12a, b. The inversion result in the time domain by JEZI (Fig. 12a) shows more estimation bias for lower layers induced by multi-reflections than that of model 2 (Fig. 10a) and has lower CCs: 0.8347, 0.8306 and 0.8211 shown in Table 5. The parameters inverted by JPMI, shown in Fig. 10b, match better with actual properties, as the CCs are 0.9731, 0.9701 and 0.9665 (Table 5).

Fig. 10 Inversion results of model 2 in Fig. 8a using the input gathers shown in Fig. 9. **a** The inversion results in time domain by JEZI. **b** The inversion results in depth domain by JPMI. The blue solid, black dotted and red solid lines are the actual parameters, initial models and inverted results, respectively

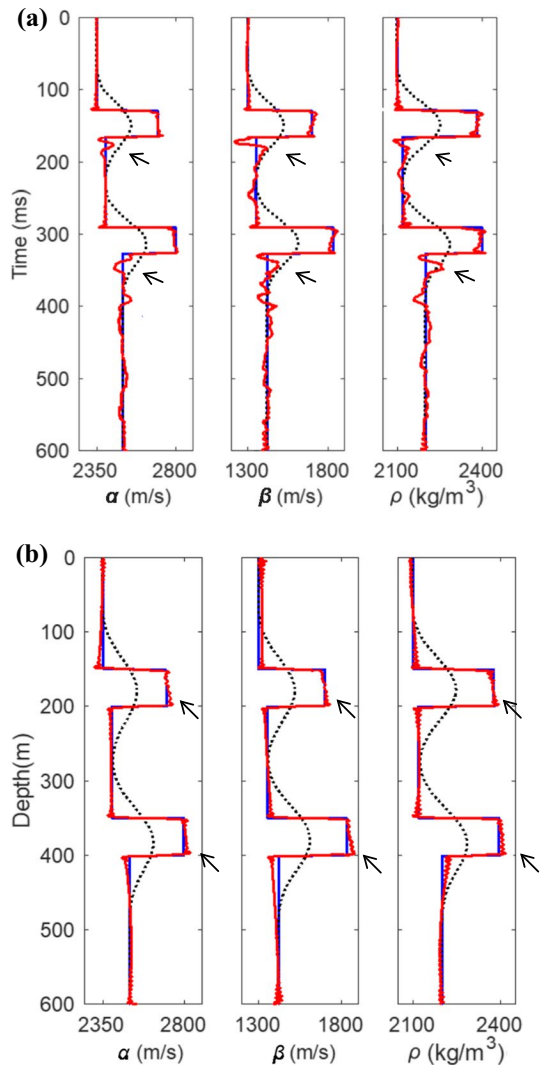


Table 4 Correlation coefficients (CCs) between the inversion results and the true properties of model 2 by using JEZI and JPMI

Method	P-wave velocity α	S-wave velocity β	Density ρ
JEZI	0.8891	0.8843	0.8689
JPMI	0.9818	0.9806	0.9786

4.3 Inversion of the PP, PS and PP-PS Data

A well-log dataset is used to test the PMI method by using PP wave, PS wave and joint PP-PS data. When using a well-log model for the inversion, the log curves should be processed first by a Backus averaging (Backus 1962), since there is an apparent difference

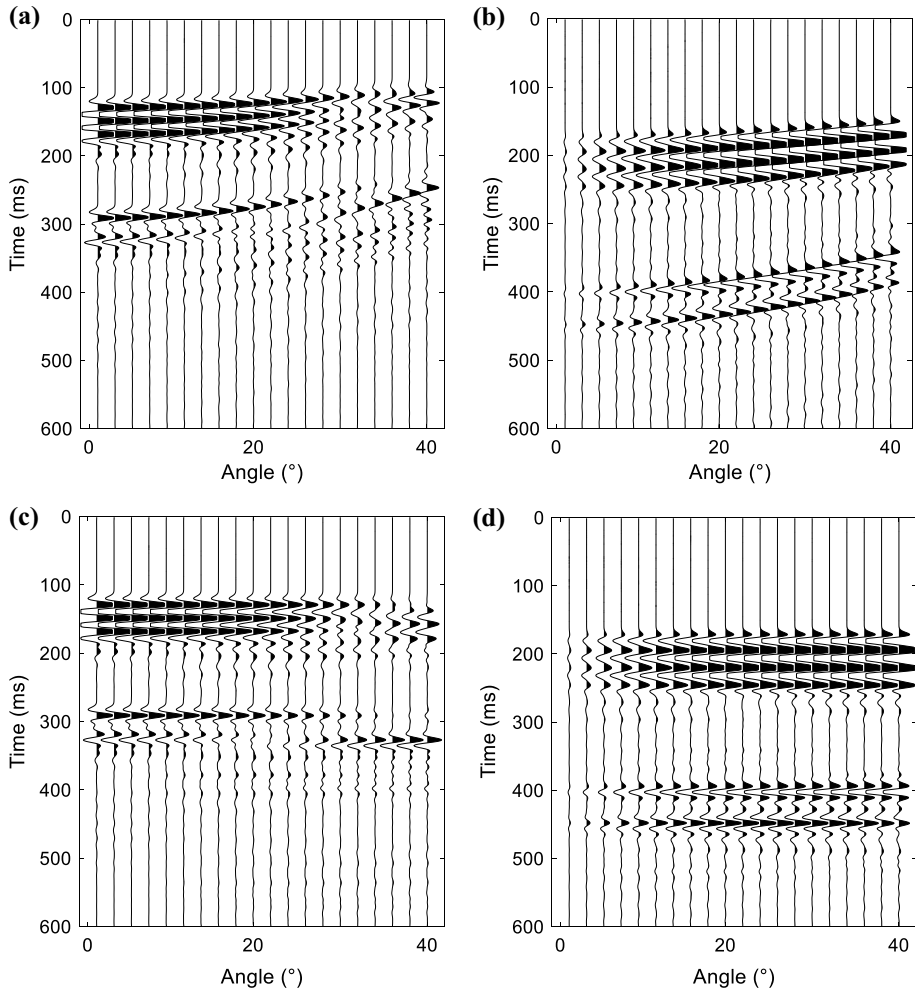


Fig. 11 Synthetic seismograms obtained with the PM for model 3 of Fig. 8b. **a, c** PP seismograms in PP time domain **a** without and **c** with normal moveout (NMO) correction and transmission loss compensation. **b, d** PS seismograms in PS time domain **b** without and **d** with NMO correction and transmission loss compensation

between the logging and seismic scales. The original well-log curves and the Backus-averaged parameters are shown in Fig. 13. The input data are the full wave-field response obtained with the PM, including PP (Fig. 14a) and PS wave (Fig. 14b) angle gathers ranging from 1° to 40°. The PMI can obtain the depth-domain inversion results directly, and correspondingly, the joint PP and PS inversion (JPMI) can be implemented without PS-to-PP time alignment. Figure 15a, b shows the inversion results of the PMI using PP and PS data, respectively, whereas Fig. 15c shows the results of the joint PP-PS inversion (JPMI). The optimal weight factors ϕ in Sect. 3.3 are used.

The correlation coefficients (CCs) between the inversion curves and the true values are given in Table 6. The PMI by using the PP wave gives better α and ρ estimations (higher CCs), but poorer β than those obtained with the PS wave. The estimation of

Fig. 12 Inversion results of model 3 in Fig. 8b using the input gathers shown in Fig. 11. **a** The inversion results in time domain by JEZI. **b** The inversion results in depth domain by JPMI. The blue solid, black dotted and red solid lines are the actual parameters, initial models and inverted results, respectively

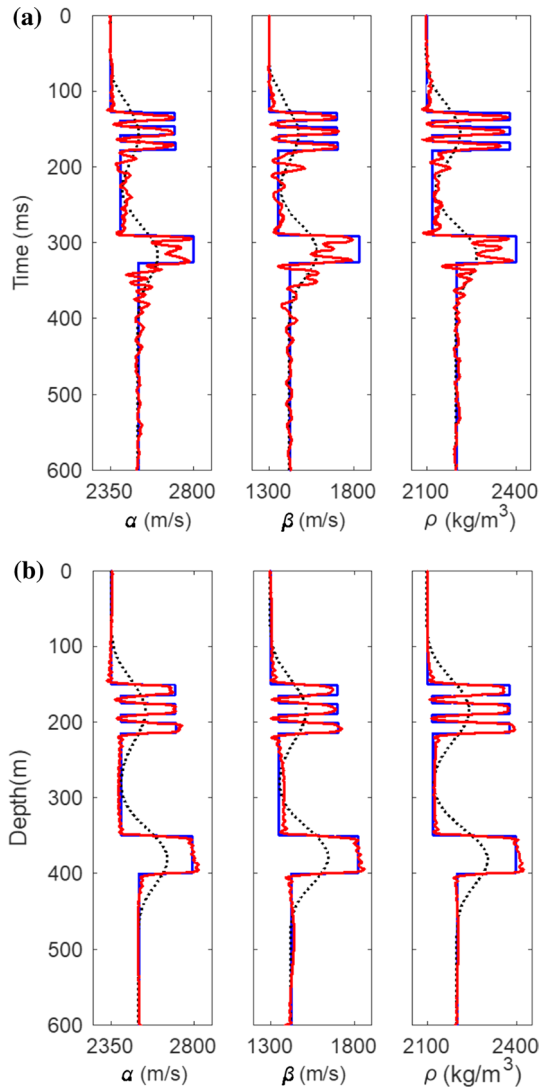


Table 5 Correlation coefficients (CCs) between the inversion results and the true properties of model 3 by using JEZI and JPMI

Method	P-wave velocity α	S-wave velocity β	Density ρ
JEZI	0.8347	0.8306	0.8211
JPMI	0.9731	0.9701	0.9665

the three parameters obtained from the PP or the PS waves is generally worse than that obtained from the joint PP-PS inversion, with values of 0.9749 (P-wave velocity), 0.9701 (S-wave velocity) and 0.9537 (density). Note that the same initial models, which are obtained by smoothing the Backus-averaged model of Fig. 13 (the span of the moving average filter is 120), are used for the three inversions. In Fig. 16, the

Fig. 13 Backus-averaged and original curves of the test well. The green curves correspond to the original well logs and the blue ones to Backus averaging

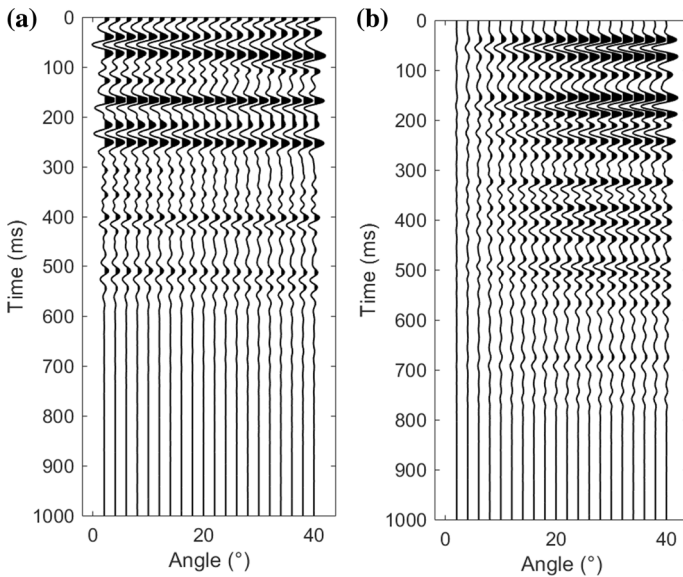
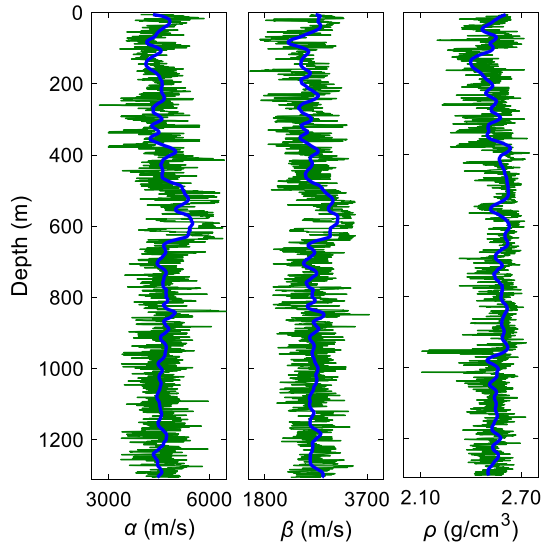


Fig. 14 Synthetic seismograms obtained with the PM for the well-log properties of Fig. 13. **a** PP seismogram in the PP time domain and **b** PS seismogram in the PS time domain, without noise

residual profiles of the PP and PS waves between the input data and the prediction data show the degree of data fitting. Only minor errors can be observed in both residuals, validating the JPMI.

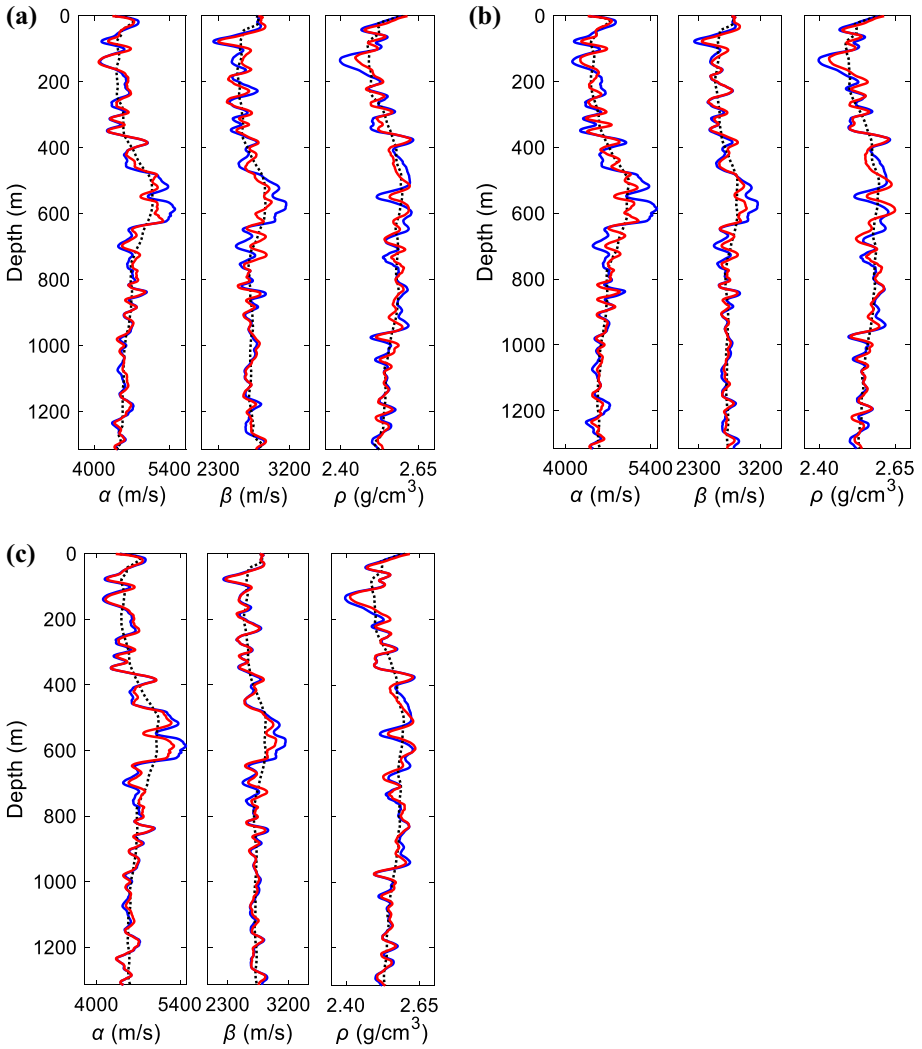


Fig. 15 Depth-domain estimations of the PM-based inversion, including the inversion results, using PP data **a**, PS data **b** and joint PP and PS data **c**. The black dotted, blue and red solid lines correspond to the initial model, actual input properties and inversion results, respectively

Table 6 Correlation coefficients (CCs) between the inversion results and the true properties of the test well by using different input gathers

Input data	P-wave velocity α	S-wave velocity β	Density ρ
Single PP wave	0.9552	0.9285	0.9158
Single PS wave	0.9107	0.9412	0.9097
Joint PP and PS	0.9749	0.9701	0.9537

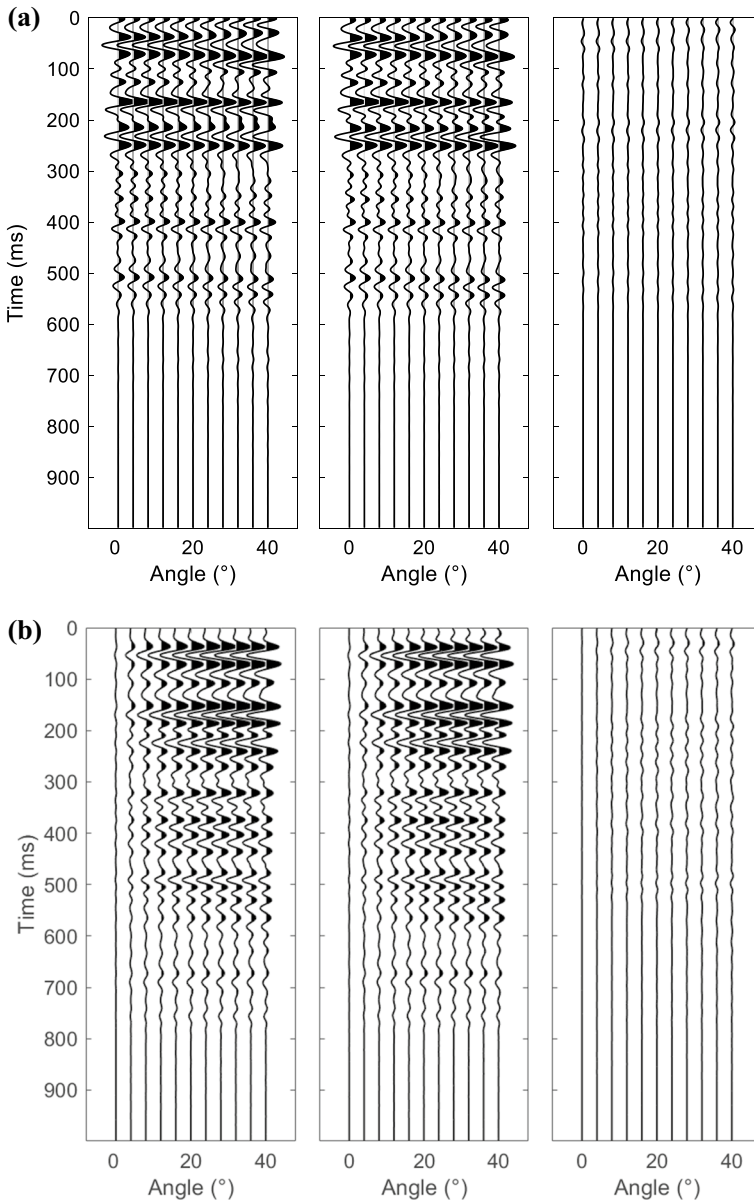


Fig. 16 Residual profiles between the input and the predicted gathers calculated by the joint inversion results of Fig. 15c. **a** PP-wave gather and **b** PS-wave gather. The left panels are the input angle gathers, the middle panels are the predicted gathers by using the joint PP and PS data inversion results of Fig. 15c, and the right panels are the residual profiles

4.4 Time Window Impact of Input Data

In a multi-component inversion, when using wave-equation-based algorithms, choosing suitable time windows (the input time ranges of the prestack data) for PP and PS data

is essential for accuracy. Based on the previous well model (Fig. 13) and the full-wave-response input gathers (Fig. 14), we set up three kinds of time windows of input gathers to test the effect of the time window settings on the inversion results. The three time windows are set to 0–600 ms for PP and 0–750 ms for PS (window 1), 0–400 ms for PP and 0–600 ms for PS (window 2), and 100–1000 ms for PP and 150–1000 ms for PS (window 3). The windows 1 and 2 have the same accurate start time. Window 1 contains all effective primary reflections, but some of the multi-reflections from deep interfaces are cut off. Window 2 includes most of the effective reflections, but the primary and multiple reflections from deep layers are both removed. Compared with the first two, window 3 is given a wrong start time. We perform the inversion by JPMI and the results

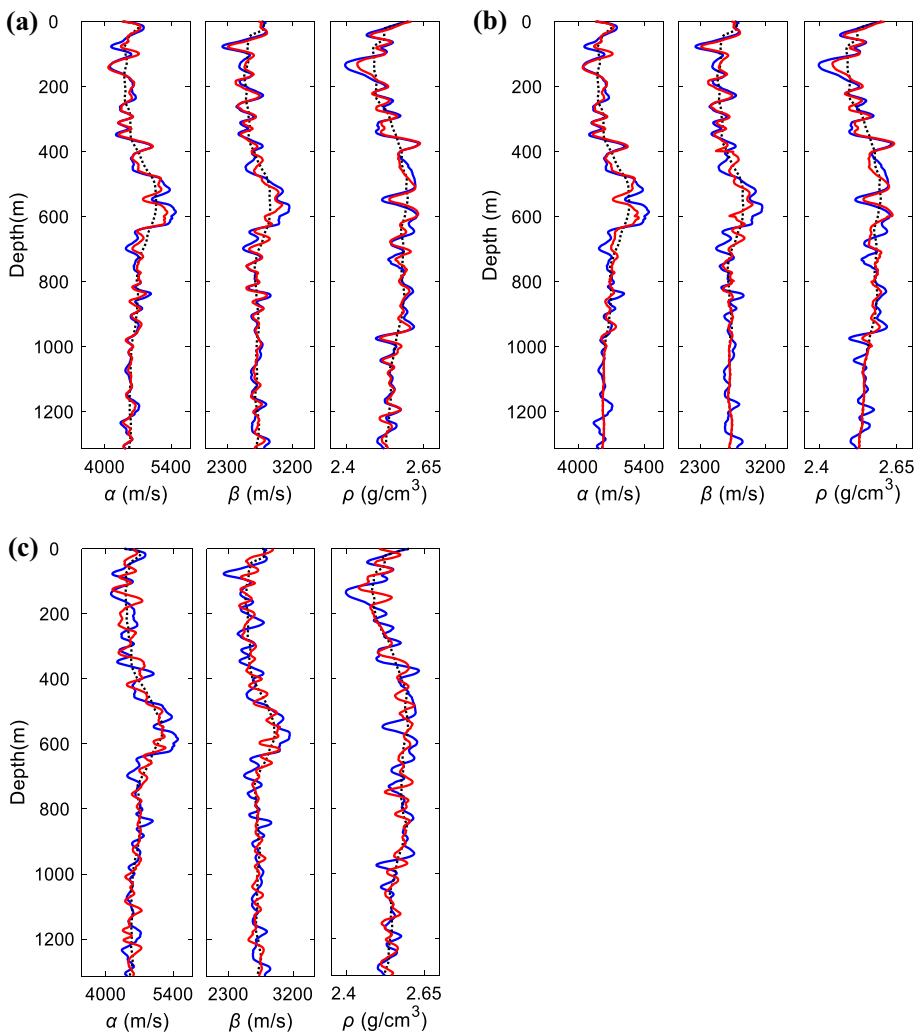


Fig. 17 Depth-domain estimations of the PM-based inversion, including the inversion results, by using the input gathers within time window 1 (a), time window 2 (b) and time window 3 (c). The black dotted, blue and red solid lines correspond to the initial model, actual input properties and inversion results, respectively

Table 7 Correlation coefficients (CCs) between the inversion results using the input data with different time windows and the true properties of the test well

Time window	P-wave velocity α	S-wave velocity β	Density ρ
Time window 1	0.9464	0.9347	0.9083
Time window 2	0.9189	0.9091	0.8918
Time window 3	0.8139	0.8004	0.7582

using the input gathers with different time windows are shown in Fig. 17, and the corresponding CCs between results and true curves are displayed in Table 7. The results of window 1 match well with well models with CCs of 0.9464 (P-wave velocity), 0.9347 (S-wave velocity) and 0.9083 (density). However, compared to the results of the full time window in Fig. 15c, there is a slight decrease in consistency at the deep section (about 1100–1300 m) of window 1. Due to missing part of the primary waves, window 2 cannot obtain valid estimations at the deep section (about 1000–1300 m) and causes lower CCs of 0.9189, 0.9091 and 0.8918. For window 3 with a wrong start time, the results of JPMI are unacceptable, which can be better seen from the poor CCs: 0.8139 (P-wave velocity), 0.8004 (S-wave velocity) and 0.7582 (density). Comparing the three inversion results, we can draw the following conclusions: the time window setting of input data can significantly affect the results. To ensure the accuracy of the inverted results, an accurate start time is needed. One should consider the sufficient length of the windows to include all the primary reflections and as much multi-reflection information as possible, which is helpful in improve the estimation accuracy.

4.5 Analysis of Noise Effects

Gaussian random noise is added to the synthetic gathers to obtain new input gathers, which are noisy angle gathers corresponding to SNR of 50, 10, 5 and 2, respectively. Figure 18a–d shows the inversion results of JPMI by using input gathers with SNR = 50, 10, 5 and 2, respectively. The initial models for the four inversions are all obtained by smoothing the Backus-averaged curves of Fig. 13, as above.

The CCs for the inverted and actual parameters are given in Table 8, and the RMS relative errors are given in Fig. 19. Figure 18a, b indicates that the inverted parameters corresponding to the gathers with SNR 50 and 10 match well with the true ones (blue lines). The CCs of the P-wave velocity are 0.9657 (SNR = 50) and 0.9577 (SNR = 10), those of the S-wave velocity are 0.9603 (SNR = 50) and 0.9519 (SNR = 10), and those of the density are 0.9336 (SNR = 50) and 0.9211 (SNR = 10). Figure 18c shows the inversion results for the gather with SNR = 5. The inverted P- and S-wave velocities for the gathers with SNR = 2 (Fig. 18d) are still acceptable, since the corresponding CCs are 0.9317 (P-wave velocity) and 0.9235 (S-wave velocity), but the CC for the density is 0.8847. Figure 19 shows the RMS errors, where S-wave velocity has the highest error. It shows that the joint inversion improves the estimation of the S-wave velocity. The RMS error of the density is the lowest, probably because density has the smallest variation among the three parameters. This analysis shows that JPMI is stable and reliable under high-noise levels. The comparison of the results also shows that decreasing the noise level leads to more accurate estimations. The CCs in Table 8 show that density is more sensitive to noise than the P- and S-wave velocities.

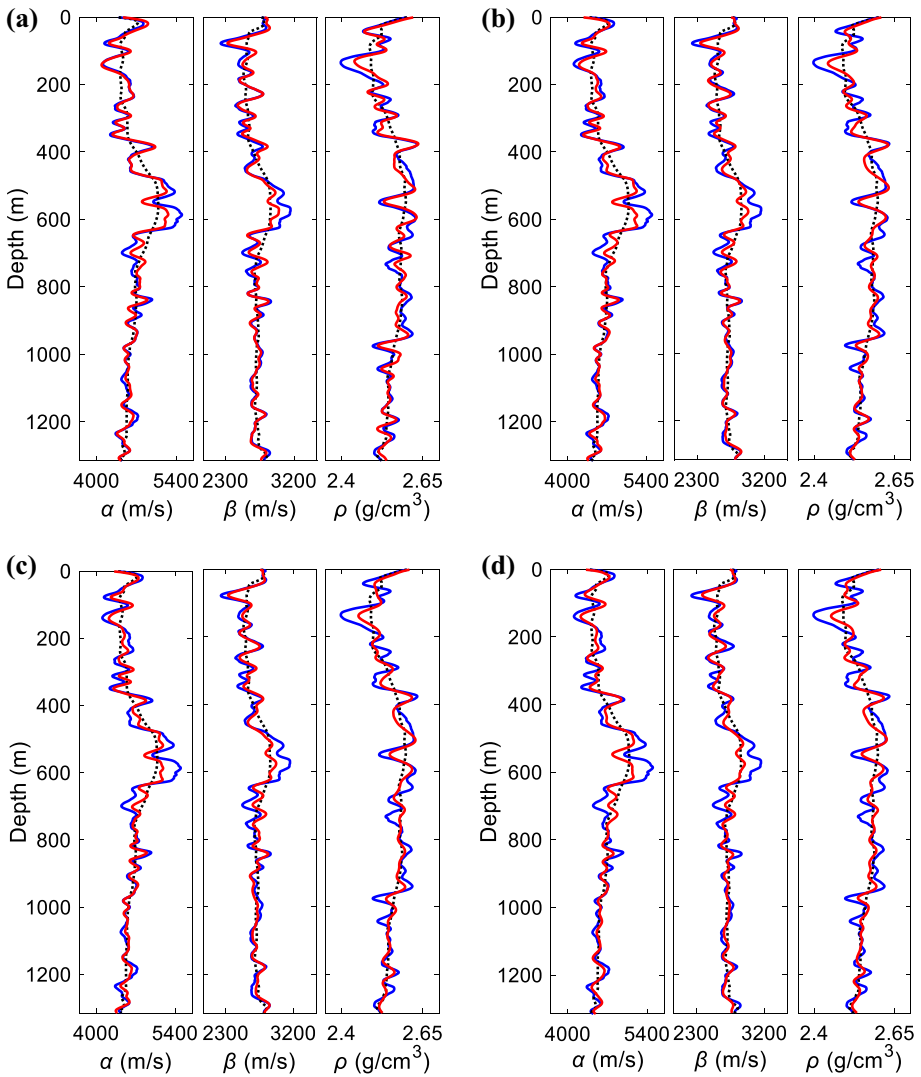
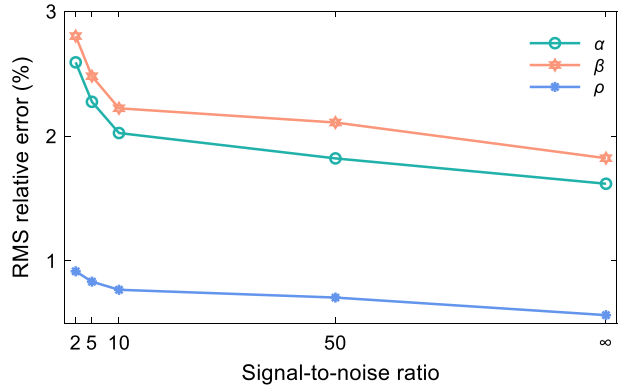


Fig. 18 The inversion results by using the noisy angle gathers. **a** SNR=50, **b** SNR=10, **c** SNR=5, **d** SNR=2. The initial, true and inverted curves are given by the black dotted, blue solid and red solid lines, respectively

Table 8 Correlation coefficients (CCs) between the joint inversion results with different noise levels and true properties of the test well

SNR	P-wave velocity α	S-wave velocity β	Density ρ
∞	0.9749	0.9701	0.9537
50	0.9657	0.9603	0.9336
10	0.9577	0.9519	0.9211
5	0.9475	0.9401	0.9086
2	0.9317	0.9235	0.8847

Fig. 19 RMS relative errors of the inverted P- and S-wave velocities and density by using the JPMI for different noise levels



4.6 Dependence on the Initial Model

Two types of models are used to test the dependence of the joint inversion on the initial models. One type is obtained by smoothing the original curves (the Backus-averaged curves). We obtain four smoothing initial models by changing the smoothing parameter (SP, the span of the moving average filter) to 50, 100, 150 and 200. Smaller SP describes better the true model. The other type is a linear model obtained by interpolation using the first and last sample points.

The inversion results for the five initial models are shown in Fig. 20, and the CCs are given in Table 9. The input noiseless gathers contain full-wave responses without PS-to-PP time alignment process in Fig. 14. If $SP = 50$, there is a good estimation of the three parameters (see Fig. 20a). The results match the actual curves well, and the CCs are 0.9844 (P-wave velocity), 0.9816 (S-wave velocity) and 0.9682 (density). Figure 20b ($SP = 100$) and Fig. 20c ($SP = 200$) show that the CCs decrease as SP increases. The estimation by using a linear initial model has the lowest quality, as shown in Fig. 20d. The RMS relative errors corresponding to the different models are shown in Fig. 21. The JPMI has the smaller dependence on the initial model, since the results with large SP agree well with the true models.

When linearizing the forward operator in Eq. (17), we assume that the terms containing high-order derivatives are negligible that it is valid when the initial model is close to the correct one. However, it is important to consider that how close the given initial guess of the model should be to the real one to guarantee the validation. A two-layer model is considered to show the effect of the initial model to the data perturbation Δd . Changing the parameters of the lower layer, we obtain the data perturbations by using the PM modeling. Figure 22a shows the comparison between the true data perturbation and the approximated one obtained by Eq. (17). The difference between the two curves (Fig. 22b) shows that the errors of the data perturbation can be caused by an inaccurate initial model and increase with the increasing deviations of the models. Therefore, a suitable initial model is important in providing effective low-frequency components to reduce the number of local extrema, reducing the errors of the computed data perturbations and thereby improving the accuracy of the solutions, since the update directions of the gradient-based methods depend on those perturbations.

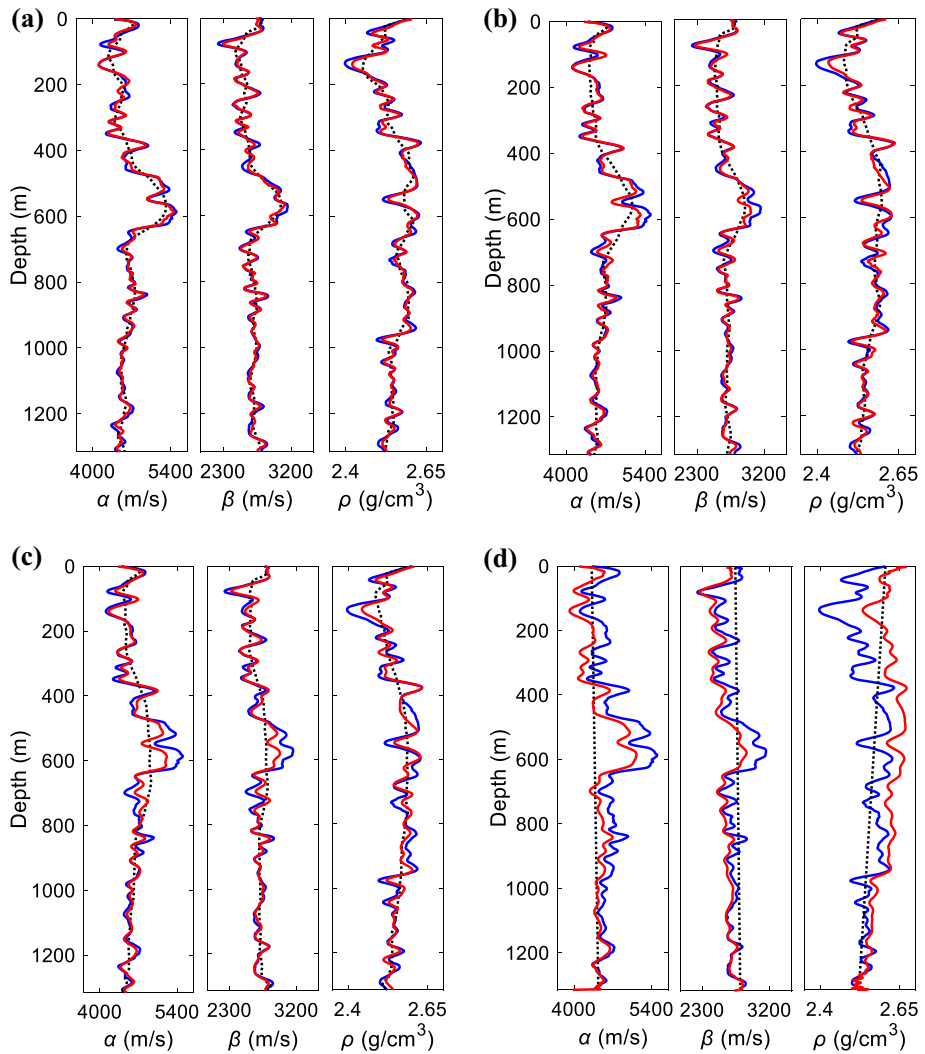


Fig. 20 The inversion results by using different initial models. **a** SP=50, **b** SP=100, **c** SP=200, **d** The linear initial model. The initial, true and inverted curves are represented by the black dotted, blue solid and red solid lines, respectively

Table 9 Correlation coefficients (CCs) between the joint inversion results by using different initial models and true properties of the test well

Initial model	P-wave velocity α	S-wave velocity β	Density ρ
SP=50	0.9844	0.9816	0.9682
SP=100	0.9779	0.9721	0.9590
SP=150	0.9664	0.9605	0.9473
SP=200	0.9513	0.9465	0.9276
Linear	0.9162	0.8801	0.7898

Fig. 21 RMS relative errors for the inverted P- and S-wave velocities and density using the JPMI in the case of different initial models

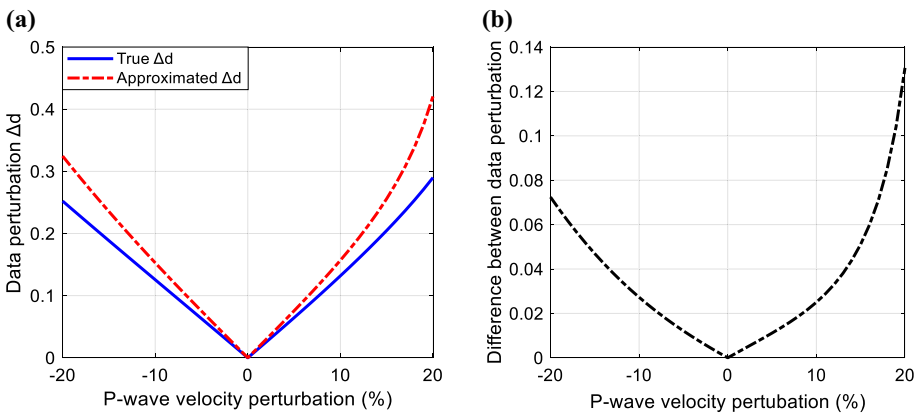
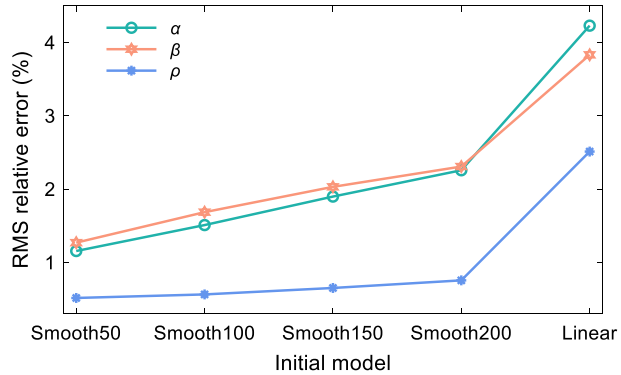


Fig. 22 **a** The comparison between the true data perturbation (blue line) and the approximated one (red dash); **b** the difference between the two perturbations

5 Real-Data Application

A joint inversion is performed on a set of seismic data from the oilfield of Eastern China. The target section ranges from about 800 to 900 ms with a relatively stable geological structure. The top and bottom interfaces of target formation are highlighted by blue and black lines in Fig. 23a. Dominated by fluvial facies, the reservoir is a set of interbedded sand-mudstone layers of multi-stage river channels and mainly develops sand bodies with thicknesses of 2–10 m. The data consist of 86 angle gathers (CDP from 430 to 515) with NMO correction, and the angle range from 1° to 39° for PP and 3°–30° for PS. The input PP and PS datasets have been compensated the transmission loss and suppressed the surface multiples and contain the primary reflections and most of internal multi-reflections. Figure 23 displays the stacked profiles and angle gathers at well position. Figure 23a, c shows the PP data in the PP time domain, and Fig. 23b, d is the PS data in the PS time domain. Figure 24 shows the logs of a well drilled at the Common-Depth Point (CDP) 468, including the actual (blue lines) and the Backus-averaged (black lines) P-wave velocity, S-wave velocity and density, as a function of depth. Using well logs and pre-stack gathers, we estimate the angle-dependent wavelets for the inversion. The previous analysis in Sect. 4.4 indicates that the time window setting of the input data

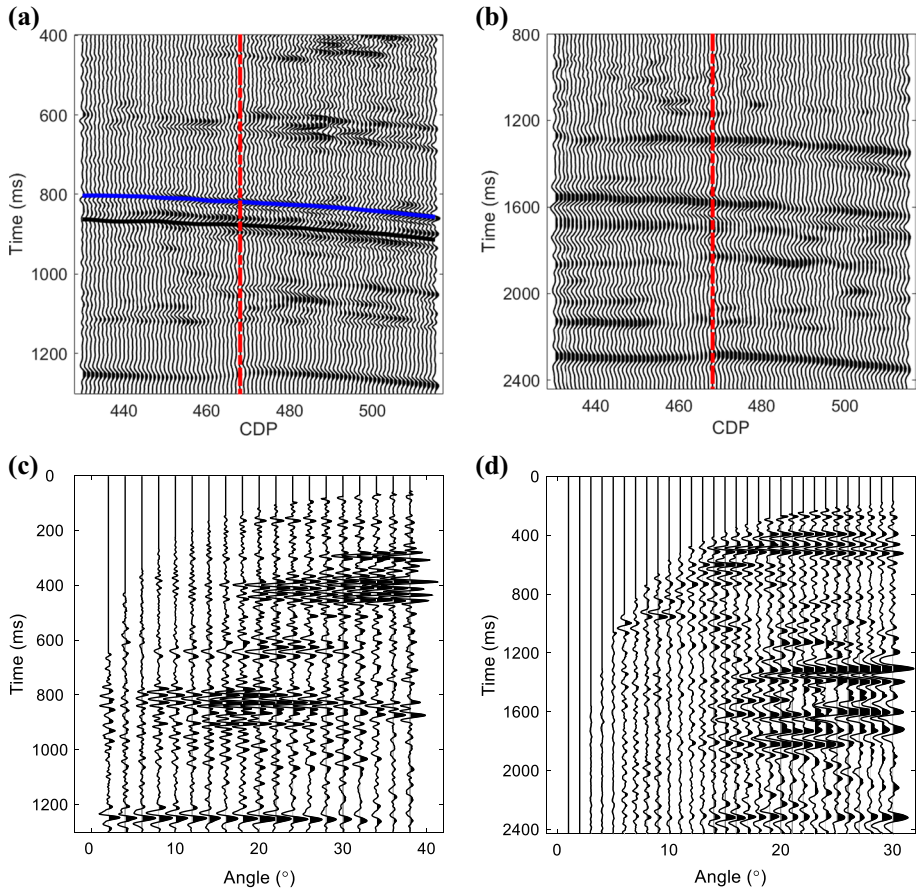
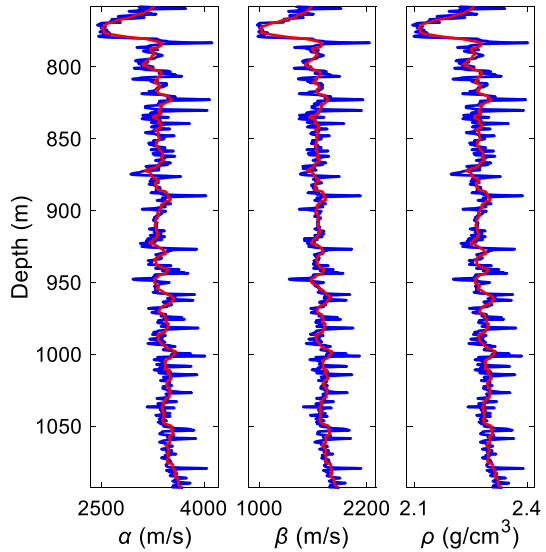


Fig. 23 **a, b** The stacked profiles and **c, d** the angle gathers. **a, c** The PP data in PP time domain. **b, d** The PS data in PS time domain. A well is drilled at CDP 468. The top and bottom interfaces of the target formation have been marked by the blue and black curves in **a**

is important to the inversion results. Therefore, in actual data applications, a feasible way is required to set the correct windows of the PP and PS input gathers. Here, like the pre-processing for ray-tracing-based inversion, we firstly adopt a depth-time matching at well position to determine the time window of PP gathers. Secondly, the dynamic time warping (DTW) algorithm (Hale 2009) is utilized to match the PP and PS waves at well position. Different from the conventional PP and PS registration, without compressing the PS data, the time matching is only used to determine the time window of PS waves. This process can provide relatively accurate windows of input data, which is helpful in reducing the errors. PMI and JPMI are used in three-parameter inversion. The L-BFGS linear optimization scheme is used to obtain the solutions. In this work, we test different dataset weights by carrying out the joint inversion at the well location to determine relatively optimal dataset factors for the three parameters. Then, these factors are adopted for all the CDPs. Since the SNR of the PS data is lower than that of PP data, to mitigate the effect of noise, the final dataset weights are set as 0.8 (P-wave velocity), 0.5 (S-wave velocity) and 0.7 (density) which are higher than those used in noise-free inversion tests. The optimal regularization parameter λ of Eq. (30) is computed adaptively.

Fig. 24 Well logs at CDP 468. The blue lines are the real well logs, and the black ones are the Backus-averaged logs



The *trivariate Gaussian* is used as the prior distribution, and the covariance matrices are calculated based on the initial models in every iteration for each CDP. Unlike the conventional ray-tracing-based inversion, the PM-based inversion requires the initial models in the depth domain. These models are obtained by interpolating the depth-domain Backus-averaged logs along the picked horizons.

Figure 25 shows the 2D inversion profiles for all CDPs by using the PMI (Fig. 25a, c, e) and JPMI (Fig. 25b, d, f). The corresponding well logs are included in the profiles. The JPMI estimations show a better matching and higher horizontal continuity, especially in the deeper formations from 950 to 1100 m. The improvements of the horizontal continuity can be well observed in the areas highlighted by arrows. According to the local conditions, the results obtained with the JPMI are more in agreement with the geological structures.

Figure 26a, b shows the inversion results at the well location by the PMI and JPMI, respectively, by using the seismic traces near the borehole. The PMI using PP data, the P-wave velocity has the best estimation, while the result of the density is not satisfactory. The three parameters have CCs of 0.8897, 0.8012 and 0.7336, and the computed RMS errors are 3.0467, 4.5828 and 1.9050, respectively. Figure 26b shows that the JPMI gives the estimations in better agreement with the well logs. Compared to the PMI, the JPMI improves the inversion accuracy, especially for the S-wave velocity and density. The CCs are 0.9280 (P-wave velocity), 0.9186 (S-wave velocity) and 0.8827 (density), and the RMS errors are 2.6398 m/s (P-wave velocity), 2.9710 m/s (S-wave velocity) and 0.9191 g/cm³ (density). The JPMI shows a better performance, especially for the S-wave velocity and density. The 2D profiles validate the horizontal continuity and the stability of the JPMI.

6 Conclusions

We have developed a joint PP and PS pre-stack seismic inversion algorithm, using a propagator matrix as forward engine, which can be implemented with PP and PS gathers without time alignment and compression. Using the L-BFGS optimization and L-curve adaptive

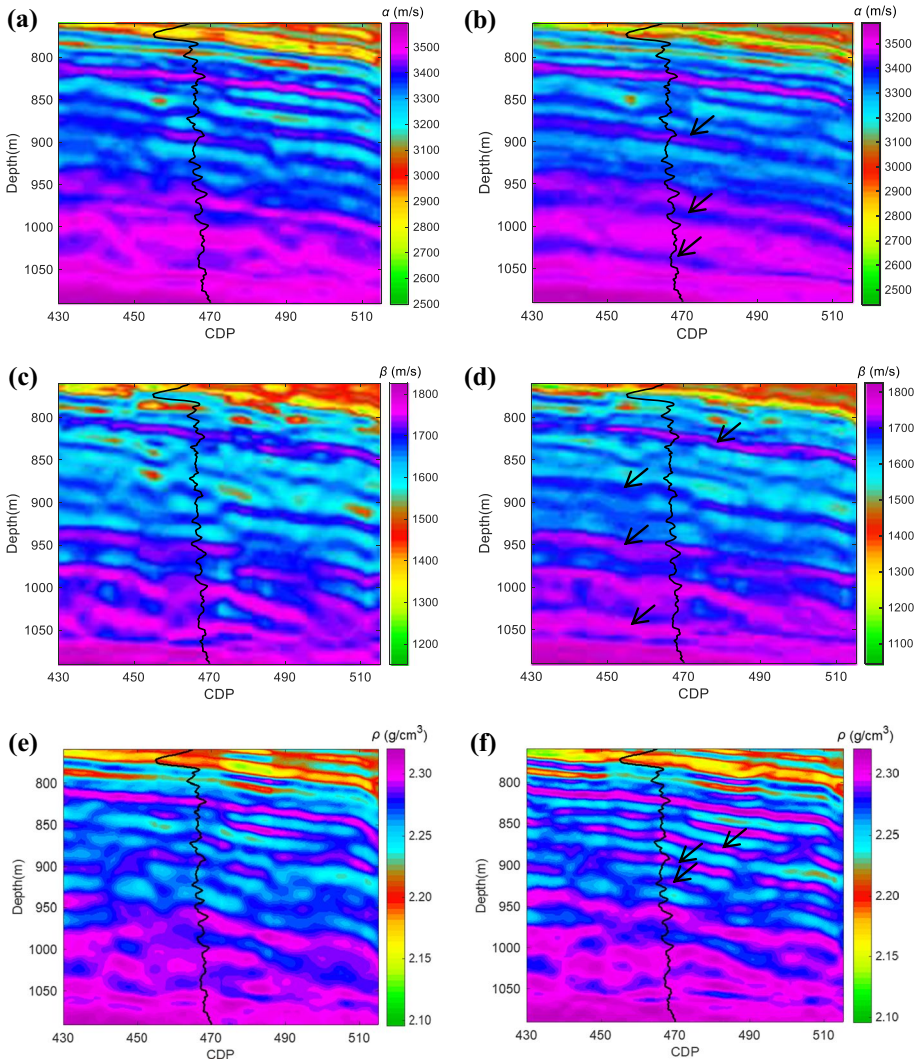
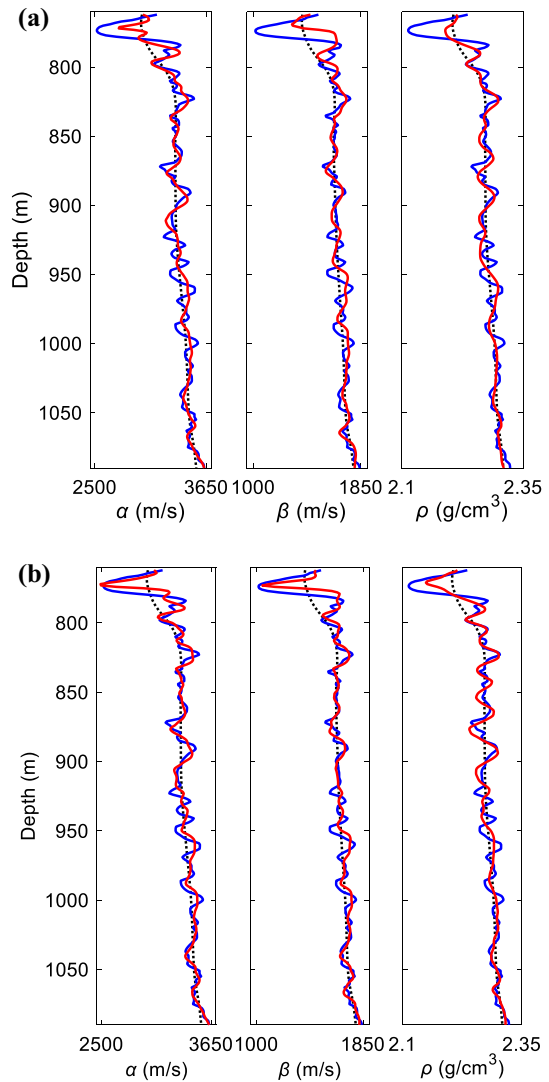


Fig. 25 Depth-domain estimation results, namely, the P-wave velocity α (a, b), S-wave velocity (c, d) and density (e, f), obtained with the PMI (a, c, e) and JPMI (b, d, f). The corresponding real well logs are given

regularization acquisition criterion, the algorithm can reliably invert the P- and S-wave velocities and density.

Forward modeling shows that seismograms obtained with ray tracing contain only primary events, being a rough approximation of the response of stratified media. On the contrary, the novel propagator matrix proposed in this work takes into account the full wavefield response and provides accurate reflection coefficients of multi-layer media. From posterior probability analysis, it is shown that considering the full wavefield improves the sensitivity of the model parameters, compared to the ray-tracing-Zoeppritz approach. Moreover, the joint PP-PS data improve the inversion of the S-wave velocity

Fig. 26 Inversion results at the well location obtained with the PM-based inversions. **a** Direct-depth-domain inversion results of the PMI. **b** Direct-depth-domain inversion results of the JPMI. The blue lines are the real logs, and the black dotted and red curves are the initial logs and the inversion results, respectively



and density, compared to the inversion based on single PP and PS data. Because of the different parameter sensitivities, the weight factor of the objective function should be set separately for the different parameters. The tests show that the proposed method is stable and reliable even at low signal-to-noise ratios, and has a relatively low dependence on the initial models.

Since ray-tracing-based modeling has concise formulas and is easy to implement, the corresponding inversion is suggested for thick-layer models, whose seismograms are slightly contaminated with multiple reflections. On the other hand, wave-equation-based inversion is more suitable in the presence of interbed layers and uses those complex internal responses and simplifies the pre-processing. However, one should note that the PM forward modeling assumes a 1D model and therefore should be used for horizontal or nearly horizontal layers.

Acknowledgements The authors are grateful to the support of the China Postdoctoral Science Foundation (Grant No. 2019M661716), the Fundamental Research Funds for the Central Universities, China (Grant No. B200202135), and the National Natural Science Foundation of China (Grant No. 41974123).

Appendix

From Eq. (2), the partial derivatives of $\mathbf{T}(0)$ and $\mathbf{T}(h)$ with respect to the parameters $m^{*=\alpha,\beta,\rho}$ are

$$\frac{\partial \mathbf{T}(0)}{\partial m^*} = i\omega \begin{bmatrix} \frac{\partial \xi_P}{\partial m^*} & \frac{\partial \xi_S}{\partial m^*} & \frac{\partial \xi_P}{\partial \gamma_P} & \frac{\partial \xi_S}{\partial \gamma_S} \\ -\frac{\partial \eta_P^*}{\partial X_P} & -\frac{\partial \eta_S^*}{\partial X_S} & -\frac{\partial \eta_P^*}{\partial X_P} & -\frac{\partial \eta_S^*}{\partial X_S} \\ \frac{\partial W_P^*}{\partial m^*} & \frac{\partial W_S^*}{\partial m^*} & -\frac{\partial W_P^*}{\partial m^*} & -\frac{\partial W_S^*}{\partial m^*} \end{bmatrix}, \tag{40}$$

and

$$\frac{\partial \mathbf{T}(h)}{\partial m^*} = \begin{bmatrix} i\omega s_P h \xi_P + \frac{\partial \xi_P}{\partial \eta_P^*} & i\omega s_S h \xi_S + \frac{\partial \xi_S}{\partial \eta_S^*} & -i\omega s_P h \xi_P + \frac{\partial \xi_P}{\partial \eta_P^*} & -i\omega s_S h \xi_S + \frac{\partial \xi_S}{\partial \eta_S^*} \\ -i\omega s_P h \gamma_P - \frac{\partial \eta_P^*}{\partial X_P} & -i\omega s_S h \gamma_S - \frac{\partial \eta_S^*}{\partial X_S} & -i\omega s_P h \xi_P + \frac{\partial \xi_P}{\partial \eta_P^*} & -i\omega s_S h \xi_S + \frac{\partial \xi_S}{\partial \eta_S^*} \\ -i\omega s_P h X_P - \frac{\partial \eta_P^*}{\partial X_P} & -i\omega s_S h X_S - \frac{\partial \eta_S^*}{\partial X_S} & i\omega s_P h X_P - \frac{\partial \eta_P^*}{\partial X_P} & i\omega s_S h X_S - \frac{\partial \eta_S^*}{\partial X_S} \\ i\omega s_P h W_P + \frac{\partial W_P^*}{\partial m^*} & i\omega s_S h W_S + \frac{\partial W_S^*}{\partial m^*} & i\omega s_P h W_P - \frac{\partial W_P^*}{\partial m^*} & i\omega s_S h W_S - \frac{\partial W_S^*}{\partial m^*} \end{bmatrix} \cdot \begin{bmatrix} i\omega e^{i\omega s_P h} & & & \\ & i\omega e^{i\omega s_S h} & & \\ & & i\omega e^{-i\omega s_P h} & \\ & & & i\omega e^{-i\omega s_S h} \end{bmatrix} \tag{41}$$

Differentiating Eq. (4) with respect to model parameter $m^{*=\alpha,\beta,\rho}$, we obtain

$$\frac{\partial \xi}{\partial m^*} = \frac{1}{\xi} \left\{ \frac{\left(\alpha s^2 \frac{\partial \alpha}{\partial m^*} + \alpha^2 s \frac{\partial s}{\partial m^*} + \beta p^2 \frac{\partial \beta}{\partial m^*} + \beta^2 p \frac{\partial p}{\partial m^*} \right) \cdot [(\alpha^2 + \beta^2)(p^2 + s^2) - 2]}{[(\alpha^2 + \beta^2)(p^2 + s^2) - 2]^2} - \frac{-(\alpha^2 s^2 + \beta^2 p^2 - 1) \cdot \left[(p^2 + s^2) \left(\alpha \frac{\partial \alpha}{\partial m^*} + \beta \frac{\partial \beta}{\partial m^*} \right) + (\alpha^2 + \beta^2) \left(p \frac{\partial p}{\partial m^*} + s \frac{\partial s}{\partial m^*} \right) \right]}{[(\alpha^2 + \beta^2)(p^2 + s^2) - 2]^2} \right\}, \tag{42}$$

and

$$\frac{\partial \gamma}{\partial m^*} = \frac{1}{\gamma} \left\{ \frac{\left(\alpha p^2 \frac{\partial \alpha}{\partial m^*} + \alpha^2 p \frac{\partial s}{\partial m^*} + \beta s^2 \frac{\partial \beta}{\partial m^*} + \beta^2 s \frac{\partial s}{\partial m^*} \right) \cdot [(\alpha^2 + \beta^2)(p^2 + s^2) - 2]}{[(\alpha^2 + \beta^2)(p^2 + s^2) - 2]^2} - \frac{(\alpha^2 p^2 + \beta^2 s^2 - 1) \cdot \left[(p^2 + s^2) \left(\alpha \frac{\partial \alpha}{\partial m^*} + \beta \frac{\partial \beta}{\partial m^*} \right) + (\alpha^2 + \beta^2) \left(p \frac{\partial p}{\partial m^*} + s \frac{\partial s}{\partial m^*} \right) \right]}{[(\alpha^2 + \beta^2)(p^2 + s^2) - 2]^2} \right\}. \tag{43}$$

Using Eq. (3), the derivatives of the intermediate variables X and Z are

$$\begin{aligned} \frac{\partial X}{\partial m^*} = & 2\alpha\rho(\xi p + s\gamma) \frac{\partial \alpha}{\partial m^*} - 4\beta\rho\xi p \frac{\partial \beta}{\partial m^*} + (\alpha^2\xi p + \alpha^2\gamma s - 2\beta^2\xi p) \frac{\partial \rho}{\partial m^*} \\ & + (\alpha^2 - 2\beta^2)\rho p \frac{\partial \xi}{\partial m^*} + (\alpha^2 - 2\beta^2)\rho\xi \frac{\partial p}{\partial m^*} + \alpha^2\rho s \frac{\partial \gamma}{\partial m^*} + \alpha^2\rho\gamma \frac{\partial s}{\partial m^*}, \end{aligned} \tag{44}$$

and

$$\begin{aligned} \frac{\partial W}{\partial m^*} = & 2\beta\rho(\gamma p + \xi s) \frac{\partial \beta}{\partial m^*} + \beta^2(\gamma p + \xi s) \frac{\partial \rho}{\partial m^*} \\ & + \beta^2\rho \left(p \frac{\partial \gamma}{\partial m^*} + \gamma \frac{\partial p}{\partial m^*} + s \frac{\partial \xi}{\partial m^*} + \xi \frac{\partial s}{\partial m^*} \right). \end{aligned} \tag{45}$$

The vertical slowness s is a function of the model parameters m^* , and the corresponding derivatives are derived according to Eq. (5):

$$\frac{\partial s}{\partial m^*} = \pm \frac{1}{2\sqrt{2}} \cdot \frac{\left(\sqrt{K_1^2 - 4K_2K_3} \mp K_1 \right) \frac{\partial K_1}{\partial m^*} - 2 \frac{\partial K_2}{\partial m^*} K_3 - 2K_2 \frac{\partial K_3}{\partial m^*}}{\sqrt{(K_1^3 - 4K_1K_2K_3) \mp (K_1^2 - 4K_2K_3)^{3/2}}}, \tag{46}$$

where

$$\frac{\partial K_1}{\partial m^*} = -\frac{2}{\alpha^3} \cdot \frac{\partial \alpha}{\partial m^*} - \frac{2}{\beta^3} \cdot \frac{\partial \beta}{\partial m^*} - 4p \frac{\partial \rho}{\partial m^*}, \tag{47a}$$

$$\frac{\partial K_2}{\partial m^*} = 2p \frac{\partial p}{\partial m^*} + \frac{2}{\alpha^3} \cdot \frac{\partial \alpha}{\partial m^*}, \tag{47b}$$

$$\frac{\partial K_3}{\partial m^*} = 2p \frac{\partial p}{\partial m^*} + \frac{2}{\beta^3} \cdot \frac{\partial \beta}{\partial m^*}. \tag{47c}$$

The partial derivatives of P-wave velocity α , S-wave velocity β and density ρ , with respect to the parameters $m^{*=\alpha,\beta,\rho}$ are

$$\frac{\partial \alpha}{\partial \alpha} = 1, \quad \frac{\partial \alpha}{\partial \beta} = 0, \quad \frac{\partial \alpha}{\partial \rho} = 0, \tag{48a}$$

$$\frac{\partial \beta}{\partial \alpha} = 0, \quad \frac{\partial \beta}{\partial \beta} = 1, \quad \frac{\partial \beta}{\partial \rho} = 0, \quad (48b)$$

$$\frac{\partial \rho}{\partial \alpha} = 0, \quad \frac{\partial \rho}{\partial \beta} = 0, \quad \frac{\partial \rho}{\partial \rho} = 1. \quad (48c)$$

Based on the derivatives above, one can obtain the partial derivatives of the reflection coefficients and then the Fréchet derivatives.

References

- Backus G (1962) Long-wave elastic anisotropy produced by horizontal layering. *J Geophys Res* 67(11):4427–4440
- Booth D, Crampin S (1983) The anisotropic reflectivity technique: theory. *Geophys J Int* 72(3):755–766
- Carcione J (1988) Viscoacoustic wave propagation simulation in the earth. *Geophysics* 53(6):769–777
- Carcione J (1995) Constitutive model and wave equations for linear, viscoelastic, anisotropic media. *Geophysics* 60(2):537–548
- Carcione J (1996) Wave propagation in anisotropic, saturated porous media: plane-wave theory and numerical simulation. *J Acoust Soc Am* 99(5):2655–2666
- Carcione J (1997) Reflection and transmission of qP-qS plane waves at a plane boundary between viscoelastically transversely isotropic media. *Geophys J Int* 129(3):669–680
- Carcione J (2001) Amplitude variations with offset of pressure-seal reflections. *Geophysics* 66(1):283–293
- Carcione J (2014) Wave fields in real media. Theory and numerical simulation of wave propagation in anisotropic, anelastic, porous and electromagnetic media. Elsevier, Amsterdam
- Cheng G, Yin X, Zong Z (2018) Third-order AVO inversion for lamé parameter based on inverse operator estimation algorithm. *J Petrol Sci Eng* 164:117–126
- Chopra S, Castagna J (2007) Introduction to this special section—AVO. *Lead Edge* 26(12):1506–1507
- Craven P, Wahba G (1978) Smoothing noisy data with spline functions. *Numer Math* 31(4):377–403
- Debski W, Tarantola A (1995) Information on elastic parameters obtained from the amplitudes of reflected waves. *Geophysics* 60(5):1426–1436
- Fatti JL (1994) Detection of gas in sandstone reservoirs using AVO analysis: a 3-D seismic case history using the Geostack technique. *Geophysics* 59(9):1362–1376
- Fryer G (1980) A slowness approach to the reflectivity method of seismogram synthesis. *Geophys J Int* 63(3):747–758
- Fryer G, Frazer L (1984) Seismic waves in stratified anisotropic media. *Geophys J Roy Astron Soc* 78(3):691–710
- Fuchs K, Müller G (1971) Computation of synthetic seismograms with the reflectivity method and comparison with observations. *Geophys J Int* 23(4):417–433
- Gisolf A, Haffinger P, Hanitzsch C, Doulgeris P, Veeken P (2014) Non-linear full wavefield inversion applied to carboniferous reservoirs in the Wingate gas field (SNS, Offshore UK). In: 76th EAGE conference and exhibition 2014
- Golub GH, Van Matt U (1997) Generalized cross-validation for large-scale problems. *J Comput Graph Stat* 6(1):1–34
- Graebner M (1992) Plane-wave reflection and transmission coefficients for a transversely isotropic solid. *Geophysics* 57(11):1512–1519
- Guo Q, Zhang H, Han F et al (2018) Prestack seismic inversion based on anisotropic Markov random field. *IEEE Trans Geosci Remote Sens* 56(2):1069–1079
- Guo Q, Zhang H, Cao H et al (2020a) Hybrid seismic inversion based on multi-order anisotropic Markov random field. *IEEE Trans Geosci Remote Sens* 58(1):407–420
- Guo Q, Ba J, Luo C et al (2020b) Stability-enhanced prestack seismic inversion using hybrid orthogonal learning particle swarm optimization. *J Petrol Sci Eng* 192:107313
- Hale D (2009) A method for estimating apparent displacement vectors from time-lapse seismic images. *Geophysics* 74(5):V99–V107
- Hansen P (2000) The L-curve and its use in the numerical treatment of inverse problems. In: Johnston P (ed) *Computational inverse problems in electrocardiology*. WIT Press, Southampton, pp 119–142

- Hansen P, O’Leary D (1993) The use of the L-curve in the regularization of discrete ill-posed problems. *SIAM J Sci Comput* 14:1487–1503
- Haskell N (1953) The dispersion of surface waves on multilayered media. *Bull Seismol Soc Am* 43(1):17–34
- Huang G, Chen X, Li J et al (2017a) Application of an adaptive acquisition regularization parameter based on an improved GCV criterion in pre-stack AVO inversion. *J Geophys Eng* 14(1):100–112
- Huang G, Li J, Luo C et al (2017b) Regularisation parameter adaptive selection and its application in the prestack AVO inversion. *Explor Geophys* 49(3):323–335
- Huang G, Chen X, Luo C et al (2018) Application of optimal transport to exact Zoeppritz equation AVO inversion. *IEEE Geosci Remote Sens Lett* 15(9):1337–1341
- Kennett B (1983a) Seismic wave propagation in stratified media. *Advances in applied mechanics*. Cambridge University Press, Cambridge
- Kennett BLN (1983b) Seismic wave propagation in stratified media. Cambridge University Press, Cambridge
- Kennett B, Kerry N (1979) Seismic waves in a stratified half space. *Geophys J Int* 57(3):557–583
- Li T, Mallick S (2013) Prestack waveform inversion of four-component, two-azimuth surface seismic data for orthorhombic elastic media parameters using a nondominated sorting genetic algorithm. *Seg Technical Program Expanded Abstracts*
- Li T, Mallick S (2015) Multicomponent, multi-azimuth pre-stack seismic waveform inversion for azimuthally anisotropic media using a parallel and computationally efficient non-dominated sorting genetic algorithm. *Geophys J Int* 200(2):1136–1154
- Li K, Yin X, Zong Z (2017) Pre-stack Bayesian cascade AVO inversion in complex-Laplace domain and its application to the broadband data acquired at East China. *J Petrol Sci Eng* 158:751–765
- Liu H, Li J, Chen X et al (2016) Amplitude variation with offset inversion using the reflectivity method. *Geophysics* 81(4):R185–R195
- Lu J, Yang Z, Wang Y et al (2015) Joint PP and PS AVO seismic inversion using exact Zoeppritz equations. *Geophysics* 80(5):R239–R250
- Lu J, Wang Y, Chen J et al (2018) Joint anisotropic amplitude variation with offset inversion of PP and PS seismic data. *Geophysics* 83(2):N31–N50
- Luo C, Li X, Huang G (2018a) Hydrocarbon identification by application of improved sparse constrained inverse spectral decomposition to frequency-dependent AVO inversion. *J Geophys Eng* 15(4):1446–1459
- Luo C, Li X, Huang G (2018b) Information on S-wave velocity and density obtained from multi-wave data. *SEG Technical Program Expanded Abstracts*
- Mallick S (1999) Some practical aspects of prestack waveform inversion using a genetic algorithm: an example from the east Texas Woodbine gas sand. *Geophysics* 64(64):326–336
- Mallick S (2007) Amplitude-variation-with-offset, elastic-impedance, and wave-equation synthetics: a modeling study. *Geophysics* 72(1):C1–C7
- Mallick S, Adhikari S (2015) Amplitude-variation-with-offset and prestack-waveform inversion: a direct comparison using a real data example from the Rock Springs Uplift, Wyoming, USA. *Geophysics* 80(2):B45–B59
- Mallick S, Frazer L (1987) Practical aspects of reflectivity modeling. *Geophysics* 52(10):1355–1364
- Mallick S, Frazer L (1990) Computation of synthetic seismograms for stratified azimuthally anisotropic media. *J Geophys Res* 95:8513–8526
- Ostrander W (1984) Plane-wave reflection coefficients for gas sands at nonnormal angles of incidence. *Geophysics* 49(10):1637–1648
- Padhi A, Mallick S (2013) Accurate estimation of density from the inversion of multicomponent prestack seismic waveform data using a nondominated sorting genetic algorithm. *Geophysics* 32(1):94–98
- Padhi A, Mallick S (2014) Multicomponent pre-stack seismic waveform inversion in transversely isotropic media using a non-dominated sorting genetic algorithm. *Geophys J Int* 5(3):589
- Pafeng J, Mallick S, Sharma H (2016) Prestack waveform inversion of three-dimensional seismic data: an example from the Rock Springs Uplift, Wyoming, USA. *Geophysics* 82(1):B1–B12
- Pan X, Zhang G (2018) Model Parameterization and PP-wave amplitude versus angle and azimuth (AVAZ) direct inversion for fracture quasi-weakness in weakly anisotropic elastic media. *Surv Geophys* 39(5):937–964
- Pan X, Zhang G, Yin X (2018a) Azimuthal seismic amplitude variation with offset and azimuth inversion in weakly anisotropic media with orthorhombic symmetry. *Surv Geophys* 39(1):99–123
- Pan X, Zhang G, Yin X (2018b) Elastic impedance variation with angle and azimuth inversion for brittleness and fracture parameters in anisotropic elastic media. *Surv Geophys* 39(5):965–992
- Rüger A (1996) Reflection coefficients and azimuthal AVO analysis in anisotropic media. Golden, Colorado School of Mine

- Sen M, Roy I (2003) Computation of differential seismograms and iteration adaptive regularization in pre-stack waveform inversion. *Geophysics* 68(6):2026–2039
- Sen M, Stoffa P (1991) Nonlinear seismic waveform inversion in one dimension using simulated annealing. *Geophysics* 56(10):1624–1638
- Shuey R (1985) A simplification of the Zoeppritz equations. *Geophysics* 50(4):609–614
- Simmons J, Backus M (1994) AVO modeling and the locally converted shear wave. *Geophysics* 59(8):1237–1248
- Thomson W (1950) Transmission of elastic waves through a stratified solid medium. *J Appl Phys* 21(2):89–93
- Wahba G (1990) Spline models for observational data. *Technometrics* 34(1):113–114
- Yang Y, Engquist B (2018) Analysis of optimal transport and related misfit functions in full-waveform inversion. *Geophysics* 83(1):A7–A12
- Yang Y, Engquist B, Sun J, Hamfeldt BF (2018) Application of optimal transport and the quadratic Wasserstein metric to full-waveform inversion. *Geophysics* 83(1):R43–R62
- Zhang F, Dai R, Liu H (2015) High order approximation for scattering matrix in layered elastic medium and its application in pre-stack seismic inversion. *J Petrol Sci Eng* 131:210–217
- Zhi L, Chen S, Li X (2016) Amplitude variation with angle inversion using the exact Zoeppritz equations—theory and methodology. *Geophysics* 81(2):N1–N15
- Zoeppritz K, Erdbebenwellen V (1919) On the reflection and penetration of seismic waves through unstable layers. *Göttinger Nachrichten* 1:66–84
- Zong Z, Yin X, Wu G (2013) Elastic impedance parameterization and inversion with Young's modulus and Poisson's ratio. *Geophysics* 78(6):N35–N42
- Zong Z, Yin X, Wu G (2015) Geofluid discrimination incorporating poroelasticity and seismic reflection inversion. *Surv Geophys* 36(5):659–681

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.