Full-waveform adjoint Q tomography in viscoelastic medium with central-frequency measurements

Wenyong Pan[1], Yanfei Wang[1,2,3], and Kristopher A. Innanen[4]

[1]Key Laboratory of Petroleum Resource Research, Institute of Geology and Geophysics, Chinese Academy of Sciences, Beijing 100029, P. R. China
[2]University of Chinese Academy of Sciences, Beijing 100049, P. R. China
[3]Innovation Academy of Earth Science, Chinese Academy of Sciences, Beijing 100029, P. R. China
[4]Department of Geoscience, University of Calgary, Calgary, AB T2N 1N4 Canada

Key Points:

- An effective formulation based on the adjoint-state method is developed for viscoelastic full-waveform adjoint tomography
- Multiparameter Hessian-vector products are used to evaluate the resolving abilities of different misfit functions for Q perturbations
- The new central-frequency misfit function can invert for the Q_P and Q_S models more accurately by reducing the trade-off uncertainties

Corresponding author: Yanfei Wang, yfwang@mail.iggcas.ac.cn

Abstract

Accurate Q (quality factor) structures can provide important constraints for characterizing subsurface hydrocarbon/water resources in exploration geophysics and interpreting tectonic evolution of the Earth in earthquake seismology. The attenuation effects on seismic amplitudes and phases can be included in forward and inverse modeling by invoking a generalized standard linear solid rheology. Compared to traditional ray-based methods, full-waveform adjoint tomography, which is based on numerical solutions of the visco-elastodynamic wave equation, has the potential to provide more accurate Q models. However, applications of adjoint Q tomography are impeded by the computational complexity of Q sensitivity kernels, and by strong velocity-Q trade-offs. In this study, following the adjoint-state method, we show that the Q (P and S wave quality factors Q_P and Q_S) sensitivity kernels can be constructed efficiently with adjoint memory strain variables. A novel central-frequency difference misfit function is designed to reduce the trade-off artifacts for adjoint Q tomography. Compared to traditional waveform-difference misfit function, this misfit function is less sensitive to velocity variations, and thus is expected to produce fewer trade-off uncertainties. The multiparameter Hessian-vector products are calculated to quantify the resolving abilities of different misfit functions. Comparative synthetic examples are given to verify the advantages of this new misfit function for adjoint Q_P and Q_S tomography. We end with a 3D viscoelastic inversion example designed to simulate a distributed acoustic sensing/vertical seismic profile survey for monitoring of CO_2 sequestration.

Plain Language Summary

Subsurface Q (quality factor) distributions are important for exploring natural resources and interpreting Earth's interior. In this study, full-waveform adjoint tomography is developed to improve our ability of mapping subsurface P and S wave quality factors (Q_P and Q_S) in viscoelastic media. With the adjoint-state method, Q sensitivity kernels can be constructed efficiently with adjoint memory strain variables. A new central-frequency difference misfit function is proposed to invert for the Q models. Numerical experiments show that this new misfit function is more sensitive to Q perturbations than standard forms and thus can reconstruct the subsurface Q models more stably by reducing the trade-off artifacts caused by the confusions with seismic velocities.

1 Introduction

As seismic waves propagate through the Earth's interior, partial energy is transformed into heat due to intrinsic attenuation, leading to amplitude dissipation and velocity dispersion of the waveforms (H. Liu et al., 1976). Attenuation effects are commonly quantified with quality factor Q. Subsurface Q variations are associated with crack density, fluid-saturation, partial melt, chemical composition (etc.) of the Earth materials (Mavko & Nur, 1979; Berryman, 1988; Hauksson & Shearer, 2006), and thus maps of Q represent potentially useful interpretive tools. In exploration/monitoring seismology, high-resolution Q profiles can provide complement constraints for characterizing fluid-filled/gas-chimney/gas-hydrate reservoirs (Innanen, 2011; Operto & Miniussi, 2018; Y. Wang et al., 2020); furthermore, knowledge of Q allows compensation procedures to produce high-fidelity seismic images by correcting for amplitude loss and phase distortions (T. Zhu et al., 2014; Shen et al., 2018). In earthquake seismology, joint interpretation of elastic velocities and Q anomalies have been shown to improve our understandings of crust/mantle structures and tectonic evolution (Romanowicz, 1995; Cao & Romanowicz, 2004; H. Zhu et al., 2013; Bao et al., 2016; Debayle et al., 2020).

Most current subsurface Q models are obtained by ray-based tomography methods with spectral-ratio (Bath, 1974), central-frequency (Quan & Harris, 1997) or t^* (Eberhart-Phillips & Chadwick, 2002) measurements. However, these ray-based approaches may

give inaccurate Q values in complex geologic areas due to the inaccurate forward modeling operators. Anelasticity of a geological volume can be approximated using a phenomenological model represented mechanically by a combination of springs and dash spots. The system constructed by several parallel standard linear elements is referred to as the generalized standard linear solid (GSLS) rheology (Carcione et al., 1988a). With the GSLS model, dissipation and dispersion of propagating waves can be simulated by solving a set of differential equations with the superposition of parallel relaxation mechanisms (Robertsson et al., 1994; Blanch et al., 1995). High-resolution Q models with arbitrary spatial variability can in principle be reconstructed through full-waveform adjoint tomography, which has emerged as a powerful tool to obtain subsurface properties in exploration (Tarantola, 1984; Virieux & Operto, 2009; Brossier et al., 2009; Yuan et al., 2015; Operto & Miniussi, 2018; Yao et al., 2018; Pan et al., 2018) and earthquake (Tromp et al., 2005; P. Chen et al., 2007; Tape et al., 2009; Fichtner & Trampert, 2011a; H. Zhu & Tromp, 2013; Tong et al., 2014; M. Chen et al., 2017; Krischer et al., 2018; Dong & Yang, 2020) seismology.

However, the progress of full-waveform adjoint Q tomography is significantly lagging behind, in comparison to that of velocity tomography. The high computational complexity of constructing Q sensitivity kernels, and the strong multiparameter couplings between Q and other elastic properties, are both important reasons for this. Effective inversion strategies and algorithms for reliable Q estimation are still underdeveloped. In this study, we have developed an effective formulation for constructing the Q sensitivity kernels with adjoint memory strain variables, and applied them in the minimization of a new central-frequency misfit function, designed to reduce trade-off artifacts. Developments of these methods are introduced in the following.

In full-waveform adjoint tomography, the sensitivity kernels of elastic properties can be calculated efficiently by cross-correlating the forward and adjoint wavefields based on the adjoint-state method (Q. Liu & Tromp, 2006; Plessix, 2006). However, Q is not explicitly included as a parameter in the rheological bodies of time domain viscoelastic wave equation. Instead, the damping effects are modeled by determining the relaxation parameters through a least-squares inversion to approximate a constant Q (Bohlen, 2002; Fichtner & van Driel, 2014). This complicates the derivation and calculation of Q sensitivity kernels. In previous studies, discrepancies exist between different frameworks for constructing the sensitivity kernels in viscoelastic full-waveform adjoint tomography. Charara et al. (2000) defined the attenuation parameters as the differences between "relaxed" and "unrelaxed" moduli and derived the corresponding sensitivity kernels based on Born approximation (Wu, 1989; Bostock et al., 2001). Assuming constant Q, Tromp et al. (2005) derived the Q sensitivity kernels based on the Kolsky-Futterman model (Kolsky, 1952; Futterman, 1962). In this study, we introduce a different, fully viscoelastic formulation; we apply an adjoint-state method (Q. Liu & Tromp, 2008; Fichtner & van Driel, 2014) to derive viscoelastically-based Q_P and Q_S sensitivity kernels, and show that they can be efficiently calculated via the adjoint memory strain variables.

The problem of interparameter trade-off significantly complicates the multiparameter full-waveform adjoint tomography (Tarantola, 1986; Fichtner & Trampert, 2011b; Operto et al., 2013; Innanen, 2014; Alkhalifah & Plessix, 2014; Pan et al., 2018; Kazei & Alkhalifah, 2019). In viscoelastic media, velocity errors tend to produce strong artifacts in the inverted Q models, which impedes the progress of adjoint Q tomography, increases the inversion uncertainties and may produce misleading interpretations for geophysicists (Brossier, 2011; Keating & Innanen, 2020). Off-diagonal blocks within the inverse multiparameter Hessian accounts for suppressing the trade-off artifacts (Pan et al., 2018). However, approximating the inverse Hessian with a small number of linear conjugate gradient iterations does not seem to reduce the trade-off artifacts obviously. This is a strong motivation to engage in designing specific misfit function that are inherently less-exposed to trade-off errors (Karaoğlu & Romanowicz, 2018; Pan & Wang, 2020). Pan and Wang (2020) showed that the amplitude-based misfit functions can resolve the Q

anomalies more effectively than traditional waveform-difference (WD) misfit function. However, it is difficult to control the absolute seismic amplitudes, which are influenced by many other factors including transmission/reflection loss, source and receiver coupling, etc. Seismic attenuation reduces high-frequencies more rapidly than low-frequencies leading to central-frequency downshifts of the seismic data (Hauge, 1981; Quan & Harris, 1997; T. Zhu et al., 2017), which is more stable and broadly-applicable across the full bandwidth. Noting this, we introduce a new central-frequency difference (CD) misfit function for adjoint Q_P and Q_S tomography in this study. The corresponding new adjoint source is derived for calculating the Q sensitivity kernels. Because central-frequency variation is mainly caused by intrinsic attenuation, the new CD misfit function is relatively insensitive to velocity perturbations and amplitude fluctuations. In order to evaluate the trade-off errors between different physical parameters, traditionally scattering radiation patterns (Tarantola, 1986; Operto et al., 2013; Alkhalifah & Plessix, 2014; Pan et al., 2016) are qualitatively interpreted. In this paper, to evaluate the advantage of our method, we quantify the velocity-Q trade-offs within WD and CD misfit functions using multiparameter Hessian-vector products (Pan et al., 2018; Krischer et al., 2018). With the quantitative and waveform-based measures, the new CD misfit function is expected to be more sensitive to Q variations than the standard WD misfit and produce more accurate Q models.

We justify all conclusions with numerical experiments. First, we analyze the sensitivity of central-frequency corresponding to velocity and Q variations and evaluate the resolving abilities of WD and CD misfit functions for Q perturbations with multiparameter Hessian-vector products. Then, a synthetic inversion example involving isolated velocity-Q anomalies is carried out, followed by a synthetic global tomography example with more realistic structures based on the Rayleigh wave phase speed map (Trampert & Woodhouse, 1995). We have found that the Q sensitivity kernels and inverted Q models obtained by WD misfit function are significantly damaged by the mappings from velocity errors. However, when the new CD misfit function is applied instead, the Q anomalies in the sensitivity kernels and inverted models can be identified more clearly suffering from fewer trade-off artifacts. Finally, we carry out a 3D synthetic data inversion experiment, designed to reflect a realistic CO_2 injection and storage monitoring effort (Yu et al., 2019). Distributed acoustic sensing (DAS) has emerged as a popular technology for seismic acquisition with the advantages of low-cost, dense sampling, higher sensitivity to lower frequencies, etc., which make it particularly appropriate for permanent deployment in borehole environments (Cox et al., 2012; Daley et al., 2016; Ajo-Franklin et al., 2019). DAS data can be incorporated relatively straightforwardly into seismic tomography algorithms, either in isolation or combination with standard geophone data (Eaid et al., 2020). In this study, we apply the proposed algorithms to 3D viscoelastic full-waveform adjoint tomography with DAS data acquired in a vertical seismic profile survey (DAS-VSP) configuration. These numerical experiments act as a feasibility study and are suggestive of the effectiveness and potential advantages of the proposed algorithms for adjoint Q tomography in both earthquake and exploration seismology.

This paper is organized as follows. The basic principle of viscoelastic full-waveform adjoint tomography is first reviewed. Forward modeling problem in viscoelastic medium based on the GSLS model is introduced in the Appendix. Then, viscoelastic sensitivity kernels based on the adjoint-state method are derived. Multiparameter Hessian-vector products and the new CD misfit function are then introduced. In the numerical modeling section, synthetic examples are given to verify the advantages of proposed algorithms.

2 Theory and Methods

2.1 Full-waveform adjoint tomography in viscoelastic medium

In a standard formulation of viscoelastic full-waveform adjoint tomography, parameters representing geological properties of the medium are iteratively updated by minimizing the direct difference between simulated data $\mathbf{u}$ and observed data $\mathbf{d}$, constrained by the visco-elastodynamic wave equation. This is embodied in the classical WD misfit function (Q. Liu & Tromp, 2006; Plessix, 2006):

$$\boldsymbol{\chi}\left(\mathbf{m},\mathbf{u},\boldsymbol{\lambda}\right)=\frac{1}{2}\sum_{\mathbf{x}_r}\int_0^{t'}\left[\mathbf{u}\left(\mathbf{m}\right)-\mathbf{d}\right]^2dt-\int_0^{t'}\int_\Omega\boldsymbol{\lambda}\cdot\left(\rho\partial_t^2\mathbf{u}-\nabla\cdot\mathbf{T}-\boldsymbol{f}^s\right)d\mathbf{x}dt, \tag{1}$$

where $\mathbf{m}$ is the model vector including different physical properties, $\mathbf{x}_r$ indicate the receiver locations, t' is the maximum recording time, Ω indicates the full subsurface volume, $\boldsymbol{\lambda}$ is the Lagrangian multiplier, "$\cdot$" means divergence, $\boldsymbol{f}^s$ is the source term and $\mathbf{T}$ is the stress tensor, determined by the entire history of strain fields via Boltzmann principle (Dahlen & Tromp, 1998; Aki & Richards, 2002):

$$T_{ij}=\kappa\delta_{ij}*\partial_t\varepsilon_{kk}+\mu\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}-\frac{2}{3}\delta_{ij}\delta_{kl}\right)*\partial_t\varepsilon_{kl}, \tag{2}$$

where "$*$" means convolution, δ_{ij} is the Kronecker delta, the subscripts i,j,k and l take on the values of x,y,z, κ and μ are the relaxation functions of bulk and shear moduli:

$$\begin{aligned}\kappa&=\kappa^R\left(1+\frac{1}{P}\sum_{p=1}^{P}\tau_\kappa e^{-t/\tau^{\sigma p}}\right)H\left(t\right),\\ \mu&=\mu^R\left(1+\frac{1}{P}\sum_{p=1}^{P}\tau_\mu e^{-t/\tau^{\sigma p}}\right)H\left(t\right),\end{aligned} \tag{3}$$

where κ^R and μ^R are the "relaxed" bulk and shear moduli at time $t=+\infty$ corresponding to low-frequency limit, P is the maximum number of relaxation mechanisms, H is the Heaviside step function, $\tau^{\sigma p}$ is the stress relaxation time of the pth Maxwell body, τ_κ and τ_μ measure the attenuation strengths of bulk and shear moduli, respectively. Equation (2) is not well-suited for solving initial value problem as the whole history of strain fields are required (Pan & Wang, 2020). The convolutional constitutive relationship can be eliminated by introducing memory variables with the superposition of several parallel relaxation mechanisms based on the generalized standard linear solid (GSLS) model. Details about how this is included in the forward modeling solutions are provided in Appendix A.

Newton optimization framework is the common approach for solving the full-waveform inverse problem. The search direction $\Delta\mathbf{m}$ at qth iteration can be obtained by solving the following Newton linear system:

$$\mathbf{H}_q\Delta\mathbf{m}_q=-K_q, \tag{4}$$

where K_q and $\mathbf{H}_q$ are the sensitivity kernel (or gradient), based on first derivative of the misfit function with respect to the model parameters, and the Hessian, based on second derivative, respectively. In large-scale inverse problems, explicit calculation of the Hessian $\mathbf{H}$ and its inverse are generally too computationally expensive to be possible. In response to this, optimization methods have been developed in which inverse Hessian approximations are used to precondition the gradient updates (Y. Wang & Yuan, 2005; Métivier et al., 2013; He & Wang, 2020). We employ a quasi-Newton l-BFGS method in our study to construct the search direction, following a "two-loop recursion" scheme (Nocedal & Wright, 2006). In the inversion process, the model parameters are updated by

$$\mathbf{m}_{q+1}=\mathbf{m}_q+\alpha_q\Delta\mathbf{m}_q, \tag{5}$$

where α_q is the step length for scaling the model updates and obtained using the backtracking line search method (Nocedal & Wright, 2006).

2.2 Viscoelastic sensitivity kernels

Within the time domain GSLS viscoelastic wave equation, Q parameters are not explicitly assigned to the rheological bodies. Instead, the relaxation parameters are selected to approximate a constant Q via an optimization procedure. This makes it cumbersome to derive the Q sensitivity kernels, and introduces discrepancies of different methods for attenuation estimation. Charara et al. (2000) defined the attenuation parameters as the differences between "relaxed" and "unrelaxed" moduli. The sensitivity kernels of "unrelaxed" moduli and attenuation parameters are derived based on Born approximation; Tromp et al. (2005) derived the Q sensitivity kernels based on the Kolsky-Futterman model (Kolsky, 1952; Futterman, 1962) and frequency domain Born scattering integral. The expressions of Q sensitivity kernels are the same with the moduli sensitivity kernels, but lead to different adjoint source in frequency domain. Pan and Wang (2020) used Tromp's approach to calculate the Q sensitivity kernels in viscoelastic adjoint tomography. However, the Kolsky-Futterman model used to derive the adjoint source is not consistent with the GSLS model used in forward modeling, and this discrepancy may produce additional uncertainty in the inversion results.

Thus, in this study, we make use of a fully self-consistent derivation and obtain the viscoelastic sensitivity kernels directly within the adjoint-state method (Q. Liu & Tromp, 2008; Fichtner & van Driel, 2014). The bulk and shear moduli quality factors (Q_κ and Q_μ) are first incorporated into the convolutional constitutive relation explicitly with the enforcement of $\tau_\kappa = Q_\kappa^{-1}$ and $\tau_\mu = Q_\mu^{-1}$ (Fichtner & van Driel, 2014), which gives

$$\begin{aligned} T_{ij} =& \kappa^R \delta_{ij} \left(1 + \frac{1}{P}\sum_{p=1}^{P} \frac{1}{Q_\kappa} e^{-t/\tau^{\sigma_p}}\right) H(t) * \partial_t \varepsilon_{kk} \\ &+ \mu^R \left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right) \left(1 + \frac{1}{P}\sum_{p=1}^{P} \frac{1}{Q_\mu} e^{-t/\tau^{\sigma_p}}\right) H(t) * \partial_t \varepsilon_{kl}. \end{aligned} \tag{6}$$

Then, inserting equation (6) into equation (1) gives the augmented Lagrangian misfit function as:

$$\begin{aligned} \boldsymbol{\chi}(\mathbf{m}, \mathbf{u}, \boldsymbol{\lambda}) = & \frac{1}{2}\sum_{\mathbf{x}_r} \int_0^{t'} \left[u_i - d_i\right]^2 dt - \int_0^{t'} \int_\Omega \lambda_i \left(\rho \partial_t^2 u_i - f_i^s\right) d\mathbf{x} dt \\ &+ \int_0^{t'} \int_\Omega \lambda_i \partial_j \left[\kappa^R \delta_{ij} \left(1 + \frac{1}{P}\sum_{p=1}^{P} \frac{1}{Q_\kappa} e^{-t/\tau^{\sigma_p}}\right) H(t) * \partial_t \varepsilon_{kk}\right] d\mathbf{x} dt \\ &- \int_0^{t'} \int_\Omega \lambda_i \partial_j \left[\frac{2}{3}\mu^R \delta_{ij} \left(1 + \frac{1}{P}\sum_{p=1}^{P} \frac{1}{Q_\mu} e^{-t/\tau^{\sigma_p}}\right) H(t) * \partial_t \varepsilon_{kk}\right] d\mathbf{x} dt \\ &+ \int_0^{t'} \int_\Omega \lambda_i \partial_j \left[2\mu^R \left(1 + \frac{1}{P}\sum_{p=1}^{P} \frac{1}{Q_\mu} e^{-t/\tau^{\sigma_p}}\right) H(t) * \partial_t \varepsilon_{ij}\right] d\mathbf{x} dt. \end{aligned} \tag{7}$$

Thus, following the adjoint-state method, we can directly derive the viscoelastic sensitivity kernels. Variation of the Lagrangian misfit function (equation (7)) due to the per-

turbations of model properties (Δu_i, $\Delta\rho$, $\Delta\kappa$, $\Delta\mu$, ΔQ_κ and ΔQ_μ) is given by

$$\begin{aligned}
\Delta\boldsymbol{\chi}\left(\mathbf{m},\mathbf{u},\boldsymbol{\lambda}\right) = & \int_0^{t'}\int_\Omega \sum_{\mathbf{x}_r}\left[u_i - d_i\right]\Delta u_i d\mathbf{x}dt \\
& - \int_0^{t'}\int_\Omega \left[\rho\partial_t^2\lambda_i - \partial_j\left(\kappa\delta_{ij}\delta_{kl} * \partial_t\tilde{\varepsilon}_{kl}\right)\right]\Delta u_i d\mathbf{x}dt \\
& + \int_0^{t'}\int_\Omega \partial_j\left[\mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right) * \partial_t\tilde{\varepsilon}_{kl}\right]\Delta u_i d\mathbf{x}dt \\
& - \int_0^{t'}\int_\Omega \Delta\rho\lambda_i\partial_t^2 u_i d\mathbf{x}dt \\
& - \int_0^{t'}\int_\Omega \left(\Delta\kappa * \partial_t\tilde{\varepsilon}_{kk}\right)\varepsilon_{ii} d\mathbf{x}dt \\
& - \int_0^{t'}\int_\Omega \left[\Delta\mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right) * \partial_t\tilde{\varepsilon}_{kl}\right]\varepsilon_{ij} d\mathbf{x}dt \\
& + \int_0^{t'}\int_\Omega \left(\Delta Q_\kappa\frac{\kappa^R}{Q_\kappa^2}\sum_{p=1}^{P}\tau^{\sigma p}\tilde{\epsilon}_{kk}^p\right)\varepsilon_{ii} d\mathbf{x}dt \\
& + \int_0^{t'}\int_\Omega \left[\Delta Q_\mu\frac{\mu^R}{Q_\mu^2}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right)\sum_{p=1}^{P}\tau^{\sigma p}\tilde{\epsilon}_{kl}^p\right]\varepsilon_{ij} d\mathbf{x}dt,
\end{aligned} \tag{8}$$

where $\tilde{\varepsilon}_{ij} = \partial_j\lambda_i$ is the Lagrangian strain field, and $\tilde{\epsilon}_{kl}^p$ are the Lagrangian memory strain variables:

$$\tilde{\epsilon}_{kl}^p = \frac{1}{P\tau^{\sigma p}}e^{-t/\tau^{\sigma p}}H\left(t\right) * \partial_t\tilde{\varepsilon}_{kl}. \tag{9}$$

The Lagrangian is stationary with respect to wavefield perturbation Δu_i in the absence of the perturbations $\Delta\rho$, $\Delta\kappa$, $\Delta\mu$, ΔQ_κ and ΔQ_μ (Q. Liu & Tromp, 2006). Setting the coefficient of Δu_i as zero gives the adjoint-state equation:

$$\rho\partial_t^2\lambda_i - \partial_j\left[\kappa\delta_{ij}\delta_{kl} * \partial_t\tilde{\varepsilon}_{kl} + \mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right) * \partial_t\tilde{\varepsilon}_{kl}\right] = f_{i,\mathrm{WD}}^\dagger, \tag{10}$$

where $f_{i,\mathrm{WD}}^\dagger = -\left(d_i - u_i\right)$ is the adjoint source of WD misfit function. Variation of the Lagrangian misfit function can be reduced to

$$\Delta\boldsymbol{\chi} = -\left(\frac{\Delta\rho}{\rho}K_\rho + \frac{\Delta\kappa}{\kappa}K_\kappa + \frac{\Delta\mu}{\mu}K_\mu + \frac{\Delta Q_\kappa}{Q_\kappa}K_{Q_\kappa} + \frac{\Delta Q_\mu}{Q_\mu}K_{Q_\mu}\right), \tag{11}$$

where K_ρ, K_κ, K_μ, K_{Q_κ} and K_{Q_μ} are the corresponding sensitivity kernels (or gradients) for ρ, κ, μ, Q_κ and Q_μ models, respectively. Their explicit expressions are listed in the following:

$$K_\rho = -\int_0^{t'}\rho u_i^\dagger\partial_t^2 u_i dt, \tag{12a}$$

$$K_\kappa = -\int_0^{t'}\left(\kappa * \partial_t\varepsilon_{kk}^\dagger\right)\varepsilon_{ii} dt, \tag{12b}$$

$$K_\mu = -\int_0^{t'}\left[\mu\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right) * \partial_t\varepsilon_{kl}^\dagger\right]\varepsilon_{ij} dt, \tag{12c}$$

$$K_{Q_\kappa} = \int_0^{t'}\left(\frac{\kappa^R}{Q_\kappa}\sum_{p=1}^{P}\tau^{\sigma p}\epsilon_{kk}^{p,\dagger}\right)\varepsilon_{ii} dt, \tag{12d}$$

$$K_{Q_\mu} = \int_0^{t'}\left[\frac{\mu^R}{Q_\mu}\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right)\sum_{p=1}^{P}\tau^{\sigma p}\epsilon_{kl}^{p,\dagger}\right]\varepsilon_{ij} dt, \tag{12e}$$

where $u_i^\dagger$ is the adjoint displacement field defined as the time-reversed Lagrangian field, $\varepsilon_{kl}^\dagger$ and $\epsilon_{kl}^{p,\dagger}$ are the adjoint strain fields and adjoint memory strain variables. The viscoelastic media is commonly described by the velocity-Q parameterization associated with V_P, V_S, Q_P and Q_S. According to the relations between these different physical parameters (see equations (15) and (16) in Pan and Wang (2020)), the sensitivity kernels for velocity (V_P and V_S) and the corresponding quality factors (Q_P and Q_S) can be derived following the chain rule, as listed in the following:

$$K_{V_P} = -\int_0^{t'} 2\left(\rho V_P^2 \delta_{ij} * \partial_t \varepsilon_{kk}^\dagger - W\kappa^R \delta_{ij} \sum_{p=1}^{P} \tau^{\sigma p} \epsilon_{kk}^{p,\dagger}\right) \varepsilon_{ij} dt, \tag{13a}$$

$$K_{V_S} = -\int_0^{t'} 4\left(\rho V_S^2 * \partial_t \varepsilon_{ij}^\dagger - \rho V_S^2 \delta_{ij} * \partial_t \varepsilon_{kk}^\dagger + \frac{W}{2}\kappa^R \delta_{ij} \sum_{p=1}^{P} \tau^{\sigma p} \epsilon_{kk}^{p,\dagger}\right) \varepsilon_{ij} dt, \tag{13b}$$

$$K_{Q_P} = \frac{V_P^2}{Q_P\left(V_P^2 - V_S^2\right)} \int_0^{t'} \left(\kappa^R \delta_{ij} \sum_{p=1}^{P} \tau^{\sigma p} \epsilon_{kk}^{p,\dagger}\right) \varepsilon_{ij} dt, \tag{13c}$$

$$\begin{aligned} K_{Q_S} = &\frac{2}{Q_S} \int_0^{t'} \left(\mu^R \sum_{p=1}^{P} \tau^{\sigma p} \epsilon_{ij}^{p,\dagger} - \frac{1}{3}\mu^R \delta_{ij} \sum_{p=1}^{P} \tau^{\sigma p} \epsilon_{kk}^{p,\dagger}\right) \varepsilon_{ij} dt \\ &- \frac{V_S^2}{Q_S\left(V_P^2 - V_S^2\right)} \int_0^{t'} \left(\kappa^R \delta_{ij} \sum_{p=1}^{P} \tau^{\sigma p} \epsilon_{kk}^{p,\dagger}\right) \varepsilon_{ij} dt, \end{aligned} \tag{13d}$$

where $\varepsilon_{ij}^\dagger$ and $\epsilon_{ij}^{p,\dagger}$ represent the adjoint strain fields and adjoint memory strain variables, and the coefficient W is

$$W = \frac{\left(Q_S - Q_P\right) V_P^2 V_S^2}{Q_P Q_S \left(V_P^2 - V_S^2\right)^2}. \tag{14}$$

Here, we ignore the influence of density in the inversion process. The velocity sensitivity kernels K_{V_P} and K_{V_S} consist of the terms associated with adjoint stress fields and adjoint memory strain variables, which are different from those in purely elastic media (Tromp et al., 2005; Pan et al., 2018). The Q sensitivity kernels K_{Q_P} and K_{Q_S} are constructed with forward strain fields and adjoint memory strain variables, which characterize the damping effects of seismic waves. Compared to the approach given in Tromp et al. (2005), this formulation is theoretically more complete because the same GSLS model is used in the forward and adjoint simulations.

2.3 Interparameter mapping and Hessian-vector product

The problem of interparameter trade-off is arguably the main practical challenge in the application of multiparameter adjoint tomography, and has been intensively investigated by geophysicists in recent years (Operto et al., 2013; Innanen, 2014; Alkhalifah & Plessix, 2014; Pan et al., 2016). In attenuating media, the inverted Q models can be easily damaged by the trade-off artifacts (Kamei & Pratt, 2013). For instance, considering simultaneous estimation of V_P and Q_P, the Newton linear system (equation (4)) can be written as:

$$\begin{bmatrix} \mathbf{H}_{V_P V_P} & \mathbf{H}_{V_P Q_P} \\ \mathbf{H}_{Q_P V_P} & \mathbf{H}_{Q_P Q_P} \end{bmatrix} \begin{bmatrix} \Delta V_P \\ \Delta Q_P \end{bmatrix} = - \begin{bmatrix} K_{V_P} \\ K_{Q_P} \end{bmatrix}, \tag{15}$$

where the off-diagonal blocks $\mathbf{H}_{V_P Q_P}$ and $\mathbf{H}_{Q_P V_P}$ in multiparameter Hessian consists of mixed Fréchet derivatives of V_P and Q_P and characterize their interparameter couplings. The first-order sensitivity kernels K_{V_P} and K_{Q_P} can be re-formulated as the summation of Hessian-vector products (Pan et al., 2018):

$$K_{V_P} = -\mathbf{H}_{V_P V_P} \Delta V_P - \mathbf{H}_{V_P Q_P} \Delta Q_P, \tag{16a}$$

$$K_{Q_P} = -\mathbf{H}_{Q_P V_P} \Delta V_P - \mathbf{H}_{Q_P Q_P} \Delta Q_P, \tag{16b}$$

where the products of diagonal Hessian blocks $\mathbf{H}_{V_PV_P}$ and $\mathbf{H}_{Q_PQ_P}$ with the model perturbations of ΔV_P and ΔQ_P form the diagonal Hessian sensitivity kernels. The products of off-diagonal Hessian blocks $\mathbf{H}_{Q_PV_P}$ and $\mathbf{H}_{V_PQ_P}$ with the model perturbations of ΔV_P and ΔQ_P form the "second-order" trade-off (or contamination) sensitivity kernels, which serve as the mappings between V_P and Q_P (Pan et al., 2018). Similarly, interparameter trade-offs exist between V_S and Q_S, V_P and V_S, ρ and Q_S, etc., but in this study, we will focus on the trade-offs between velocity and Q.

The trade-offs of different physical parameters are commonly analyzed qualitatively with radiation patterns (Tarantola, 1986; Operto et al., 2013; Alkhalifah & Plessix, 2014; Pan et al., 2016), which measure the scattering amplitudes of seismic waves due to local model perturbations. However, radiation patterns are limited in their ability to incorporate dispersion effects due to Q perturbations (Keating & Innanen, 2020). Multiplying multiparameter Hessian with local model perturbations approximates the matrix columns, which are referred to as point-spread functions and provides one effective tool to quantify the trade-offs (Fichtner & Leeuwen, 2015; Pan et al., 2018; Sager et al., 2018). In this study, we calculate the products of multiparameter Gauss-Newton Hessian with model perturbation vectors to analyze the velocity-Q trade-offs. The Hessian-vector products can be calculated efficiently using the adjoint-state method (Métivier et al., 2013; Pan et al., 2018) (see Appendix B). For instance, product of off-diagonal block $\mathbf{H}_{Q_PV_P}$ with model perturbation ΔV_P can be calculated by

$$\begin{aligned}\mathbf{H}_{Q_PV_P}\Delta V_P = \Bigg\langle & \frac{2V_P^2}{Q_P\left(V_P^2-V_S^2\right)}\left(\kappa^R\delta_{ij}\sum_{p=1}^{P}\tau^{\sigma_p}\epsilon_{kk}^{p,\dagger}\right)\varepsilon_{ij} \\ & \times\left(\rho V_P^2\delta_{i'j'}*\partial_t\varepsilon_{k'k'}^{\dagger}-W\kappa^R\delta_{i'j'}\sum_{p=1}^{P}\tau^{\sigma_p}\epsilon_{k'k'}^{p,\dagger}\right)\Delta V_P\varepsilon_{i'j'}\Bigg\rangle,\end{aligned} \tag{17}$$

where $\langle\cdot\rangle$ means summations of time and space samples for sake of compactness.

2.4 Central-frequency difference misfit function

Previous studies have revealed that when using WD misfit function, velocity errors produce strong trade-off artifacts in the inverted Q models (Mulder & Hak, 2009; Brossier, 2011; Pan & Wang, 2020), with both velocity and Q variations contributing to the waveform residuals. In principle, the inverse Hessian corrects many of the trade-off errors, but, in applying the truncated Newton approach, and using a small number of conjugate gradient iterations, we have observed limited reduction of trade-off artifacts.

Intrinsic attenuation leads to obvious amplitude reduction and central-frequency downshift of the seismic data. Designing specific misfit functions measuring amplitude and central-frequency variations represents one alternative approach to reduce the trade-off artifacts without increasing computational cost. Because seismic amplitudes are also influenced by many other factors including instrument responses, radiation patterns, etc, the amplitude-based misfit functions are easily affected by unexpected amplitude fluctuations. Therefore, in this study, we introduce a new central-frequency difference (CD) misfit function for adjoint Q tomography:

$$\boldsymbol{\chi}_{\mathrm{CD}}(Q)=\frac{1}{2}\sum_{\mathbf{x}_r}\left[f_i^c(Q)-f_{i,\mathrm{obs}}^c\right]^2, \tag{18}$$

where f_i^c and $f_{i,\mathrm{obs}}^c$ indicate the central-frequencies of synthetic and observed data, respectively. We determine central-frequency of the signal by (Berkhout, 1984; Barnes, 1993)

$$f_i^c=\frac{1}{N}\int_0^{+\infty}f\mathcal{A}_i^2(f)\,df, \tag{19}$$

where f indicates frequency, $\mathcal{A}_i$ represents amplitude spectrum of the seismic data and the coefficient N in the denominator is

$$N = \int_0^{+\infty} \mathcal{A}_i^2(f)\, df. \tag{20}$$

Variation of the misfit function with respect to Q perturbation is given by

$$\Delta\boldsymbol{\chi}_{\mathrm{CD}}(Q) = \sum_{\mathbf{x}_r} \left[f_i^c(Q) - f_{i,\mathrm{obs}}^c\right] \Delta f_i^c, \tag{21}$$

where Δf_i^c indicates the variation of central-frequency due to Q perturbation:

$$\begin{aligned}\Delta f_i^c = &\left[\int_0^{+\infty} f\mathcal{A}_i(f)\, df\right] \frac{2\mathcal{R}\left[\tilde{u}_i(f)\,\Delta\tilde{u}_i(f)\right]}{N\mathcal{A}_i(f)} \\ &- \left[\int_0^{+\infty} f\mathcal{A}_i^2(f)\, df \int_0^{+\infty} \mathcal{A}_i(f)\, df\right] \frac{2\mathcal{R}\left[\tilde{u}_i(f)\,\Delta\tilde{u}_i(f)\right]}{N^2\mathcal{A}_i(f)},\end{aligned} \tag{22}$$

where $\tilde{u}_i$ is the synthetic data in frequency domain, the symbol $\mathcal{R}$ means real part and $\Delta\tilde{u}_i$ indicates the data variation caused by Q perturbation. Inserting equation (22) into equation (21) gives:

$$\begin{aligned}\Delta\boldsymbol{\chi}_{\mathrm{CD}} = &\sum_{\mathbf{x}_r} \left[f_i^c - f_{i,\mathrm{obs}}^c\right] \frac{2\mathcal{R}\left[\tilde{u}_i(f)\,\Delta\tilde{u}_i(f)\right]}{N\mathcal{A}_i(f)} \\ &\times \left[\int_0^{+\infty} f\mathcal{A}_i(f)\, df - \frac{1}{N}\int_0^{+\infty} f\mathcal{A}_i^2(f)\, df \int_0^{+\infty} \mathcal{A}_i(f)\, df\right].\end{aligned} \tag{23}$$

Thus, the adjoint source for calculating Q sensitivity kernels using the central-frequency misfit function $\boldsymbol{\chi}_{\mathrm{CD}}$ in frequency domain can be obtained as:

$$\begin{aligned}\tilde{f}_{i,\mathrm{CD}}^{s,\dagger}(f) = &\sum_{\mathbf{x}_r} \frac{2\mathcal{R}\left[\tilde{u}_i(f)\left(f_i^c - f_{i,\mathrm{obs}}^c\right)\right]}{N\mathcal{A}_i(f)} \\ &\times \left[\int_0^{+\infty} f\mathcal{A}_i(f)\, df - \frac{1}{N}\int_0^{+\infty} f\mathcal{A}_i^2(f)\, df \int_0^{+\infty} \mathcal{A}_i(f)\, df\right].\end{aligned} \tag{24}$$

Because central-frequency of the seismic data is mainly controlled by Q variations, the new CD misfit function is expected to invert for the Q models independent of velocity variations. This efficacy can be quantified by applying multiparameter Hessian-vector products to the problem of resolution analysis for Q inversion emerging from this new misfit function.

3 Numerical Experiments

In this section, we first analyze the seismic waveforms with velocity and Q variations and evaluate the sensitivities of different misfit functions for Q perturbations using multiparameter Hessian-vector products. Then, inversion experiments are carried out for full-waveform adjoint Q tomography in 2D/3D viscoelastic medium. The numerical experiments are performed using open-source package SeisElastic2D_1.1 (Pan et al., 2020) based on the spectral-element forward modeling method (Komatitsch & Tromp, 2005).

3.1 Sensitivity analysis of seismic waveforms and multiparameter Hessian-vector products

Seismic forward modeling experiments in viscoelastic media are carried out by perturbing Q and velocity models. The 2D initial viscoelastic model is homogeneous with V_P=2V_S=5.50 km/s, Q_P=Q_S=150 and density ρ=3000 kg/m^3, as shown in Figure 1a.

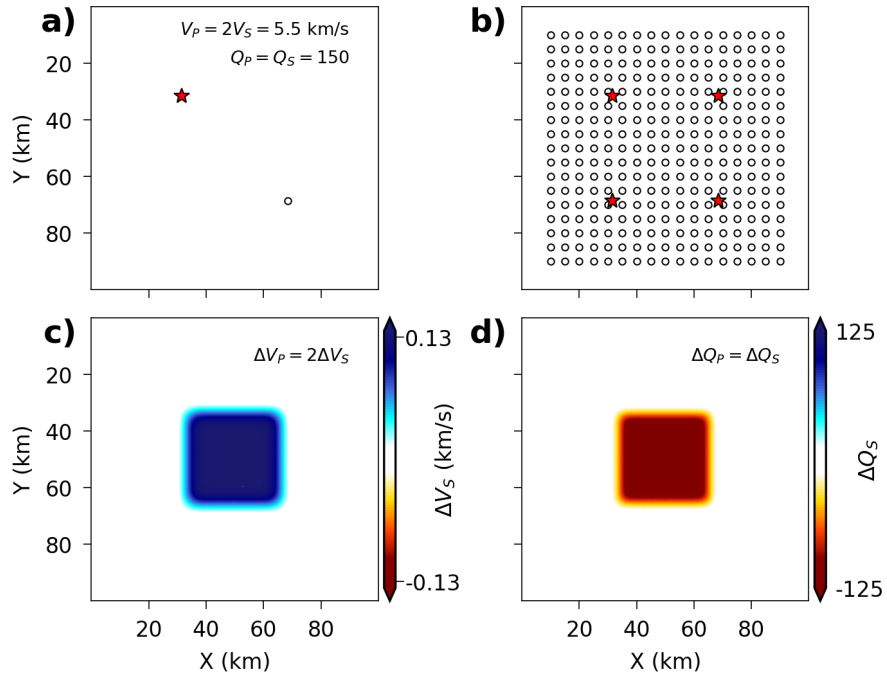

Figure 1. (a) shows the initial homogeneous model for seismic waveform analysis with single source-receiver pair; The source and receiver are arranged at (X=31.5 km, Y=31.5 km) and (X=68.5 km, Y=68.5 km), respectively. (b) shows the acquisition geometry for calculating the multiparameter Hessian-vector products; The red stars and white circles indicate the locations of sources and receivers. Locations of the 4 sources are (X=31.5 km, X=31.5 km), (X=31.5 km, Y=68.5 km), (X=68.5 km, Y=31.5 km) and (X=68.5 km, Y=68.5 km), respectively. (c) and (d) show the perturbation vectors of velocity (ΔV_P=2ΔV_S=0.275 km/s) and Q (ΔQ_P=ΔQ_S=−125).

The maximum sizes of the model in X and Y directions are both 100 km. A Ricker wavelet with dominant frequency of 0.50 Hz is used as the source time function. The initial homogeneous model is used as benchmark for the forward modeling experiments and referred to as Model I. We first change the background Q_P and Q_S values from 150 to 25, which is referred to as Model II. Based on Model II, we apply −5% and −10% perturbations to V_P and V_S models, which are referred to as Model III and IV, respectively. The density model is kept unchanged. Properties of the 4 viscoelastic models are illustrated in Table 3.1.

Then, forward modeling experiments are performed with single source-receiver pair. The recorded vertical (z) component data are plotted for comparison, as shown in Figure 2a. Figures 2b and 2c present the corresponding amplitude spectra of P and S waves, respectively. The central-frequencies calculated using equation (19) are also labeled. As can be seen that when decreasing Q_P and Q_S from 150 to 25, the waveform amplitudes reduce significantly. Central-frequencies of P and S waves are also reduced from 0.46 and 0.45 to 0.42 and 0.38, respectively. For Model III and IV, the waveforms lag obviously due to velocity perturbations. However, the central-frequencies almost do not change, meaning that velocity perturbations produce limited influences on central-frequencies of the seismic data. Similar phenomena can also be observed in horizontal (x) component of the recordings (Figure S1).

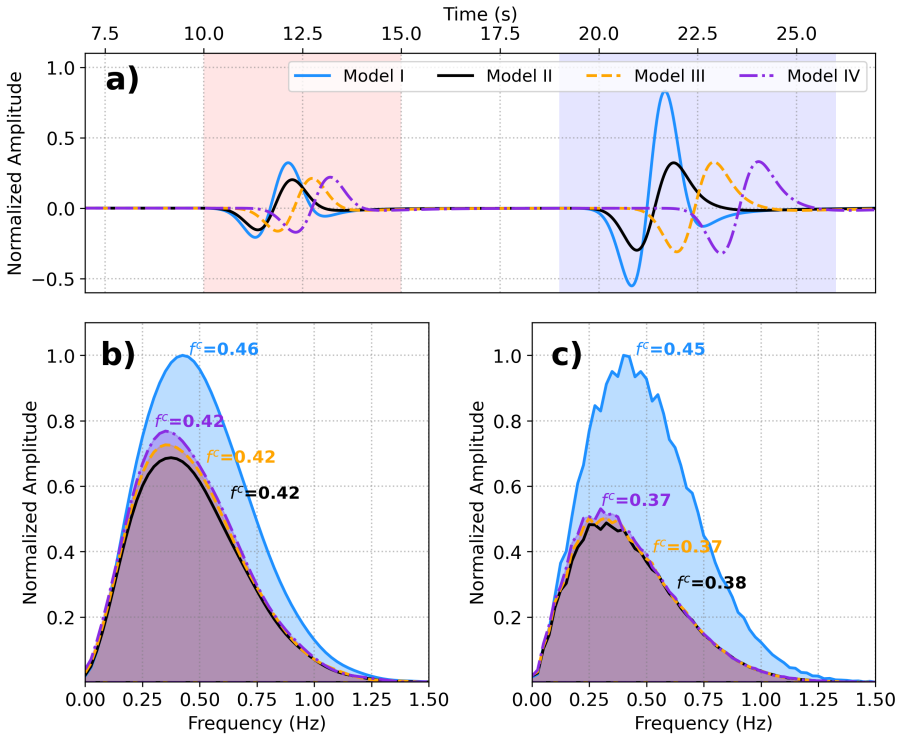

Figure 2. (a) shows the comparison of seismic waveforms (z component) calculated using Model I (solid-blue line), II (solid-black line), III (dashed-orange line) and IV (dash-dotted-purple line). The red and blue shading areas indicate the time windows for extracting P and S waves. (b) shows the corresponding amplitude spectra of P waves; (c) shows the corresponding amplitude spectra of S waves. f^c (Hz) indicate central-frequencies of P and S waves. Colors of the amplitude spectra and f^c are consistent with those of the models.

Table 1. *Properties of the Viscoelastic Models for Forward Modeling*

Models	Velocity (km/s)	Quality factor	Density (kg/m^3)
I	$V_P = 5.50$ $V_S = 2.75$	$Q_P = 150$ $Q_S = 150$	3000
II	$V_P = 5.50$ $V_S = 2.75$	$Q_P = 25$ $Q_S = 25$	3000
III	$V_P = 5.23$ $V_S = 2.61$	$Q_P = 25$ $Q_S = 25$	3000
IV	$V_P = 4.95$ $V_S = 2.48$	$Q_P = 25$ $Q_S = 25$	3000

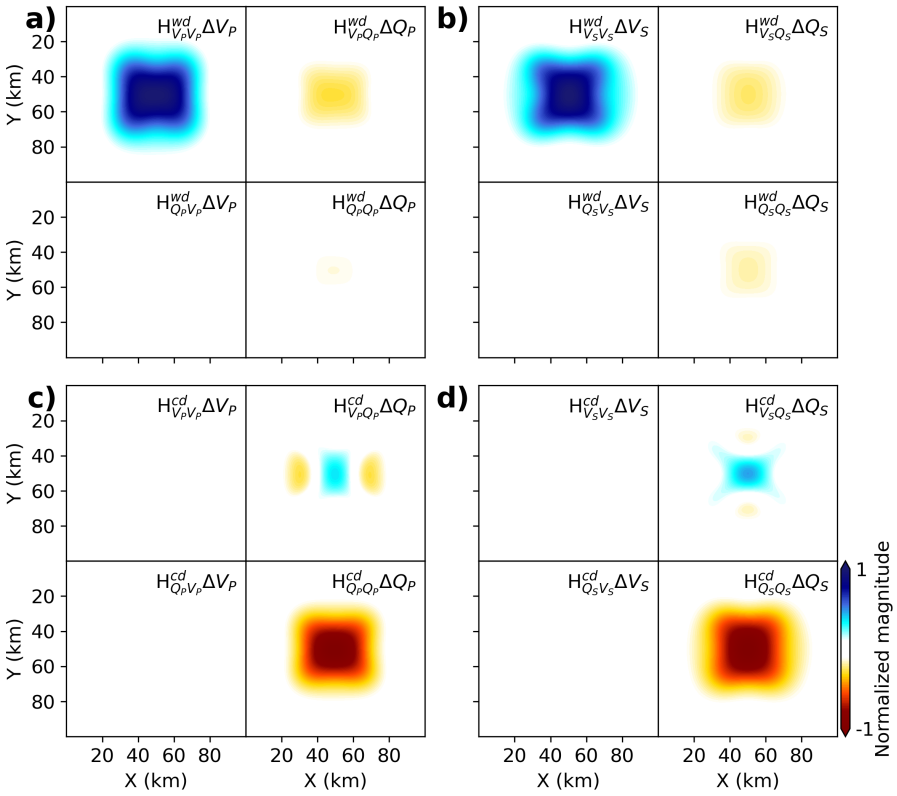

Figure 3. (a) shows the Hessian-vector products ($\mathbf{H}^{\mathrm{wd}}_{V_PV_P}\Delta V_P$, $\mathbf{H}^{\mathrm{wd}}_{V_PQ_P}\Delta Q_P$, $\mathbf{H}^{\mathrm{wd}}_{Q_PV_P}\Delta V_P$ and $\mathbf{H}^{\mathrm{wd}}_{Q_PQ_P}\Delta Q_P$) associated with ΔV_P and ΔQ_P calculated using WD misfit function; (b) shows the Hessian-vector products ($\mathbf{H}^{\mathrm{wd}}_{V_SV_S}\Delta V_S$, $\mathbf{H}^{\mathrm{wd}}_{V_SQ_S}\Delta Q_S$, $\mathbf{H}^{\mathrm{wd}}_{Q_SV_S}\Delta V_S$ and $\mathbf{H}^{\mathrm{wd}}_{Q_SQ_S}\Delta Q_S$) associated with ΔV_S and ΔQ_S calculated using WD misfit function; (c) shows the Hessian-vector products ($\mathbf{H}^{\mathrm{cd}}_{V_PV_P}\Delta V_P$, $\mathbf{H}^{\mathrm{cd}}_{V_PQ_P}\Delta Q_P$, $\mathbf{H}^{\mathrm{cd}}_{Q_PV_P}\Delta V_P$ and $\mathbf{H}^{\mathrm{cd}}_{Q_PQ_P}\Delta Q_P$) associated with ΔV_P and ΔQ_P calculated using CD misfit function; (d) shows the Hessian-vector products ($\mathbf{H}^{\mathrm{cd}}_{V_SV_S}\Delta V_S$, $\mathbf{H}^{\mathrm{cd}}_{V_SQ_S}\Delta Q_S$, $\mathbf{H}^{\mathrm{cd}}_{Q_SV_S}\Delta V_S$ and $\mathbf{H}^{\mathrm{cd}}_{Q_SQ_S}\Delta Q_S$) associated with ΔV_S and ΔQ_S calculated using CD misfit function.

Next, we proceed forward further by calculating the multiparameter Hessian-vector products for inversion sensitivity analysis with different misfit functions. The distributions of 4 sources and 289 receivers are presented in Figure 1b. We apply velocity (-5%) and Q (ΔQ_P=ΔQ_S=-125) perturbations at center parts of the models, as presented in Figures 1c and 1d. The recorded P waves are used to calculate the products of multiparameter Hessian with perturbation vectors ΔV_P and ΔQ_P. The recorded S waves are used to calculate the multiparameter Hessian-vector products associated with ΔV_S and ΔQ_S. The Hessian-vector products calculated using WD and CD misfit functions are presented in Figure 3. When using WD misfit function, the Hessian-vector products $\mathbf{H}^{\mathrm{wd}}_{V_PV_P}\Delta V_P$ and $\mathbf{H}^{\mathrm{wd}}_{V_SV_S}\Delta V_S$ dominate the panels, as shown in Figures 3a and 3b. This means that even for -5% velocity perturbations, the WD misfit function is much more sensitive to velocity perturbations. However, when using CD misfit function, the panels are dominated by Hessian-vector products $\mathbf{H}^{\mathrm{cd}}_{Q_PQ_P}\Delta Q_P$ and $\mathbf{H}^{\mathrm{cd}}_{Q_SQ_S}\Delta Q_S$, suggesting that the CD misfit function is more sensitive to Q perturbations. These observations in the numerical experiments suggest that the new CD misfit function will be more effective to invert for Q models suffering from limited influences of velocity errors.

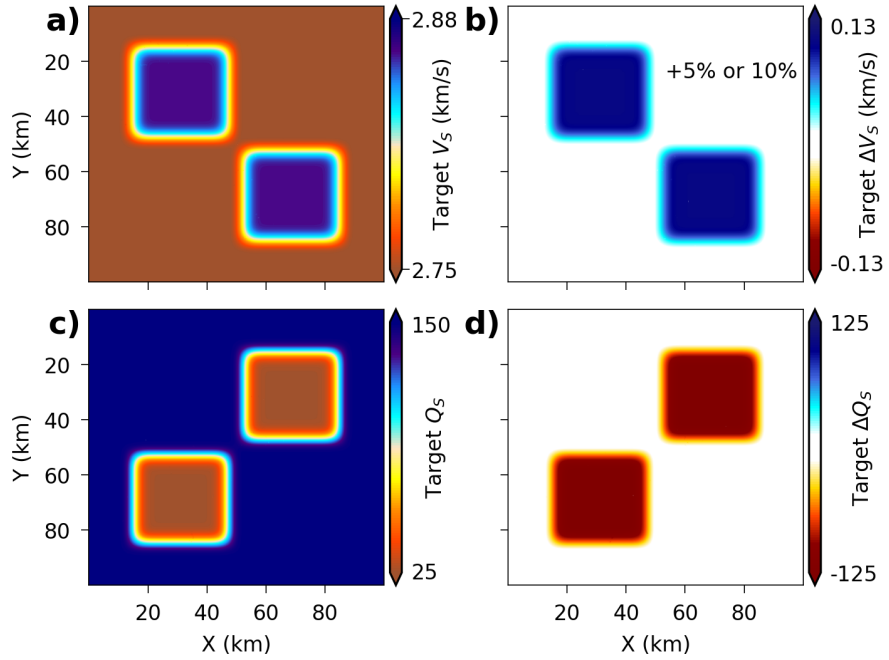

Figure 4. (a) The target V_S model structure; (b) The target model perturbation ΔV_S; (c) The target Q_S model structure; (c) The target model perturbation ΔQ_S.

3.2 Isolated velocity-Q anomaly example

A synthetic example with isolated velocity-Q anomalies is designed to examine the performances of WD and CD misfit functions for Q inversion. The initial velocity and Q models are the same with those in previous example. The target V_S model is created by embedding two positive perturbations in the homogeneous background, as shown in Figure 4a. Figure 4b shows the target model perturbation ΔV_S. The initial and target V_P models are created with V_P/V_S=2. The target Q_S model is created by embedding two strong anomalies (Q=25) in the homogeneous background, as shown in Figure 4c. Figure 4d shows the target model perturbation ΔQ_S. The initial and target Q_P models are created with Q_P=Q_S. The Q and velocity anomalies are isolated, which helps evaluate and visualize the interparameter mappings. Furthermore, to examine the influences of velocity errors, the velocity perturbations vary from +5% to +10% of the background velocities.

Figure 5 shows the Q_P and Q_S sensitivity kernels calculated using WD and CD misfit functions for comparison. With +5% velocity perturbations, the Q sensitivity kernels calculated using WD misfit function are blurred by the trade-off artifacts, as indicated by the arrows in Figures 5a and 5b. When increasing velocity perturbations to +10%, these artifacts in the Q sensitivity kernels become stronger, as shown in 5b and 5d. However, when using the new CD misfit function, for +5% and +10% velocity perturbations, the Q anomalies in the sensitivity kernels can be identified clearly suffering from fewer trade-off artifacts. Then, we carry out inversion experiments for Q estimation in the presence of +5% velocity perturbations. The P and S waves in both z and x component data are extracted to invert for Q_P and Q_S models, respectively. The recovered perturbations ΔQ_P^{wd} and ΔQ_S^{wd} by WD misfit function are contaminated by trade-off artifacts seriously, as indicated by the arrows in Figures 5i and 5j. However, with CD misfit function, the velocity errors produce very limited influences on the recovered perturbations ΔQ_P^{cd} and

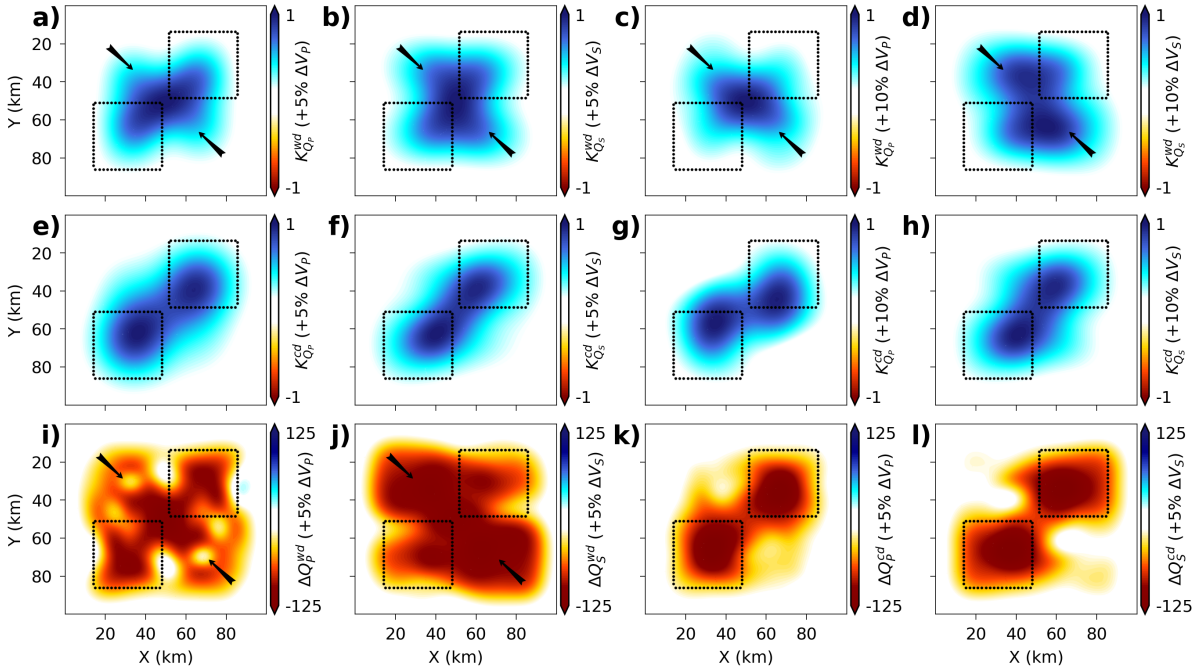

Figure 5. (a) and (b) are the Q_P and Q_S sensitivity kernels $K^{\mathrm{wd}}_{Q_P}$ and $K^{\mathrm{wd}}_{Q_S}$ calculated using WD misfit function with +5% velocity perturbations; (c) and (d) are the Q_P and Q_S sensitivity kernels $K^{\mathrm{wd}}_{Q_P}$ and $K^{\mathrm{wd}}_{Q_S}$ calculated using WD misfit function with +10% velocity perturbations; (e) and (f) are the sensitivity kernels $K^{\mathrm{cd}}_{Q_P}$ and $K^{\mathrm{cd}}_{Q_S}$ calculated using CD misfit function with +5% velocity perturbations; (g) and (h) are the sensitivity kernels $K^{\mathrm{cd}}_{Q_P}$ and $K^{\mathrm{cd}}_{Q_S}$ calculated using CD misfit function with +10% velocity perturbations; (i) and (j) are the inverted ΔQ^{wd}_P and ΔQ^{wd}_S model perturbations using WD misfit function with +5% velocity perturbations; (k) and (l) are the corresponding inverted ΔQ^{cd}_P and ΔQ^{cd}_S model perturbations using CD misfit function. The dotted-black lines indicate boundaries of the Q anomalies.

ΔQ^{cd}_S. The Q anomalies are clearly resolved, as shown in Figures 5k and 5l. These observations verify the advantages of CD misfit function for reliable Q inversion.

3.3 Global tomography example

In this subsection, following Yuan et al. (2016) and Tromp and Bachmann (2019), we design a synthetic global tomography example with realistic structures based on the Rayleigh wave phase speed map provided by Trampert and Woodhouse (1995). Figure 6a shows the target V_S model, which is discretized with 50 and 100 elements in Latitude and Longitude directions corresponding to one meter per degree (see section 5.2 in Yuan et al. (2016)). The initial V_S model is homogeneous with a constant value of 4.5 km/s. Figure 6b shows the target model perturbation ΔV_S. The target and initial V_P models are created with V_P/V_S=2. The initial Q_S model is homogeneous with a constant value of 150. The target Q_S model is created by embedding one strong anomaly (Q_S=25) in the homogeneous background, as shown in Figure 6c. Figure 6d shows the target model perturbation ΔQ_S. The target and initial Q_P models are created with Q_P=Q_S. A number of 32 sources and 293 receivers are selected as the acquisition network, as indicated by the red stars and white circles in Figure 6. Ray coverage of the network is plotted in

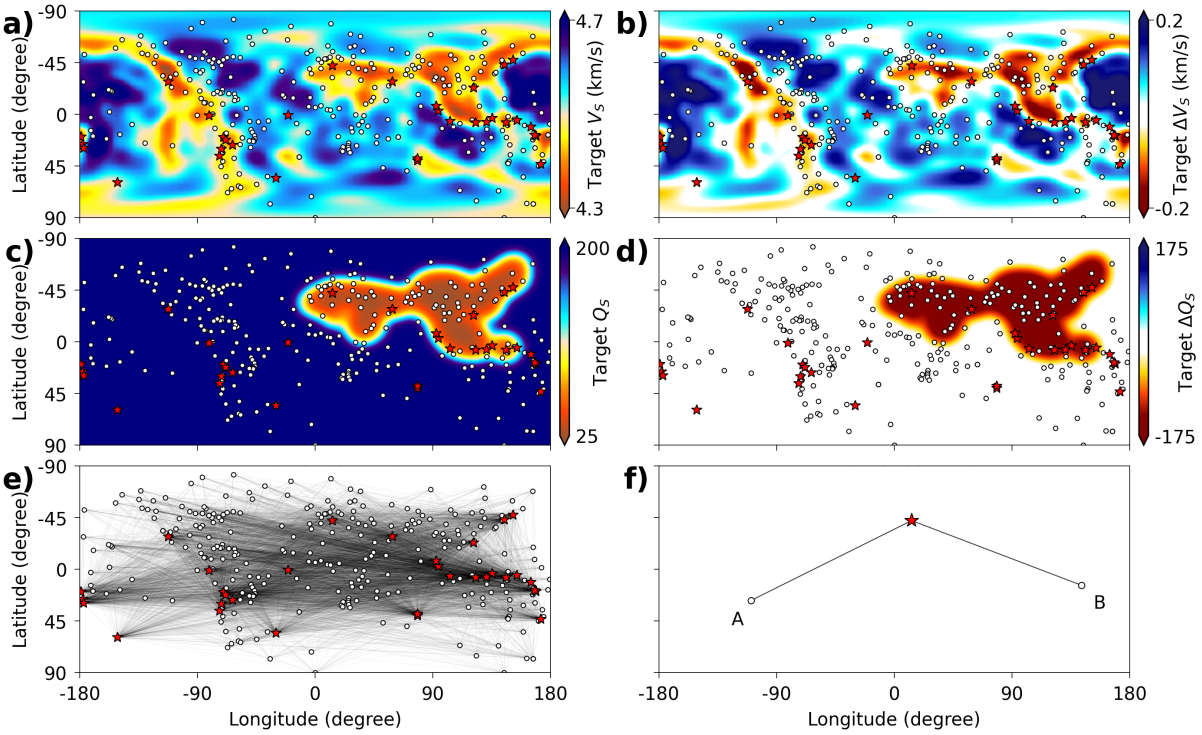

Figure 6. (a) and (b) are the target V_S velocity model and perturbation ΔV_S; The red stars and white circles indicate the locations of sources and stations. (c) and (d) are the target Q_S model and perturbation ΔQ_S; (e) shows the ray coverage of the seismic acquisition; (f) shows the locations of one source and two receivers (A and B) for waveform comparison.

Figure 6e. A 0.60 kHz dominant frequency Ricker wavelet is used as the source to generate the observed data.

For the inversion experiments, in stage I, we first invert for V_P and V_S models using cross-correlation (CC) traveltime misfit function (Luo & Schuster, 1991; Yuan et al., 2016) by expanding the frequency band from [0 kHz, 0.3 kHz] to [0 Hz, 0.6 kHz], which helps reduce non-linearity of the inverse problem. P and S waves are extracted from z and x component data to invert for V_P and V_S models separately. Then, the velocity structures are refined using WD misfit function. The final reconstructed velocity perturbations are presented in Figure 7. Reductions of the CC and WD data misfits are provided in Figure S2. The velocity structures around the Q anomalies appear to be under-estimated because of wrong Q models, as indicated by the black arrows. Because the ray coverage is limited, the whole velocity structures are not perfectly recovered.

In stage II of the inversion experiments, we use WD and CD misfit functions to invert for Q models. The Q_P and Q_S sensitivity kernels are calculated for comparison, as shown in Figure 8. In the Q sensitivity kernels calculated using WD misfit function, the Q anomalies can be identified. However, the structures are distorted to some extent. The trade-off artifacts caused by residual velocity errors also appear, as indicated by the red arrows in Figures 8a and 8b. Figures 8e and 8f show the corresponding inverted Q model perturbations by WD misfit function, wherein the Q structures are distorted and strong trade-off artifacts are clearly visible, as indicated by the red arrows. These false Q structures may result in misleading interpretations of subsurface models. The Q sensitivity kernels and model perturbations obtained using CD misfit function are presented in Fig-

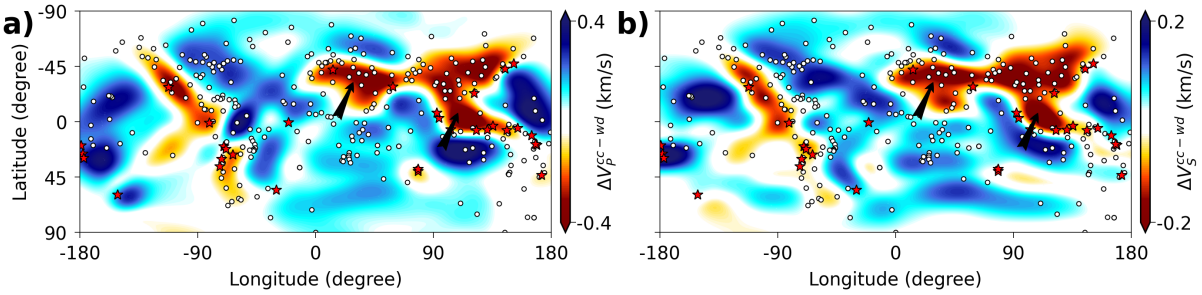

Figure 7. (a) and (b) are the inverted velocity model perturbations $\Delta V_P^{\text{cc-wd}}$ and $\Delta V_S^{\text{cc-wd}}$ using CC and WD misfit functions in stage I.

ures 8c-8d and Figures 8g-8h. Even though boundaries of the Q anomalies suffer from some smearing effects, the whole structures match with the target Q models more closely without containing strong artifacts.

In Figure 9, the recorded z component P and S waves calculated using the target, initial and inverted models are plotted. The locations of one source and two receivers (A and B) are presented in Figure 6f. Because the seismic waves propagating from the source to receiver B pass trough the strongly attenuative zones, the amplitudes and central-frequencies reduce more obviously that those of the waves recorded at A. Compared to the observed data, the synthetic data calculated using initial models experience amplitude and central-frequency reductions. Furthermore, compared to WD misfit function, the waveforms and central-frequencies of the data calculated using the inverted Q models by CD misfit function match with those of the observed data more closely. Comparisons of the x component waveforms are provided in Figure S3. The CD data misfits also reduce much more faster than the WD data misfits for Q inversion, as shown in Figure S2. From these numerical experiments and observations, we have proved that the new CD misfit function shows superiority to overcome the parameter trade-off difficulty and recover subsurface Q models more reliably.

3.4 3D viscoelastic inversion example

Finally, the proposed algorithms are applied to 3D viscoelastic full-waveform tomography with data emulating a distributed acoustic sensing (DAS) experiment acquired in a walk-away vertical seismic profile configuration (DAS-VSP). This synthetic example is designed to emulate a carbon capture and storage monitoring with DAS-VSP data in practical conditions (Hall et al., 2019).

Figures 10a and 10b show the target and initial 3D V_S models with complex topographic variations. The maximum sizes of the model in X, Y and Z directions are 1000 m, 1000 m and 500 m, respectively. The target and initial 3D V_P models are created with V_P/V_S=3. Figure 10c shows the target 3D Q_S model. The initial 3D Q_S model is homogeneous with Q_S=200. The target 3D Q_S model is created by embedding one attenuating layer (Q_S=20) overlapping with the low velocity formation at near-surface. The target and initial 3D Q_P models are created with Q_P=Q_S. The density model is homogeneous with a constant value of 3000 kg/m^3 and is kept unchanged in the inversion. Figure 11a shows the 3D mesh grids for discretizing the model and describing the complex topography. Figure 11b shows the acquisition settings. Twenty-five sources are arranged regularly on top surface of the model, as indicated by the red spheres in Figure 11b. The fiber-optic is deployed in the borehole with a recording interval of 1 m from 20 m to 500

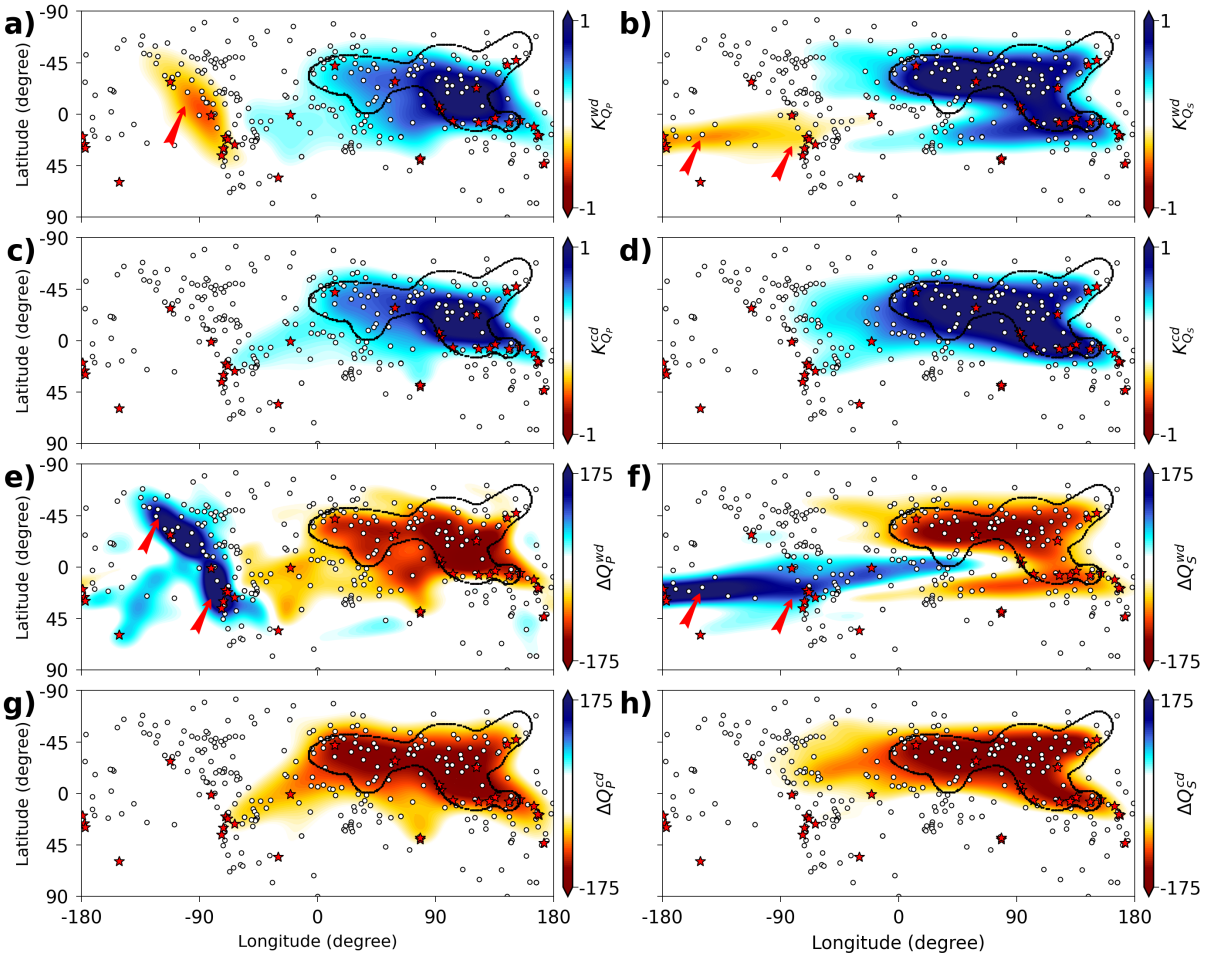

Figure 8. (a) and (b) are the Q_P and Q_S sensitivity kernels $K^{\mathrm{wd}}_{Q_P}$ and $K^{\mathrm{wd}}_{Q_S}$ calculated using WD misfit function; (c) and (d) are the Q_P and Q_S sensitivity kernels $K^{\mathrm{cd}}_{Q_P}$ and $K^{\mathrm{cd}}_{Q_S}$ calculated using CD misfit function; (e) and (f) are the inverted Q model perturbations ΔQ^{wd}_P and ΔQ^{wd}_S using WD misfit function; (g) and (h) are the inverted Q model perturbations ΔQ^{cd}_P and ΔQ^{cd}_S using CD misfit function. The dotted-black lines depict the boundaries of Q anomalies. The red stars and white circles indicate the locations of sources and stations.

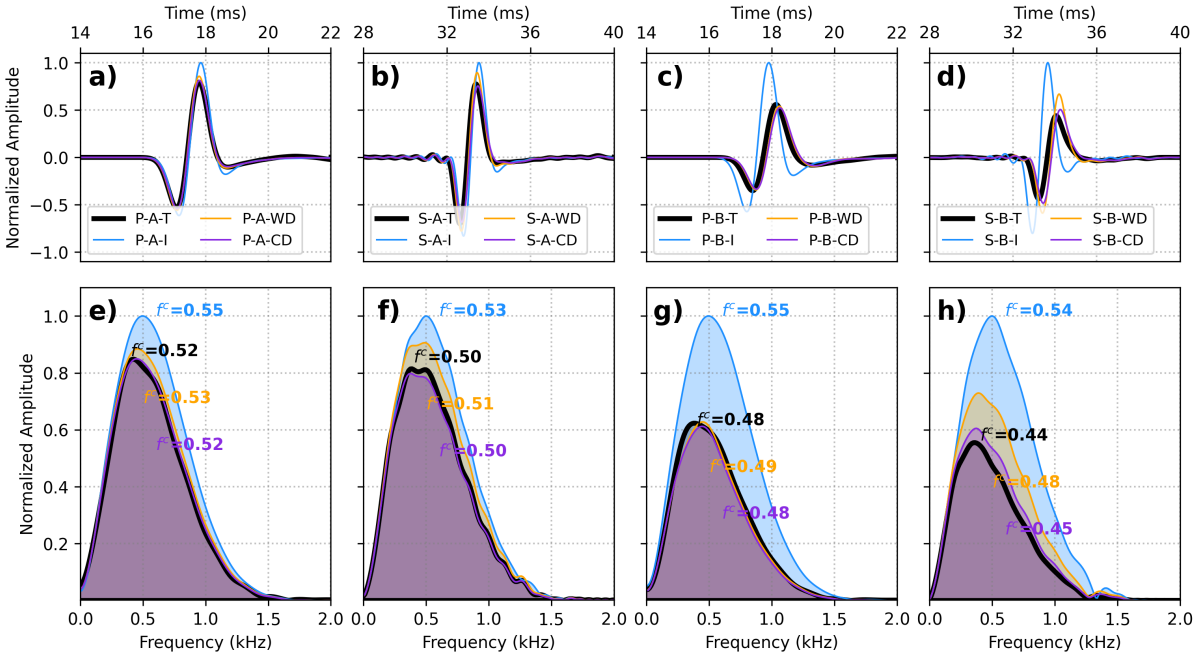

Figure 9. (a) shows the P waves recorded at A calculated using target (black PA-T), initial models (blue PA-I) and the inverted models by WD (orange PA-WD) and CD (purple PA-CD) misfit functions; (b) shows the S waves recorded at A calculated using target (black S-A-T), initial models (blue S-A-I) and the inverted models by WD (orange S-A-WD) and CD (purple S-A-CD) misfit functions; (c) shows the P waves recorded at B calculated using target (black PB-T), initial models (blue PB-I) and the inverted models by WD (orange PB-WD) and CD (purple PB-CD) misfit functions; (d) shows the S waves recorded at B calculated using target (black S-B-T), initial models (blue S-B-I) and the inverted models by WD (orange S-B-WD) and CD (purple S-B-CD) misfit functions; (e), (f), (g) and (h) are the corresponding amplitude spectra of the P and S waves in (a), (b), (c) and (d). f^c (kHz) indicate central-frequencies of P and S waves. Colors of the amplitude spectra and f^c are consistent with those of the waves.

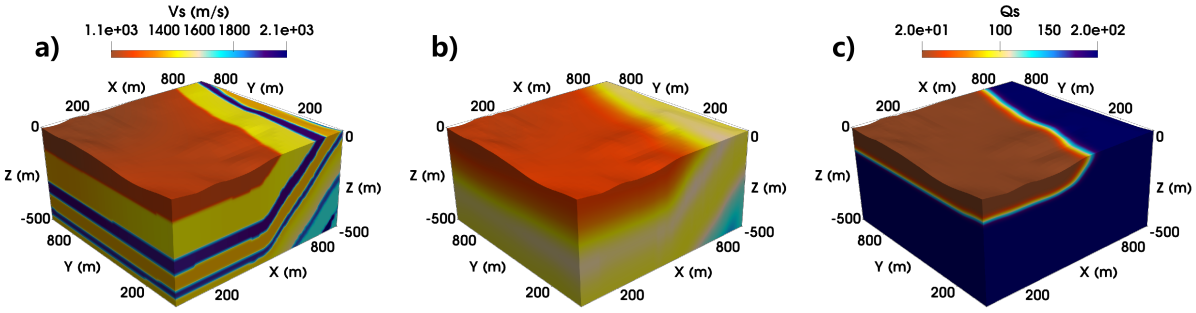

Figure 10. (a) The target 3D V_S velocity model with complex topographic variations; (b) The initial 3D V_S velocity model; (c) The target 3D Q_S model.

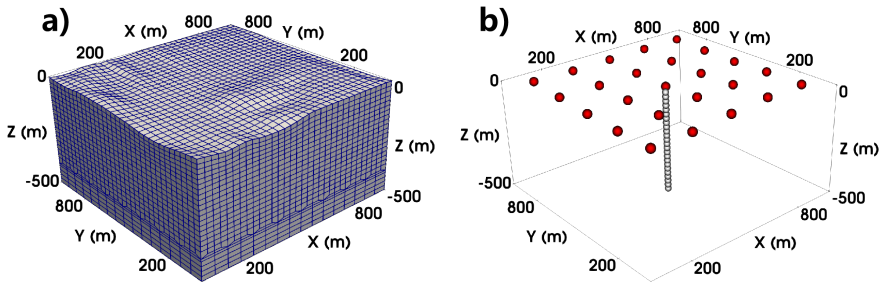

Figure 11. (a) The 3D unstructured mesh; (b) The distributions of sources (red spheres) and optical-fiber (white spheres).

m, as indicated by the white spheres in Figure 11b. Figure 12 shows the 3D slices extracted from the target model perturbations ΔV_S and ΔQ_S.

The source used for forward modeling is a 30 Hz dominant frequency Ricker wavelet. We first transform the data of vertical strain-rate into z component displacement data using the frequency-wavenumber method (Daley et al., 2016), which is then used as the observed data for inversion. The forward and adjoint simulations in 3D viscoelastic medium are performed on a local workstation with 16 cores (3.10 GHz) and 2 NVIDIA GeoRTX2080Ti Graphic Processing Unit (GPU) cards.

In the inversion experiments, we first invert for V_P and V_S models using CC and WD misfit functions within frequency band [1 Hz, 30 Hz] in stage I. The 3D slices extracted from the reconstructed velocity model perturbations ΔV_P and ΔV_S are presented in Figure 13. Reductions of the normalized CC and WD data misfits are presented in Figure S4. Because the source-receiver illumination is limited, only the velocity structures around the borehole are well recovered. Some formations are distorted due to the influences of wrong Q models. In stage II, we apply the CD misfit function to invert for the near-surface Q_P and Q_P anomalies. The reconstructed model perturbations ΔQ_P and ΔQ_S are presented in Figure 14. Reductions of the normalized CD data misfits are presented in Figure S4. Even though some artifacts appear in the deeper parts of the inverted ΔQ_P and ΔQ_S model perturbations, the strong Q_P and Q_S anomalies at near-surface are resolved clearly. Figure 15 shows the comparisons of shot gathers and traces (at the depths of -250 m and -300 m) calculated using target, initial and inverted models. Waveforms and central-frequencies calculated using the inverted velocity and Q mod-

els match those of the observed data closely. These results indicate that the proposed algorithms are applicable to DAS-VSP data for Q_P and Q_S inversion.

4 Discussions

In this paper, the engine for viscoelastic forward and adjoint simulations is based on the GSLS rheology in time domain. However, the attenuation effects on propagating seismic waves can be modeled using many other physical mechanisms in time or frequency domain. Thus, the velocity and Q sensitivity kernels in equation (13) are only applicable to viscoelastic inversion experiments based on the GSLS model. The formulas of calculating the sensitivity kernels should depend on the physical mechanism in forward simulation. Furthermore, it is necessary to evaluate the uncertainties and modeling errors in viscoelastic adjoint tomography with different physical mechanisms.

Theoretically, central-frequency of the seismic data is mainly controlled by Q variations. In the numerical experiments, we find that velocity perturbations result in fluctuations of the amplitude spectra. In practical seismic data, the amplitude spectra are always dithered due to noise. These effects can result in errors for calculating central-frequencies using equation (19), which will lead to unexpected uncertainties for Q inversion. Smoothing the amplitude spectra can remove these fluctuations and dithering effects. Variation of the central-frequency can also be measured by cross-correlating the amplitude spectra of synthetic and observed data, which is similar to the CC traveltime misfit function but leads to different adjoint source for calculating Q sensitivity kernels. Stability and accuracy of these two different approaches for determining central-frequency difference will be examined in our future studies.

5 Conclusions

In this study, we develop theory and methods for full-waveform adjoint Q tomography in 2D/3D viscoelastic medium. The Q sensitivity kernels are constructed efficiently with adjoint memory strain variables based on the adjoint-state method. Multiparameter Hessian-vector products provide an effective tool for quantifying the trade-offs between velocity and Q. Compared to traditional waveform-difference misfit function, the proposed new central-frequency misfit function is more sensitive to Q variations and can invert for the Q models more reliably by reducing the trade-off artifacts caused by velocity errors. Synthetic inversion examples in 2D/3D viscoelastic medium verify that the proposed algorithms are effective and advantageous for adjoint Q tomography, which is applicable in both exploration and earthquake seismology.

Acknowledgments

This research is supported by National Natural Science Foundation of China (Grant No. E1115401), IGGCAS Research Start-up Funds (Grant No. E0515402), IGGCAS grant 2019031 and CAS innovation program ZDBS-LY-DQC003. K. A. Innanen is supported by CREWES project and NSERC grants (CRDPJ 461179-13 and CRDPJ 543578-19). The forward modeling codes in 2D and 3D viscoelastic medium are available at `https://github.com/PanIGGCAS/SeisElastic2D_1.1/specfem2d` and `https://github.com/geodynamics/specfem3d`. The full-waveform adjoint tomography codes can be found at `https://github.com/PanIGGCAS/SeisElastic2D_1.1/SeisElastic2D`. The data, models and plotting scripts used in this research can be downloaded from `http://doi.org/10.6084/m9.figshare.13655186`.

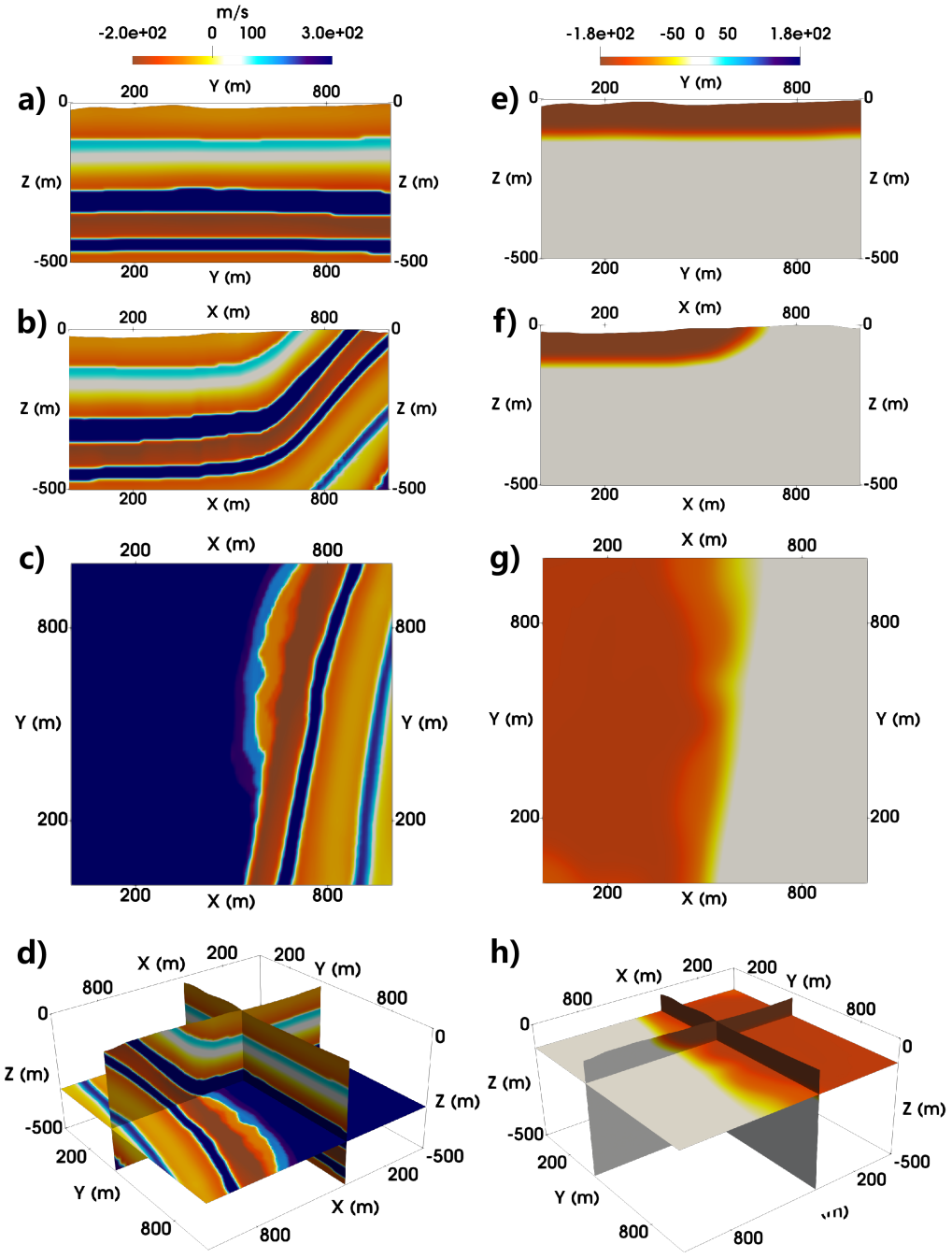

Figure 12. (a) shows the Y-Z slice of the target model perturbation ΔV_S at X=400 m; (b) shows the X-Z slice of the target model perturbation ΔV_S at Y=400 m; (c) shows the X-Y slice of the target model perturbation ΔV_S at Z=-320 m; (d) shows the 3D view of these slices in ΔV_S; (e), (f) and (g) are the Y-Z, X-Z and X-Y slices of target model perturbation ΔQ_S at X=400 m, Y=400 m and Z=-100 m, respectively; (h) is the 3D view of these slices in ΔQ_S.

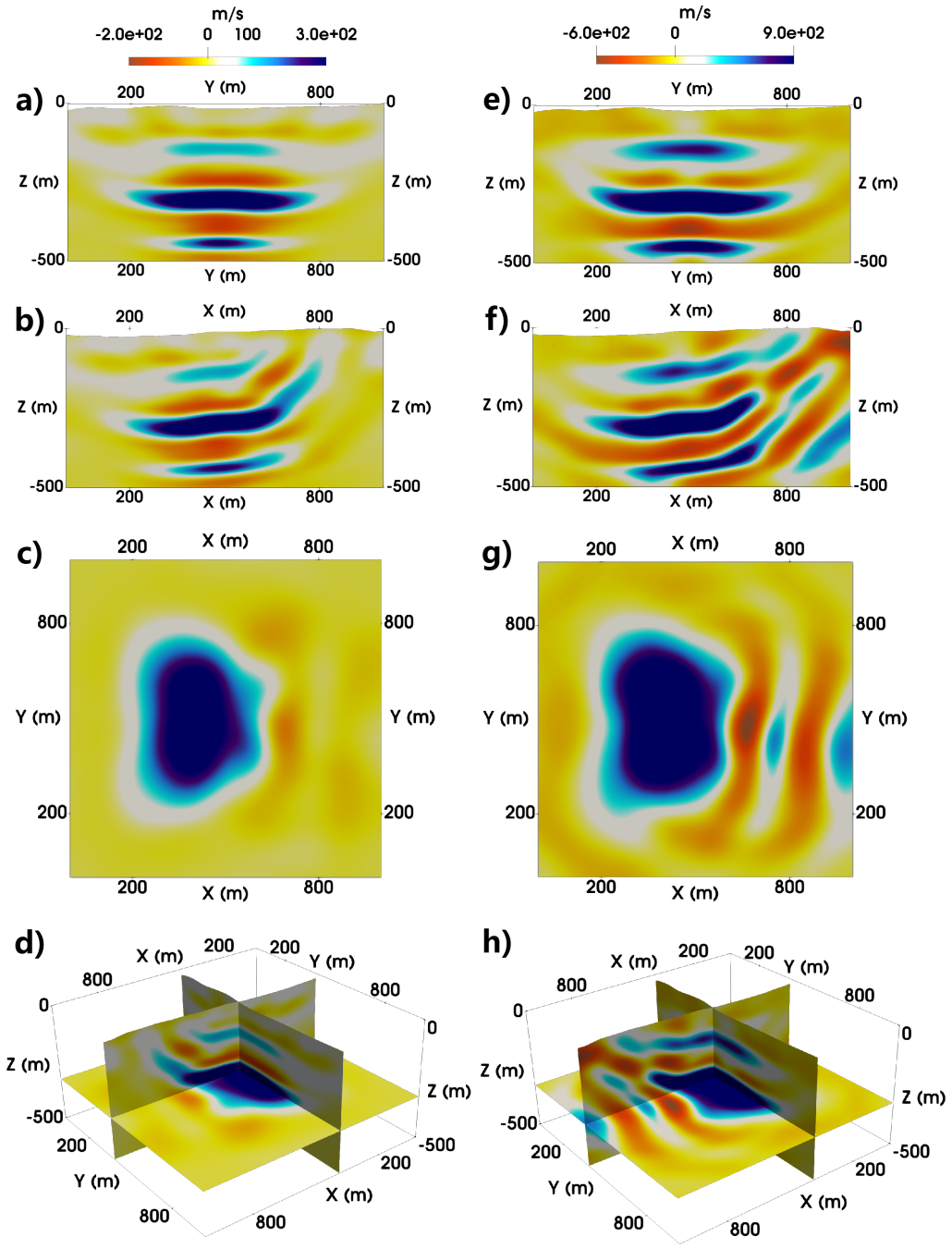

Figure 13. (a) shows the Y-Z slice of the inverted model perturbation ΔV_P at X=400 m; (b) shows the X-Z slice of the inverted model perturbation ΔV_P at Y=400 m; (c) shows the X-Y slice of the inverted model perturbation ΔV_P at Z=-320 m; (d) shows the 3D view of these slices in ΔV_P; (e), (f) and (g) are the Y-Z, X-Z and X-Y slices of inverted model perturbation ΔV_S at X=400 m, Y=400 m and Z=-320 m, respectively; (h) is the 3D view of these slices in ΔV_S.

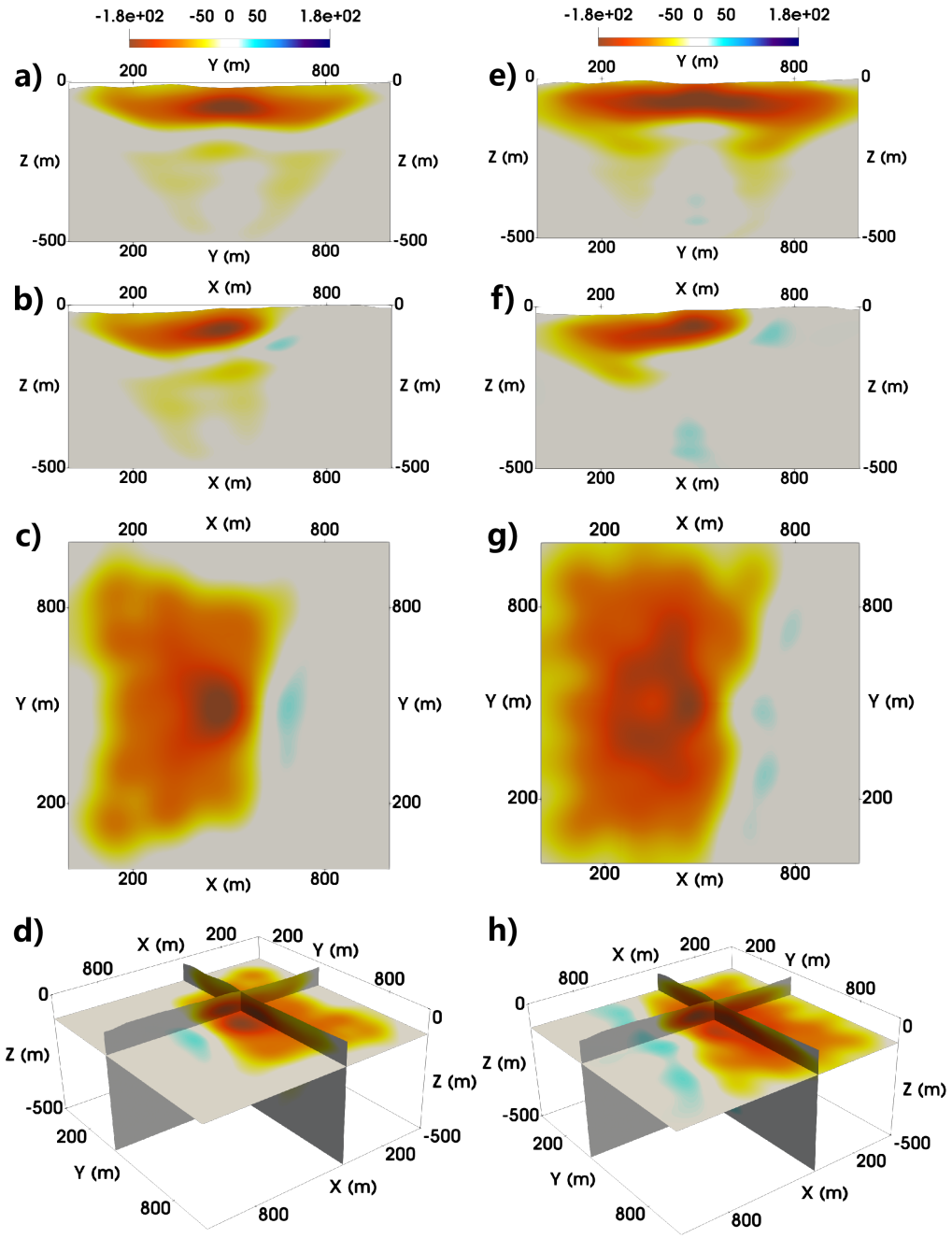

Figure 14. (a) shows the Y-Z slice of the inverted model perturbation ΔQ_P at X=400 m; (b) shows the X-Z slice of the inverted model perturbation ΔQ_P at Y=400 m; (c) shows the X-Y slice of the inverted model perturbation ΔQ_P at Z=-100 m; (d) shows the 3D view of these slices in ΔQ_P; (e), (f) and (g) are the Y-Z, X-Z and X-Y slices of inverted model perturbation ΔQ_S at X=400 m, Y=400 m and Z=-100 m, respectively; (h) is the 3D view of these slices in ΔQ_S.

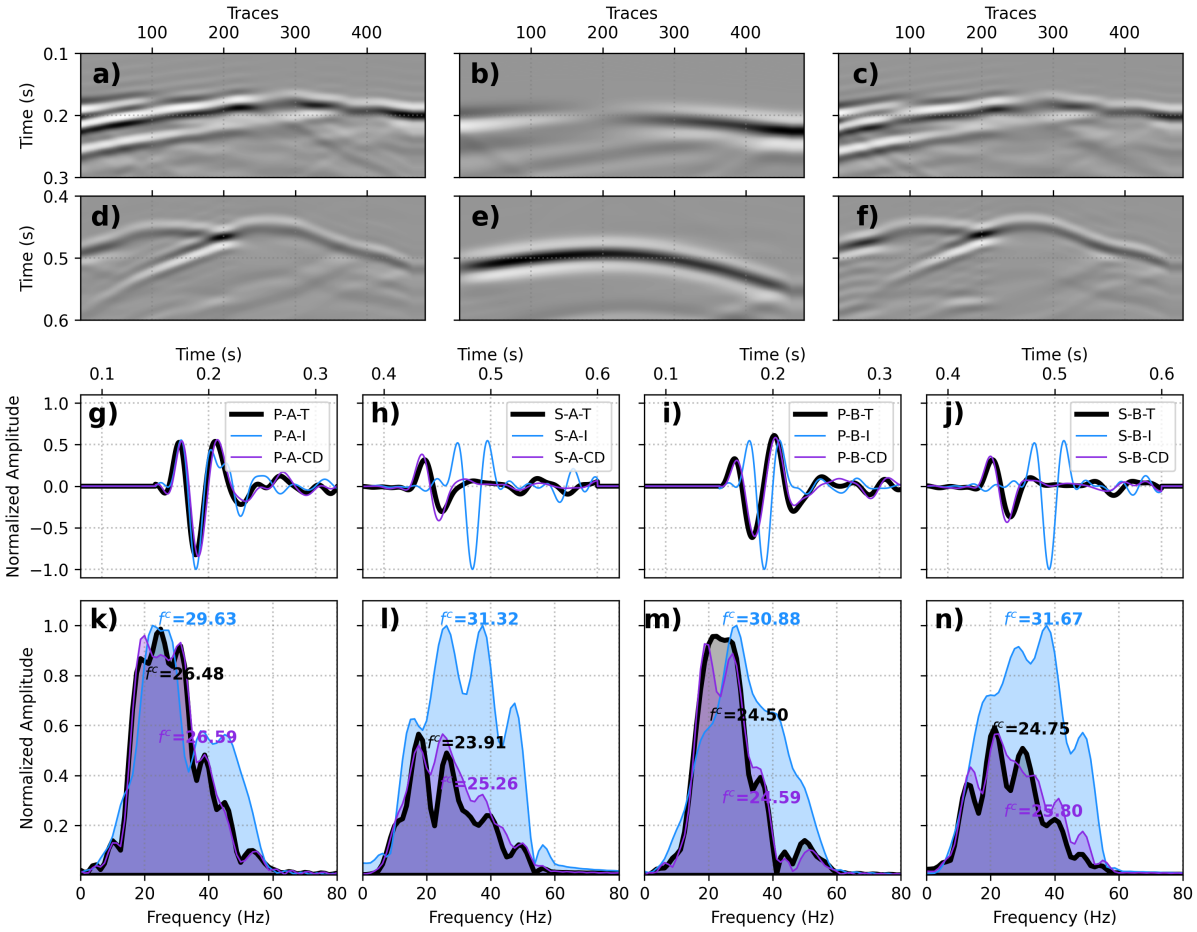

Figure 15. (a), (b) and (c) show the shot gathers of P waves calculated using target, initial and inverted models with source at (X=100 m, Y=200 m, Z=0 m); (d), (e) and (f) show the corresponding shot gathers of S waves; (g) shows the P waves recorded at A (Z=-250 m) calculated using target (black PA-T), initial (blue PA-I) and inverted models (purple PA-CD); (h) shows the S waves recorded at A calculated using target (black S-A-T), initial (blue S-A-I) and inverted models (purple S-A-CD); (i) shows the P waves recorded at B (Z=-300 m) calculated using target (black PB-T), initial (blue PB-I) and inverted models (purple PB-CD); (j) shows the S waves recorded at B calculated using target (black S-B-T), initial (blue S-B-I) and inverted (purple S-B-CD); (k), (l), (m) and (m) are the corresponding amplitude spectra of the P and S waves in (g), (h), (i) and (j). f_c (Hz) indicate central-frequencies. Colors of the amplitude spectra and f^c are consistent with those of the waves.

Appendix A Review of forward modeling in viscoelastic medium based on GSLS model

Different physical mechanisms and approaches (Kolsky, 1952; Biot, 1956; Futterman, 1962; Emmerich & Korn, 1987; Carcione et al., 1988a, 1988b; Robertsson et al., 1994; T. Zhu & Carcione, 2014; N. Wang et al., 2019) have been suggested to model the anelastic behaviors of propagating seismic waves in real Earth's medium. The system constructed by the parallel connection of several standard linear elements is referred to as the generalized standard linear solid rheology (H. Liu et al., 1976; Aki & Richards, 2002). In this study, the GSLS model is adopted as the forward modeling engine for adjoint Q tomography in time domain. The convolution integral in the constitutive relationship (equation (2)) can be eliminated by taking partial derivative of time on its both sides, which yields (Blanch et al., 1995)

$$\begin{aligned}\partial_t T_{ij} = \kappa^R \delta_{ij}\delta_{kl}\left[\left(\tau_\kappa + 1\right)\partial_t \varepsilon_{kl} - \sum_{p=1}^{P}\tau_\kappa \epsilon_{kl}^p\right] \\ + \mu^R\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk} - \frac{2}{3}\delta_{ij}\delta_{kl}\right)\left[\left(\tau_\mu + 1\right)\partial_t \varepsilon_{kl} - \sum_{p=1}^{P}\tau_\mu \epsilon_{kl}^p\right],\end{aligned} \tag{A1}$$

where ϵ_{kl}^p are the introduced memory strain variables describing anelastic characteristics of the seismic waves (Pan & Wang, 2020):

$$\epsilon_{kl}^p = \frac{1}{P\tau^{\sigma p}} e^{-t/\tau^{\sigma p}} H\left(t\right) * \partial_t \varepsilon_{kl}, \tag{A2}$$

and time derivative of the memory variables satisfies the following equation (Pan & Wang, 2020):

$$\partial_t \epsilon_{kl}^p = \frac{1}{P\tau^{\sigma p}}\partial_t \varepsilon_{kl} - \frac{1}{\tau^{\sigma p}}\epsilon_{kl}^p. \tag{A3}$$

Thus, wave propagation in viscoelastic medium can be modeled by numerically solving the equation of motion and equations (A1) and (A3) with the superposition of several parallel relaxation mechanisms. The quality factor Q is defined in frequency domain as (Blanch et al., 1995; Bohlen, 2002)

$$\tilde{Q}_M^{-1}\left(\omega, \tau_M, \tau^{\sigma p}\right) = \frac{\mathcal{I}\left[M\left(\omega\right)\right]}{\mathcal{R}\left[M\left(\omega\right)\right]} = \left[\tau_M \sum_{p=1}^{P}\frac{\omega\tau^{\sigma p}}{1+\left(\omega\tau^{\sigma p}\right)^2}\right]\left[1 + \tau_M \sum_{p=1}^{P}\frac{\left(\omega\tau^{\sigma p}\right)^2}{1+\left(\omega\tau^{\sigma p}\right)^2}\right]^{-1}, \tag{A4}$$

where M indicates Fourier transform of the time derivative of the modulus relaxation function, ω denotes angular frequency, $\mathcal{I}$ means imaginary part.

To approximate a nearly constant value of Q_{ref} within the seismic frequency band by several Maxwell bodies, the relaxation parameters τ_M and $\tau^{\sigma p}$ can be obtained by minimizing the distance between Q_{ref}^{-1} and $\tilde{Q}_M^{-1}$ (equation (A4)). Emmerich and Korn (1987) solved this problem using the classical linear optimization approach. Fichtner and van Driel (2014) obtained the relaxation parameters with Monte Carlo approaches and found that the optimal relaxation parameters can be determined with the enforcement of $\tau_M = Q_M^{-1}$ for the whole set of Q values. In this study, following Blanc et al. (2016), we introduce the objective function:

$$\mathcal{J}\left(\{\tau_M, \tau^{\sigma p}\}; P\right) = \int_{\omega_0}^{\omega'}\left\{\sum_{p=1}^{P}\frac{\tau_M}{P}\frac{\omega\tilde{Q}_{\text{ref}}\left(\omega\right)\left[\left(\tau^{\sigma p}\right)^{-1} - \omega\tilde{Q}_{\text{ref}}^{-1}\left(\omega\right)\right]}{\left(\tau^{\sigma p}\right)^{-2} + \omega^2} - 1\right\}^2 d\omega, \tag{A5}$$

where ω_0 and ω' indicate the minimum and maximum angular frequencies. The relaxation parameters are solved using the *SolvOpt* algorithm (Kappel & Kuntsevich, 2000) based on the iterative Shor's method (Shor, 1985) with positivity constraints of $\tau_M >$

0 and $\tau^{\sigma p} > 0$ (see section 2.3 in Blanc et al. (2016)). Similarly, with the enforcement of $\tau_M = Q_M^{-1}$, Q can be explicitly introduced in the constitutive relation and equation of motion, which enables and facilitates the derivation of Q sensitivity kernels in adjoint tomography. In this study, a number of 3 relaxation mechanisms is used to approximate constant Q in the forward modeling experiments.

Appendix B Hessian-vector product calculation in viscoelastic medium with the adjoint-state method

Products of multiparameter Hessian with an arbitrary vector can be calculated efficiently with the first-order adjoint-state method (Métivier et al., 2013; Pan et al., 2018). We can first consider minimizing the following augmented Lagrangian misfit function:

$$\tilde{\boldsymbol{\chi}}\left(\mathbf{m},\mathbf{u},\boldsymbol{\Lambda}\right)=\sum_{\mathbf{x}_r}\int_0^{t'}\int_\Omega\left[\mathbf{u}^\ddagger v-\boldsymbol{\Lambda}\cdot\left(\rho\partial_t^2\mathbf{u}-\nabla\cdot\mathbf{T}-\boldsymbol{f}^s\right)\right]d\mathbf{x}dt, \tag{B1}$$

where "‡" means complex conjugate, $\boldsymbol{\Lambda}$ is a new Lagrangian multiplier and v is an arbitrary function. Variation of misfit function with respect to the perturbations of model parameters ($\Delta\kappa$, $\Delta\mu$, ΔQ_κ and ΔQ_μ) and wavefield (Δu_n) can be obtained as

$$\begin{aligned}
\Delta\tilde{\boldsymbol{\chi}}\left(\mathbf{m},\mathbf{u},\boldsymbol{\Lambda}\right)=&\int_0^{t'}\int_\Omega\sum_{\mathbf{x}_r}\left[v-\rho\partial_t^2\Lambda_i+\partial_j\left(\kappa\delta_{ij}\delta_{kl}*\partial_t\tilde{\Gamma}_{kl}\right)\right]\Delta u_n d\mathbf{x}dt\\
&+\int_0^{t'}\int_\Omega\partial_j\left[\mu\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}-\frac{2}{3}\delta_{ij}\delta_{kl}\right)*\partial_t\tilde{\Gamma}_{kl}\right]\Delta u_n d\mathbf{x}dt\\
&-\int_0^{t'}\int_\Omega\left[\left(\Delta\kappa*\partial_t\tilde{\Gamma}_{kk}\right)\Gamma_{ii}-\left(\Delta Q_\kappa\frac{\kappa^R}{Q_\kappa^2}\sum_{p=1}^{P}\tau^{\sigma p}\tilde{\Pi}_{kk}^p\right)\varepsilon_{ii}\right]d\mathbf{x}dt\\
&-\int_0^{t'}\int_\Omega\left[\Delta\mu\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}-\frac{2}{3}\delta_{ij}\delta_{kl}\right)*\partial_t\tilde{\Gamma}_{kl}\right]\varepsilon_{ij}d\mathbf{x}dt\\
&+\int_0^{t'}\int_\Omega\left[\Delta Q_\mu\frac{\mu^R}{Q_\mu^2}\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}-\frac{2}{3}\delta_{ij}\delta_{kl}\right)\sum_{p=1}^{P}\tau^{\sigma p}\tilde{\Pi}_{kl}^p\right]\varepsilon_{ij}d\mathbf{x}dt,
\end{aligned} \tag{B2}$$

where $\tilde{\Gamma}_{ij}=\partial_j\Lambda_i$ is the Lagrangian strain field and $\tilde{\Pi}_{kl}^p$ are the corresponding Lagrangian memory strain variables. In equation (B2), we ignore density ρ for compactness. The adjoint-state equation can be obtained by setting the coefficient of Δu_n as zero:

$$\rho\partial_t^2\Lambda_i-\partial_j\left[\kappa\delta_{ij}\delta_{kl}*\partial_t\tilde{\Gamma}_{kl}+\mu\left(\delta_{ik}\delta_{jl}+\delta_{il}\delta_{jk}-\frac{2}{3}\delta_{ij}\delta_{kl}\right)*\partial_t\tilde{\Gamma}_{kl}\right]=v, \tag{B3}$$

where v serves as the adjoint source and solution of equation (B3) can be obtained by convolving Green's function G_{in} with v:

$$\Lambda_i\left(t\right)=\int_0^{t'}G_{in}\left(t-t''\right)v\left(t\right)dt''. \tag{B4}$$

To calculate the Hessian-vector product, we can first derive the gradients of misfit function (equation (B1)) with respect to the model parameters. For example, gradient for κ is:

$$\nabla_\kappa\tilde{\boldsymbol{\chi}}=-\sum_{\mathbf{x}_r}\int_\Omega\int_0^{t'}\int_0^{t'}\Gamma_{ii}\left(t\right)\partial_t\partial_k G_{kn}\left(t'-t-t''\right)v\left(t''\right)dt''dtd\mathbf{x}. \tag{B5}$$

Product of Gauss-Newton Hessian with model perturbation vector can be calculated by replacing v in equation (B5) with Jacobian-vector product. For example, product of the Jacobian matrix with model perturbation vector $\Delta\kappa$ is

$$\mathbf{J}_\kappa\left(t''\right)=-\int_0^{t'}\partial_t\partial_{k'}G_{k'n'}\left(t''-t'''\right)\Delta\kappa\Gamma_{i'i'}\left(t'''\right)dt'''. \tag{B6}$$

Then, replacing v in equation (B5) with equation (B6) gives the Hessian-vector product $\mathbf{H}_{\kappa\kappa}\Delta\kappa$:

$$\begin{aligned}\mathbf{H}_{\kappa\kappa}\Delta\kappa = \sum_{\mathbf{x}_r}\int_{\Omega}\int_0^{t'}\int_0^{t'}\Gamma_{ii}\left(t\right)\partial_t\partial_k G_{kn}\left(t'-t-t''\right)\\ \times\int_0^{t'}\partial_t\partial_{k'}G_{k'n'}\left(t''-t'''\right)\Delta\kappa\Gamma_{i'i'}\left(t'''\right)dt'''dt''dtd\mathbf{x}.\end{aligned} \tag{B7}$$

More detailed introduction for calculating the Hessian-vector product can be found in Appendix B of Pan et al. (2018). The multiparameter Hessian-vector products associated with other physical parameters or misfit functions can be derived following the same procedure.

References

Ajo-Franklin, J. B., Dou, S., Lindsey, N. J., Monga, I., Tracy, C., Robertson, M., ... Li, X. (2019). Distributed acoustic sensing using dark fiber for near-surface characterization and broadband seismic event detection. *Scientific Reports*, *9*, 1–14.

Aki, K., & Richards, P. G. (2002). *Quantitative seismology* (2nd ed.). University Science Books.

Alkhalifah, T., & Plessix, R. E. (2014). A recipe for practical full-waveform inversion in anisotropic media: An analytic parameter resolution study. *Geophysics*, *79*(3), R91–R101.

Bao, X., A, D. C., Jin, G., James, B. J., & Shen, Y. (2016). Imaging Rayleigh wave attenuation with USArray. *Geophysical Journal International*, *206*(1), 241–259.

Barnes, A. E. (1993). Instantaneous spectral bandwidth and dominant frequency with applications to seismic reflection data. *Geophysics*, *58*(3), 419–428.

Bath, M. (1974). *Spectral analysis in geophysics: Developments in solid earth geophysics*. Elsevier Science Publishing Co.

Berkhout, A. J. (1984). *Seismic resolution: Resolving power of acoustical echo techniques*. Geophysical Press.

Berryman, J. G. (1988). Seismic wave attenuation in fluid-saturated porous media. *Pure Appl. Geophys.*, *128*, 423–432.

Biot, M. A. (1956). Theory of propagation of elastic waves in a fluid-saturated porous solid, I, Low-frequency range. *J. Acoust. Soc. Am.*, *28*(0), 168–178.

Blanc, E., Komatitsch, D., Chaljub, E., Lombard, B., & Xie, Z. (2016). Highly accurate stability-preserving optimization of the Zener viscoelastic model, with application to wave propagation in the presence of strong attenuation. *Geophysical Journal International*, *205*(1), 427–439.

Blanch, J. O., Robertsson, J. O. A., & Symes, W. W. (1995). Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique. *Geophysics*, *60*(1), 176–184.

Bohlen, T. (2002). Parallel 3-D viscoelastic finite difference seismic modelling. *Computer & Geosciences*, *28*, 845–870.

Bostock, M. G., Rondenay, S., & Shragge, J. (2001). Multiparameter two-dimensional inversion of scattered teleseismic body waves 1. Theory for oblique incidence. *Journal of Geophysical Research*, *106*(12), 30771–30782.

Brossier, R. (2011). Two-dimensional frequency-domain visco-elastic full waveform inversion: Parallel algorithms, optimization and performance. *Computer & Geosciences*, *37*, 444–445.

Brossier, R., Operto, S., & Virieux, J. (2009). Robust frequency-domain full-waveform inversion using the $l1$ norm. *Geophysical Research Letters*, *36*, L20310.

Cao, A., & Romanowicz, B. (2004). Hemispherical transition of seismic attenuation at the top of the Earth's inner core. *Earth Planet. Sci. Lett.*, *228*, 243–253.

Carcione, J. M., Kosloff, D., & Kosloff, R. (1988a). Viscoacoustic wave propagation simulation in the earth. *Geophysics*, *53*, 769–777.

Carcione, J. M., Kosloff, D., & Kosloff, R. (1988b). Wave propagation simulation in a linear viscoelastic medium. *Geophysical Journal International*, *95*(3), 597–611.

Charara, M., Barnes, C., & Tarantola, A. (2000). *Full waveform inversion of seismic data for a viscoelastic medium, in Methods and Applications of Inversion.* Springer.

Chen, M., Niu, F., Tromp, J., Lenardic, A., Lee, C. A., Cao, W., & Ribeiro, J. (2017). Lithospheric foundering and underthrusting imaged beneath Tibet. *Nature Communications*, *8*(0), 15659.

Chen, P., Zhao, L., & Chapman, C. H. (2007). Full 3D tomography for the crustal structure of the Los Angeles region. *Bulletin of the Seismological Society of America*, *97*(4), 1094–1120.

Cox, B., Wills, P., Kiyashchenko, D., Jeff, M., Jorge, L., Bourne, S., ... Roy, J. (2012). Distributed Acoustic Sensing for Geophysical Measurement, Monitoring and Verification. *CSEG Recorder*, *37*, 7–13.

Dahlen, F. A., & Tromp, J. (1998). *Theoretical global seismology.* Princeton University Press.

Daley, T. M., Miller, D. E., Dodds, K., Cook, P., & Freifeld, B. M. (2016). Field testing of modular borehole monitoring with simultaneous distributed acoustic sensing and geophone vertical seismic profiles at Citronelle, Alabama. *Geophysical Prospecting*, *64*, 1318–1334.

Debayle, E., Bodin, T., & Durand, S. (2020). Seismic evidence for partial melt below tectonic plates. *Nature*, *586*(7830), 555–559.

Dong, X., & Yang, D. (2020). Reanimating lithosphere of North China: new insights from full-waveform seismic tomography. *Science Bulletin*, *65*(10), 775–777.

Eaid, M. V., Keating, S. D., & Innanen, K. A. (2020). Multiparameter seismic elastic full-waveform inversion with combined geophone and shaped fiber-optic cable data. *Geophysics*, *85*(6), R537–R552.

Eberhart-Phillips, D., & Chadwick, M. (2002). Three-dimensional attenuation model of the shallow Hikurangi subduction zone in the Raukumara Peninsula, New Zealand. *J. Geophys. Res.*, *107*(B2), 2033.

Emmerich, H., & Korn, M. (1987). Incorporation of attenuation into time-domain computations of seismic wave fields. *Geophysics*, *52*, 1252–1264.

Fichtner, A., & Leeuwen, T. (2015). Resolution analysis by random probing. *J. Geophys. Res. Solid Earth*, *120*, 5549–5573.

Fichtner, A., & Trampert, J. (2011a). Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, *185*(2), 775–798.

Fichtner, A., & Trampert, J. (2011b). Resolution analysis in full waveform inversion. *Geophysical Journal International*, *187*(3), 1604–1624.

Fichtner, A., & van Driel, M. (2014). Models and Fréchet kernels for frequency-(in)dependent Q. *Geophysical Journal International*, *198*, 1878–1889.

Futterman, W. I. (1962). Dispersive body waves. *Journal of Geophysical Research*, *67*, 5279–5291.

Hall, K. W., Bertram, K. L., Bertram, M. B., Innanen, K. A., & Lawton, D. C. (2019). Simultaneous accelerometer and optical fibre multi-azimuth walk-away VSP experiment, Newell County, Alberta, Canada. *SEG Technical Program Expanded Abstracts*, 5340–5344.

Hauge, P. (1981). Measurements of attenuation from vertical seismic profiles. *Geophysics*, *46*, 1548–1558.

Hauksson, E., & Shearer, P. S. (2006). Attenuation models (Q_P and Q_S) in three

dimensions of the southern California crust: Inferred fluid saturation at seismogenic depths. *J. Geophys. Res.*, *111*, B05302.

He, Q., & Wang, Y. (2020). Inexact Newton-type methods based on Lanczos orthonormal method and application for full waveform inversion. *Inverse Problems*, *36*, 115007.

Innanen, K. A. (2011). Inversion of the seismic AVF/AVA signatures of highly attenuative targets. *Geophysics*, *76*(1), R1–R14.

Innanen, K. A. (2014). Seismic AVO and the inverse Hessian in precritical reflection full waveform inversion. *Geophysical Journal International*, *199*, 717–734.

Kamei, R., & Pratt, G. R. (2013). Inversion strategies for visco-acoustic waveform inversion. *Geophysical Journal International*, *194*(2), 859–884.

Kappel, F., & Kuntsevich, A. (2000). An implementation of Shor's r-algorithm. *Comput. Optim. Appl.*, *15*(2), 193–205.

Karaoğlu, H., & Romanowicz, B. (2018). Global seismic attenuation imaging using full-waveform inversion: a comparative assessment of different choices of misfit functionals. *Geophysical Journal International*, *212*(2), 807–826.

Kazei, V., & Alkhalifah, T. (2019). Scattering radiation pattern atlas: What anisotropic elastic properties can body waves resolve? *J. Geophys. Res. Solid Earth*, *124*(3), 2781–2811.

Keating, S., & Innanen, K. A. (2020). Parameter crosstalk and leakage between spatially separated unknowns in viscoelastic full-waveform inversion. *Geophysics*, *85*(4), R397–R408.

Kolsky, H. (1952). The propagation of stress pulses in viscoelastic solids. *Philosophical Magazine*, *1*, 693–710.

Komatitsch, D., & Tromp, J. (2005). Introduction to the spectral-element method for 3-D seismic wave propagation. *Geophysical Journal International*, *139*, 806–822.

Krischer, L., Fichtner, A., Boehm, C., & Igel, H. (2018). Automated Large-Scale Full Seismic Waveform Inversion for North America and the North Atlantic. *J. Geophys. Res. Solid Earth*, *123*, 5902–5928.

Liu, H., Anderson, D. L., & Kanamori, H. (1976). Velocity dispersion due to anelasticity; implications for seismology and mantle composition. *Geophys. J. R. Astron. Soc.*, *47*(1), 41–58.

Liu, Q., & Tromp, J. (2006). Finite-Frequency Kernels Based on Adjoint Methods. *Bulletin of Seismological Society of America*, *96*(6), 2383–2397.

Liu, Q., & Tromp, J. (2008). Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods. *Geophysical Journal International*, *174*, 265–286.

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

Mavko, G. M., & Nur, A. (1979). Wave attenuation in partially saturated rocks. *Geophysics*, *44*(0), 161–178.

Métivier, L., Brossier, R., Virieux, J., & Operto, S. (2013). Full waveform inversion and the truncated Newton method. *SIAM Journal On Scientific Computing*, *35*, B401–B437.

Mulder, W. A., & Hak, B. (2009). An ambiguity in attenuation scattering imaging. *Geophysical Journal International*, *178*, 1614–1624.

Nocedal, J., & Wright, S. J. (2006). *Numerical optimization*. Springer.

Operto, S., Gholami, Y., Prieux, V., Ribodetti, A., Brossier, R., Métivier, L., & Virieux, J. (2013). A guided tour of multiparameter full waveform inversion with multicomponent data: from theory to practice. *The Leading Edge*, *32*, 1040–1054.

Operto, S., & Miniussi, A. (2018). On the role of density and attenuation in 3D multi-parameter visco-acoustic VTI frequency-domain FWI: an OBC case

study from the North Sea. *Geophysical Journal International*, *213*, 2037–2059.

Pan, W., Geng, Y., & Innanen, K. A. (2018). Interparameter trade-off quantification and reduction in isotropic-elastic full-waveform inversion: synthetic experiments and Hussar land data set application. *Geophysical Journal International*, *213*, 1305–1333.

Pan, W., Innanen, K. A., Margrave, G. F., Fehler, M. C., Fang, X., & Li, J. (2016). Estimation of elastic constants for HTI media using Gauss-Newton and full-Newton multiparameter full-waveform inversion. *Geophysics*, *81*(5), R275–R291.

Pan, W., Innanen, K. A., & Wang, Y. (2020). SeisElastic2D: An open-source package for multiparameter full-waveform inversion in isotropic-, anisotropic- and visco-elastic media. *Computers & Geosciences*, *145*, 104586.

Pan, W., & Wang, Y. (2020). On the influence of different misfit functions for attenuation estimation in viscoelastic full-waveform inversion: synthetic study. *Geophysical Journal International*, *221*(2), 1292–1319.

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

Quan, Y., & Harris, J. M. (1997). Seismic attenuation tomography using the frequency shift method. *Geophysics*, *62*, 895–905.

Robertsson, J. O., Blanch, J. O. A., & Symes, W. W. (1994). Viscoelastic finite-difference modeling. *Geophysics*, *59*, 1444–1456.

Romanowicz, B. (1995). A global tomographic model of shear attenuation in the upper mantle. *J. Geophys. Res. Solid Earth*, *100*(B7), 12375–12394.

Sager, K., Ermert, L., Boehm, C., & Fichtner, A. (2018). Towards full waveform ambient noise inversion. *Geophysical Journal International*, *212*(1), 566–590.

Shen, Y., Biondi, B., & Clapp, R. (2018). Q-model building using one-way wave-equation migration Q analysis — Part 1: Theory and synthetic test. *Geophysics*, *83*(2), S93–S109.

Shor, N. (1985). *Minimization Methods for Non-Differentiable Functions*. Springer-Verlag.

Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2009). Adjoint tomography of the southern California crust. *Science*, *325*, 988–992.

Tarantola, A. (1984). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, *49*, 1259–1266.

Tarantola, A. (1986). A strategy for nonlinear elastic inversion of seismic reflection data. *Geophysics*, *51*, 1893-1903.

Tong, P., Zhao, D., Yang, D., Yang, X., Chen, J., & Liu, Q. (2014). Wave-equation-based travel-time seismic tomography-Part 1: Method. *Solid Earth*, *5*(2), 1151–1168.

Trampert, J., & Woodhouse, J. H. (1995). Global phase-velocity maps of Love and Rayleigh-waves between 40 and 150 seconds. *Geophysical Journal International*, *122*(2), 675–690.

Tromp, J., & Bachmann, E. (2019). Source encoding for adjoint tomography. *Geophysical Journal International*, *218*(3), 2019–2044.

Tromp, J., Tape, C., & Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal, and banana-doughnut kernels. *Geophysical Journal International*, *160*, 195-216.

Virieux, A., & Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, *74*(6), WCC1–WCC26.

Wang, N., Li, J., Borisov, D., Gharti, H. N., Shen, Y., Zhang, W., & Savage, B. (2019). Modeling Three-Dimensional Wave Propagation in Anelastic Models With Surface Topography by the Optimal Strong Stability Preserving Runge-Kutta Method. *J. Geophys. Res. Solid Earth*, *124*(1), 890–907.

Wang, Y., Ning, Y., & Wang, Y. (2020). Fractional Time Derivative Seismic Wave Equation Modeling for Natural Gas Hydrate. *Energies*, *13*(22), 5901.

Wang, Y., & Yuan, Y. (2005). Convergence and regularity of trust region methods for nonlinear ill-posed inverse problems. *Inverse Problems*, *21*, 821–838.

Wu, R. (1989). The perturbation method in elastic wave scattering. *Pure and Applied Geophysics*, *131*(4), 605–637.

Yao, G., Silva, N. V., Warner, M., & Kalinicheva, T. (2018). Separation of migration and tomography modes of full-waveform inversion in the plane wave domain. *J. Geophys. Res. Solid Earth*, *123*, 1486–1501.

Yu, X., Ahmadinia, M., Shariatipour, S. M., Lawton, D., Osadetz, K., & Saeedfar, A. (2019). Impact of Reservoir Permeability, Permeability Anisotropy and Designed Injection Rate on CO_2 Gas Behavior in the Shallow Saline Aquifer at the CaMI Field Research Station, Brooks, Alberta. *Natural Resources Research*, *29*(4), 2735–2752.

Yuan, Y. O., Simons, F. J., & Bozdağ, E. (2015). Multiscale adjoint waveform tomography for surface and body waves. *Geophysics*, *80*(5), R281–R302.

Yuan, Y. O., Simons, F. J., & Tromp, J. (2016). Double-difference adjoint seismic tomography. *Geophysical Journal International*, *206*(3), 1599–1618.

Zhu, H., Bozdağ, E., Duffy, T. S., & Tromp, J. (2013). Seismic attenuation beneath Europe and the North Atlantic: Implications for water in the mantle. *Earth and Planetary Science Letters*, *381*, 1–11.

Zhu, H., & Tromp, J. (2013). Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean. *Science*, *341*, 871–875.

Zhu, T., Ajo-Franklin, J. B., & Daley, T. M. (2017). Spatiotemporal changes of seismic attenuation caused by injected CO2 at the Frio-II pilot site, Dayton, TX, USA. *J. Geophys. Res. Solid Earth*, *122*, 7156–7171.

Zhu, T., & Carcione, J. M. (2014). Theory and modelling of constant-Q P- and S-waves using fractional spatial derivatives. *Geophysical Journal International*, *196*, 1787–1795.

Zhu, T., Harris, J. M., & Biondi, B. (2014). Q-compensated reverse-time migration. *Geophysics*, *79*(3), R471–R484.