Inversion strategies for Q estimation in viscoacoustic full-waveform inversion

Peng Yong, Romain Brossier, Ludovic Métivier

Inversion strategies for Q estimation in visco-acoustic full-waveform inversion

Peng Yong [*1,3], Romain Brossier[1] and Ludovic Métivier[2,1]

[1] Univ. Grenoble Alpes, ISTerre, F-38000 Grenoble, France

[2] Univ. Grenoble Alpes, CNRS, LJK, F-38000 Grenoble, France

[3] Currently at Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

Email address: yongpeng@mail.ioa.ac.cn

romain.brossier@univ-grenoble-alpes.fr

ludovic.metivier@univ-grenoble-alpes.fr

(May 7, 2024)

Running head: Visco-acoustic full-waveform inversion

ABSTRACT

Estimation of an attenuation parameter, represented by the quality factor Q, holds paramount importance in seismic exploration. One of the main challenges in Q estimation through visco-acoustic full-waveform inversion is effectively decoupling Q from velocity. In this study, our objective is to enhance Q inversion by addressing critical aspects, including gradient preconditioning, workflow, and misfit design. By developing a new preconditioner that approximates the diagonal of the Hessian, we facilitate automatic parameter tuning across different classes, ensuring comparable magnitudes of preconditioned gradients for velocity and Q. Moreover, Our investigations confirm the efficacy of the two-stage hierarchical strategy in mitigating velocity-Q trade-offs, enabling more accurate Q estimation by first

focusing on velocity reconstruction before jointly estimating velocity and Q. The analysis and numerical examples also highlight the importance of broadband data and long-offset acquisition for a reliable Q estimation. Additionally, leveraging amplitude information can improve Q estimation to some extent, but careful consideration of frequency band and noise effects is necessary. We have explored two misfit functions that capture amplitude variation with frequency in the time-frequency domain, noting their sensitivity to noise. To address this, we propose a differential strategy that can effectively mitigate the effects of low-frequency noise. This comprehensive study on enhancing Q estimation in visco-acoustic FWI offers valuable insights for multi-parameter inversion in realistic scenarios.

INTRODUCTION

Over the past decade, full-waveform inversion (FWI) has demonstrated its capability to reconstruct high-resolution velocity models (Sirgue et al., 2010; Virieux et al., 2017). Currently, there is a growing interest in extending mono-parameter FWI, primarily focused on velocity, to multi-parameter reconstruction to consider the more complex and realistic physics of wave propagation through the Earth (Operto et al., 2013). Attenuation parameter estimation, characterized by the quality factor Q, holds great importance in exploration-scale studies, providing valuable information about subsurface fluid/gas content (Carcione, 2015). Additionally, it plays a key role in Q-compensated migration for migration focusing (Zhu et al., 2014). The scattering radiation patterns of Q and velocity only differ by a phase shift (Mulder and Hak, 2009; Hak and Mulder, 2011; da Silva et al., 2019a). For this reason, effectively decoupling Q from velocity presents significant challenges (Malinowski et al., 2011; Dutta and Schuster, 2016; da Silva et al., 2019a; Jiang, 2019; Keating and Innanen, 2019; Yang et al., 2020; Kamath et al., 2021; Xing, 2022; Pan et al., 2023b). In this study, we aim at enhancing Q estimation in visco-acoustic FWI.

Reducing interparameter trade-off or crosstalk is one of the intrinsic difficulties in multi-parameter FWI. Some efforts have been made to choose a suitable parameterization based on scattering radiation patterns (Operto et al., 2013; Alkhalifah and Plessix, 2014; Kamath and Tsvankin, 2016; He and Plessix, 2017). However, the effectiveness of this strategy may be reduced in complex geology environments with limited diffraction angle ranges. Since the Hessian operator contains information about interactions between different parameter classes, applying the inverse Hessian operator to the gradient during the optimization process in a truncated-Newton-like step scheme can limit trade-offs (Métivier et al., 2015; Pan

et al., 2017; Yang et al., 2018; Gao et al., 2021; Yong et al., 2022; Xie et al., 2023). Nevertheless, the ill-posedness of the Hessian matrix and the computational cost of constructing Hessian-vector products pose challenges for its application to 3D field data.

Velocity, as the first-order kinematic parameter, strongly influences the phase information in seismograms. The attenuation parameter can also affect the seismic phase through the dispersion relation, but its effects are most evident in the amplitude information (Wang, 2009; Carcione, 2015). In practice, a two-stage workflow is commonly adopted to mitigate cross-talk between velocity and attenuation parameters, with the first stage focusing on dominant velocity parameter reconstruction and the second stage targeting attenuation parameter estimation. One may consider a straightforward sequential workflow that updates velocity only in stage 1 and Q only in stage 2. However, from our numerical experience, addressing data misfit with the secondary parameter (Q) alone can readily lead to over-updating when the primary parameter (velocity) lacks accuracy. This observation aligns with findings from earlier studies (Prieux et al., 2013; Wu et al., 2021). In this work, we will focus on two other kinds of two-stage workflows: simultaneous inversion and hierarchical inversion. Simultaneous inversion, recommended by Wu et al. (2021), refers to the joint reconstruction of velocity and Q at both stages. In this strategy, the reconstructed Q model from the first stage is discarded. Conversely, hierarchical inversion updates only the velocity parameter, making it as kinematically accurate as possible in the first stage, while both velocity and Q are jointly reconstructed in the second stage (Prieux et al., 2013; Kamei and Pratt, 2013).

When jointly updating multiple parameters in FWI, gradients with values of varying magnitudes for different parameter classes (velocity and Q) are often encountered, even with a unity-based normalization technique that makes the involved parameters dimensionless

and ranges their values from zero to one (Prieux et al., 2013; Yang et al., 2018). A scaling strategy is commonly used to precondition the gradient for better control of the magnitudes of updates (Kamei and Pratt, 2013; Yang et al., 2018). However, previous studies reveal that Q estimation significantly varies when the scaling number is modified (Kamei and Pratt, 2013; Prieux et al., 2013; Operto and Miniussi, 2018). Careful selection and optimization of the scaling parameter are essential to achieve robust and consistent results in Q estimation, while finding a suitable scaling parameter is not trivial. To address the issue of unbalanced illumination for different parameters, we develop a preconditioner by approximating the diagonal of the Hessian. This approach makes it possible to automatically tune parameters of different classes, ensuring that the values of preconditioned gradients for velocity and Q are at a similar order of magnitude. By implementing this preconditioner, we aim for stabler and more reliable updates during joint parameter estimation in FWI without the need for special tuning of the scaling parameter.

In exploration seismology, the constant Q model over frequency has been widely adopted to approximate seismic attenuation. In such cases, the attenuation effect on waveform linearly increases with frequency (Carcione, 2015). As a result, the centroid of the signal spectrum experiences a downshift during wave propagation. To estimate Q, Quan and Harris (1997) propose using the centroid frequency shift in seismic tomography, which is relatively insensitive to geometrical spreading, reflection, and transmission effects. Note that this method requires the signal to have a broad bandwidth, leading to significant centroid frequency shifts. It has been used to define the misfit function of FWI for more reliable Q estimation (Dutta and Schuster, 2016; Pan et al., 2023a), but its applicability is limited to single-phase data, such as identifiable first-arrivals. To handle realistic multi-event signals, we explore the use of a time-frequency analysis method based on the Gabor transform

(Gabor, 1946) to compute the instantaneous centroid frequency (Yong et al., 2021). We specifically evaluate its performance within the frequency band commonly considered in FWI applications. In addition, we propose a misfit function employing frequency-weighted amplitude information for Q inversion, which can reduce the sensitivity of the instantaneous centroid frequency to velocity errors (Yong et al., 2021). Although frequency-weighted amplitude (FWA) can capture the distinctive spectral modifications caused by attenuation and provide more reliable Q estimation, we observe that this measurement is highly sensitive to noise, even Gaussian noise. To address this, we take the time derivative of the FWA measurement, inspired by the double-difference strategy (Zhang and Thurber, 2003), with the assumption that the time-spectrum of noise varies slowly.

The structure of the paper is organized as follows: In the theory section, we first analyze the effect of dissipation and dispersion on seismic wave propagation, especially in the low-frequency band, to better understand the expected Q estimation from waveform inversion. Then, we recall the basics of FWI based on the adjoint-state method, with a focus on the preconditioner and a unity-based normalization strategy for multi-parameter inversion. Next, we present three misfit functions for Q reconstruction, aiming at minimizing the differences in waveforms (conventional L^2 misfit), instantaneous centroid frequencies, and frequency-weighted amplitudes of observed and predicted data. In the numerical examples section, we utilize a portion of the synthetic BP model, representative of the geology of the Western Gulf of Mexico made available by the BP company, to validate the developed preconditioner and compare two kinds of two-stage workflows. The numerical tests demonstrate that the hierarchical strategy yields a faster convergence rate and provides a more accurate Q estimation. Following this, we implement the hierarchical strategy in a more realistic synthetic 2D Valhall case study, representing the shallow water environment in the

North sea. In this study, we first highlight the significance of broadband data and long-offset acquisition for Q estimation. Subsequently, we conduct a comparative analysis of the previously mentioned misfit functions using this 2D Valhall model. Finally, discussions and conclusions are presented.

METHODOLOGY

In this section, we first detail the time-domain equations of acoustic wave modeling in a linear isotropic viscoelastic medium, based on the generalized Zener body model. Next, we make a brief analysis about the effect of constant Q model on the low-frequency seismic data to better understand the Q inversion. Then, we present the general framework of FWI and three misfit functions for Q estimation.

Generalized Zener body

Time-domain seismic wave propagation in a constant Q medium can be expressed by the generalized Zener body model (Moczo and Kristek, 2005; Yang et al., 2016; da Silva et al., 2019b), and the 3D first-order visco-acoustic wave equation can be written as (Yang et al., 2018)

$$\begin{aligned}
&\partial_t \mathbf{v}(\mathbf{x},t) = \frac{1}{\rho(\mathbf{x})}\nabla p(\mathbf{x},t), \quad \mathbf{v} = (v_x, v_y, v_z),\\
&\partial_t p(\mathbf{x},t) = M_U(\mathbf{x})\left(\nabla\cdot\mathbf{v}(\mathbf{x},t) - Q^{-1}(\mathbf{x})\sum_{\ell=1}^{n} Y_\ell \xi_\ell(\mathbf{x},t)\right),\\
&\partial_t \xi_\ell(\mathbf{x},t) + \omega_\ell \xi_\ell(\mathbf{x},t) = \omega_\ell \nabla\cdot\mathbf{v}(\mathbf{x},t), \quad \ell = 1,2,\ldots,n.
\end{aligned} \tag{1}$$

In equation (1), $p(\mathbf{x},t)$ is the pressure field, $\mathbf{v}$ denotes the particle velocity, and ξ_ℓ are the memory variables, each one associated with one reference frequency ω_ℓ. The parameter ρ is the density and the unrelaxed modulus M_U is related to the wave speed v via $M_U = \rho v^2$. Finally, Q is the quality factor. Here, v denotes the P-wave velocity in a purely elastic material. The actual wave propagation velocity in a viscoelastic material is frequency-dependent, which will be analyzed in the following subsection.

Note that Y_ℓ are dimensionless auxiliary scalar variables, determined by solving the following optimization problem for adjusting the constant Q model over a given frequency band $\Omega = [\omega_{min}, \omega_{max}]$

$$Y_\ell = \arg\min_{y_\ell} \int_{\omega\in\Omega} \left(\sum_{\ell=1}^{n} y_\ell \frac{\omega\omega_\ell}{\omega_\ell^2 + \omega^2} - 1 \right)^2 \mathrm{d}\omega. \tag{2}$$

For three relaxation mechanisms ($\ell = 3$), as used in this work, ω_ℓ are chosen as ω_{min}, $\sqrt{\omega_{min}\omega_{max}}$ and ω_{max}.

Analysis on the effect of Q

In this part, we first introduce the definition of Q from the stress-strain relation. Then, we build the link between Q and complex-valued velocity, to analyze the effect of viscosity on wave propagation.

For a linear isotropic viscoelastic material, the stress-strain relation can be related to the generalized Hooke's law as

$$\sigma(t) = M(t) * \varepsilon(t), \tag{3}$$

where $*$ denotes time convolution and $M(t)$ is the stress response to the Dirac δ-function in strain. In the frequency domain, the convolution becomes a multiplication and the relation becomes

$$\sigma(\omega) = M(\omega)\varepsilon(\omega), \tag{4}$$

where ω denotes the frequency. The quality factor $Q(\omega)$ is defined by

$$Q(\omega) = \frac{Re(M(\omega))}{Im(M(\omega))}, \tag{5}$$

where Re and Im take real and imaginary parts of the complex-valued modulus. The physical meaning of Q is the number of wavelengths a wave propagates through the medium before its amplitude has decreased by a factor of $e^{-\pi}$. The frequency-dependent modulus $M(\omega)$ corresponding to the visco-acoustic wave equation (1) can be expressed as

$$M(\omega) = M_U - M_U Q^{-1} \sum_{\ell=1}^{n} Y_\ell \frac{\omega_\ell}{\omega_\ell + i\omega}. \tag{6}$$

The unrelaxed (elastic) modulus $M_U = \lim_{\omega \to \infty} M(\omega)$ represents an instantaneous elastic response of the viscoelastic material (Moczo and Kristek, 2005).

The complex-valued velocity $c(\omega)$ is related to the frequency-dependent modulus $M(\omega)$ and the density ρ by:

$$c(\omega) = \sqrt{\frac{M(\omega)}{\rho}}. \tag{7}$$

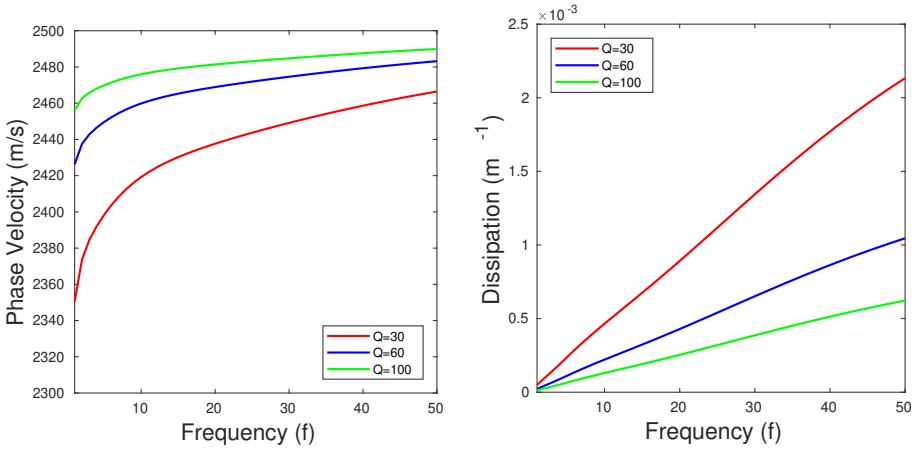
The numerical phase velocity is the real part of $c(\omega)$, which is used to describe the dispersion effect. The imaginary part of $c(\omega)$ is related to the dissipation effect.

To understand the dissipation effect of viscosity, let us consider a plane wave solution of the wave equation:

$$U(r + \Delta r, \omega) = U(r, \omega) e^{ik\Delta r}, \quad k = \frac{\omega}{c(\omega)} = \alpha + i\beta, \tag{8}$$

where k is the wavenumber. Considering the phase-shifting property, we know that the amplitude will decrease by a factor of $e^{-\beta \Delta r}$ after propagating a distance of Δr. Here, β is used to characterize the dissipation effect (Wang, 2009).

Figure 1 presents the phase velocity change (dispersion) and amplitude attenuation (dissipation) of different Q values, under the constant-Q assumption. The smaller the value



(a) (b)

Figure 1: The (a) dispersion and (b) dissipation effects within different Q values in a medium $v = 2500$ m/s.

of Q, the stronger dispersion and dissipation are. In addition, the phase velocity change caused by Q is relatively smaller than the velocity change observed between different geology strata. This indicates that the seismic diffraction (strong interaction) caused by Q is small, and the main effect of Q should be a weak interaction, which accumulatively increases along the wave path. In addition, we can infer that, when the velocity model (also, density and anisotropy) is not known with sufficient accuracy, it is difficult to retrieve Q models with resolution as high as velocity, due to the small effect on phase velocity. It is important to also point out that the amplitude attenuation linearly increases with frequency in constant Q media. This property could be exploited to design misfit function for more robust Q estimation by FWI.

General formulation of FWI

FWI updates the model by minimizing the misfit function $\mathbb{J}$ defined between observed and predicted data, which can be formulated as a PDE-constrained optimization problem of the

form:

$$\min_{\mathbf{m}} \mathbb{J}(u[\mathbf{m}], d) \quad \text{subject to} \quad A(\mathbf{m})w = s, \tag{9}$$

where $\mathbf{m}$ represents the model parameters of interest in the model space $\mathcal{M} \subset \mathbb{R}^n$, $d := d(\mathbf{x}_r, t)$ is the observed data at receiver location $\mathbf{x}_r$, and $u = Rw$ represents the predicted data extracted from the wavefield at the receiver location. The operator $A(\mathbf{m})$ corresponds to the forward modeling operator (equation (1) in this paper), and s is the source term.

For large-scale FWI applications, the commonly used approach is gradient-based local optimization for updating the model parameters. The update equation at the $(k+1)$-th iteration is given by

$$\mathbf{m}^{k+1} = \mathbf{m}^k - \alpha^k P^{-1} \nabla \mathbb{J}(\mathbf{m}^k), \tag{10}$$

where α^k represents the step length at the k-th iteration, determined by the linesearch method with classical Wolfe conditions. P denotes the preconditioner designed to accelerate the convergence rate. The gradient of the misfit function, $\nabla \mathbb{J}(\mathbf{m}^k)$, can be calculated via the adjoint-state method

$$\nabla \mathbb{J}(\mathbf{m}) = \int_0^T \lambda(t) \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}} w(t) \mathrm{d}t, \tag{11}$$

where λ is known as the adjoint wavefield. The adjoint field λ is calculated by back-propagating the receiver data residuals (Tarantola, 1984; Plessix, 2006) with the adjoint-state equation

$$A^T(\mathbf{m})\lambda = -R^T r, \tag{12}$$

Here, the adjoint operator $A^T(\mathbf{m})$ has a similar structure to the wave-modeling operator $A(\mathbf{m})$. The adjoint source r, positioned at the receiver locations, is related to the derivative

of the misfit function with respect to the predicted data.

In the case of multi-parameter inversion, the preconditioner P should consider the effects of the Hessian to correctly scale the gradients of different parameter classes. The preconditioner used in this work is based on source energy illumination, ensuring that the preconditioned gradients of different parameters are in the same (or similar) order of magnitude. The detailed derivation can be found in Appendix A. The preconditioner P is defined as follows

$$P_i = \sqrt{\int_0^T \left(w(t) \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(t) \right)^2 \mathrm{d}t}. \tag{13}$$

In addition, considering the varying magnitudes and physical units of velocity, density, and Q, directly combining them into a single vector is impractical both in terms of physical interpretation and computational feasibility. To address this issue, a unity-based normalization technique (Prieux et al., 2013; Yang et al., 2018) is employed. The normalization equation is given as

$$\tilde{m}_i = \frac{m_i - m_i^{min}}{m_i^{max} - m_i^{min}}, \tag{14}$$

Here, i represents the ith parameter class of interest. The values m_i^{min} and m_i^{max} correspond to the lower and upper bounds of the physical parameter. By applying this normalization strategy, all model parameters are confined in the range [0, 1]. Furthermore, utilizing the chain rule, both the gradient and preconditioner are appropriately scaled.

Three misfit functions for Q inversion

In conventional FWI, the misfit function is defined to minimize the least-squares difference between observed and predicted waveforms

$$\mathbb{J}_{WF} = \frac{1}{2}\int_{\mathbb{R}} (u(t) - d(t))^2 \mathrm{d}t, \tag{15}$$

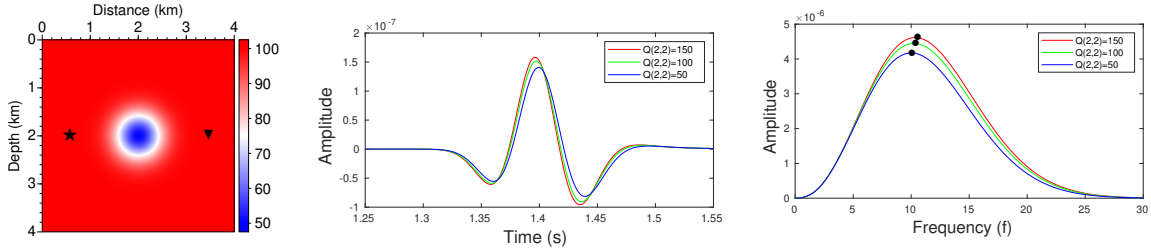
where $d(t)$ represents the recorded data and $u(t)$ represents the predicted data. The adjoint source of waveform measurement can be given by

$$r_{WF} = u(t) - d(t). \tag{16}$$

The attenuation effect on seismic waveform linearly increases with frequency (Carcione, 2015). As a result, the centroid of a signal spectrum is affected by a downshift during wave propagation. Quan and Harris (1997) introduce the concept of centroid frequency to characterize this peculiar change, which can be calculated by

$$\tilde{\omega}_u^c = \frac{\int_{\mathbb{R}} |\omega| A_u^2 \mathrm{d}\omega}{\int_{\mathbb{R}} A_u^2 \mathrm{d}\omega}, \quad A_u = |\tilde{u}(\omega)|, \tag{17}$$

where $|\tilde{u}(\omega)|$ is the amplitude spectrum resulting from the Fourier transform of a given signal $u(t)$. Figure 2 illustrates the phenomenon of centroid frequency downshift resulting from a biased attenuation of frequencies. Let us mention that, in principle, the applicability of equation (17) is limited to single-event data. Considering the non-stationarity of seismic data that different events may have different centroid frequencies, it can be more suitable to extract the instantaneous centroid frequency (ICF) to exploit this feature. The ICF of a



(a) (b) (c)

Figure 2: An illustration of centroid frequency downshift: (a) an anomaly Q model with $v = 2500\ m/s$ and $\rho = 1\ g/cm^3$ and Q model is defined as $Q = 100 + Q_0 \exp(-4((z-2)^2 + (x-2)^2))$, (b) recorded data with three different Q models, (c) corresponding amplitude spectra. The source and receiver, situated at a distance of 3 km from each other, are denoted by star and triangle in (a), respectively. The horizontal coordinate of the black points denotes the centroid frequency. Please note the centroid frequency downshift.

signal $u(t)$ can be defined as

$$\omega_u^c(t) = \frac{\int_{\mathbb{R}} |\omega| A_u^2(t,\omega) \mathrm{d}\omega}{\int_{\mathbb{R}} A_u^2(t,\omega) \mathrm{d}\omega}, \quad A_u(t,\omega) = |\hat{u}(t,\omega)|, \tag{18}$$

where $\hat{u}(\omega, t)$ is the time-frequency spectrum of $u(t)$ obtained through a Gabor transform. The specific Gabor transform used in this work is defined in Strang and Nguyen (1996) as

$$\hat{f}(t,\omega) = \mathcal{G}[f](t,\omega) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f(\xi) h_\sigma^\dagger(\xi - t) e^{-i\omega\xi} \mathrm{d}\xi, \tag{19}$$

$$f(t) = \mathcal{G}^{-1}[\hat{f}](t) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}^2} \hat{f}(\xi,\omega) h_\sigma(t-\xi) e^{i\omega t} \mathrm{d}\xi \mathrm{d}\omega. \tag{20}$$

Here, † denotes complex conjugate, and the window function $h_\sigma(t)$ is a normalized Gaussian given by $h_\sigma = (\pi\sigma^2)^{-\frac{1}{4}} e^{\frac{-t^2}{2\sigma^2}}$, where σ controls the radius of the window function. To account

for the ICF shift caused by dissipation, a new misfit for FWI can be defined as follows

$$\mathbb{J}_{ICF} = \frac{1}{2}\int_{\mathbb{R}} W(t)(\omega_u^c(t) - \omega_d^c(t))^2 \mathrm{d}t, \tag{21}$$

Here, the weight function $W(t)$ is defined as $\log(1 + \int_{\mathbb{R}} A_d(\omega, t)\mathrm{d}\omega)$, which is an integration of the time-frequency spectrum over frequency. This choice of weight function is inspired by the one used for phase measurement in Fichtner et al. (2008), and is designed to ensure the robustness of measurements. Additionally, the logarithmic function is employed to alleviate the potential dominance of large amplitudes. The denominator in equation (18), acting as a normalization factor, diminishes the influence of absolute amplitude on the ICF measurement. This prioritizes the ICF measurement's focus on the centroid frequency change induced by dissipation, however it may be sensitive to velocity errors due to the limited-band seismic data used in FWI. The adjoint source of the ICF measurement is written as

$$r_{ICF} = 2\mathcal{G}^{-1}\left[W(t)(\omega_u^c(t) - \omega_d^c(t))\frac{|\omega| - \omega_u^c(t)}{\int_{\mathbb{R}} A_u^2(t,\omega)\mathrm{d}\omega}\hat{u}(t,\omega)\right]. \tag{22}$$

To improve the robustness of Q estimation, we propose using a new measurement function called the frequency-weighted amplitude (FWA) function. The FWA of a signal $u(t)$ is defined as

$$\mathcal{A}_u(t) = \int_{\mathbb{R}} |\omega| A_u(\omega, t)\mathrm{d}\omega, \tag{23}$$

The corresponding L^2 norm misfit using FWA information can be written as

$$\mathbb{J}_{FWA} = \frac{1}{2}\int_{\mathbb{R}} (\mathcal{A}_u(t) - \mathcal{A}_d(t))^2 \mathrm{d}t. \tag{24}$$

The FWA misfit function emphasizes the decay of high-frequency amplitudes. In FWI, we can first obtain high-resolution velocity models that can account for the main traveltime and amplitude information. Q estimation can then be performed by leveraging the feature that energy loss linearly increases with frequency. The adjoint source for FWA is given by

$$r_{FWA} = \mathcal{G}^{-1}\left[(\mathcal{A}_u(t) - \mathcal{A}_d(t))|\omega|\frac{\hat{u}(\omega,t)}{A_u(\omega,t)}\right]. \tag{25}$$

We generate the data presented in Figure 3(a) using three 2D homogeneous models with different P-wave velocity and attenuation values (Vp=2.5 km/s, Qp=∞; Vp=2.6 km/s, Qp=∞; Vp=2.5 km/s, Qp=60). A 5 Hz Ricker wavelet is utilized as time source function. The same source-receiver geometry is used to record the data, and the offset is 3 km.

Conventional FWI attempts to minimize the difference between waveforms (Figure 3(b)). It can be observed that the waveform difference may be more sensitive to velocity change. The ICF of the data recorded in the lossless media (green and red lines in Figure 3(c)) have the same shape with a slight time shift. Due to the limited low-frequency band, the centroid frequency shift caused by viscosity is small (blue lines in Figure 3(c) and Figure 3(d)), which is consistent with the previous analysis. Figure 3(d) also shows that small velocity error may generate a large effect on the Q inversion using ICF measurement. The FWA measurement, which enhances the high-frequency decay, might better capture the effect of viscosity for band-limited data, and might be more efficient and robust for Q estimation.

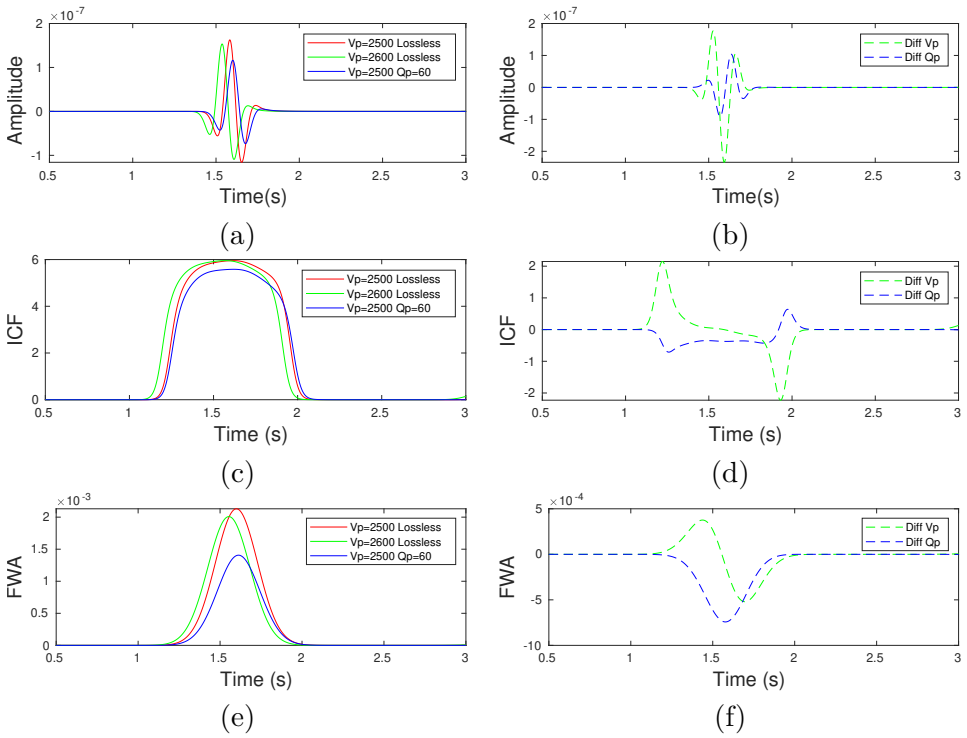


Figure 3: An intuitive illustration of three measurements: (a) presents the recorded waveform in three different parameter settings. The ICF and FWA information are displayed in (c) and (e). The dash green (blue) lines in (b,d,f) are the difference between the solid green (blue) and red lines in (a,c,e), respectively.

NUMERICAL EXPERIMENTS

We utilize a section of the synthetic model made available by BP where the gas cloud is situated at a shallow depth. This model is commonly employed for the study of Q modeling, migration, and inversion (e.g., Zhu et al., 2014; Pan et al., 2023a). In our investigation, we employ this portion to examine two types of two-stage workflows and assess the performance of the developed preconditioner. Subsequently, we present a more realistic example using the Valhall model to demonstrate the impact of long-offset acquisition and broadband data. Additionally, we study the misfit functions for Q estimation and introduce the differential strategy to enhance the robustness of the frequency-weighted amplitude (FWA) measurement against noise.

BP model

In this numerical experiment, we aim to compare two two-stage sequential strategies for multi-parameter inversion involving Vp and Q: simultaneous inversion and hierarchical inversion. Simultaneous inversion refers to the joint reconstruction of velocity and Q at both stages. In this approach, the reconstructed Q model from the first stage is discarded. Consequently, the inversion of the second stage begins with a well-reconstructed velocity model and an initial Q model. Conversely, hierarchical inversion updates only the velocity parameter in the first stage, while both velocity and Q are jointly reconstructed in the second stage. Additionally, we shall illustrate the significance of the preconditioner in our study.

Figures 4 and 5 depict the true and initial models of velocity and $1/Q$, respectively. For the generation of observed data, we utilize a Ricker wavelet with a dominant frequency of 8 Hz, resulting in 32 shots. At a depth of 60 m, we position 200 receivers at intervals of

25 m. In Figure 6 (a-b), the gradients of velocity and $1/Q$ are presented, revealing that the gradient values of different parameter classes exhibit a discrepancy of approximately one order of magnitude. To address this, Figure 6 (c-d) illustrates how the developed preconditioner effectively mitigates the significant disparities observed in the gradients of the multi-parameter FWI.

We use the ℓ-BFGS method ($\ell = 5$) developed in the SEISCOPE optimization tool box (Métivier and Brossier, 2016) to update the models of both velocity and $1/Q$. In Figure 7 (a-b), the reconstructed velocity and $1/Q$ at the first stage of simultaneous inversion are displayed. Notably, the velocity model exhibits a well-reconstructed appearance. Upon comparison with the true velocity model, we observe that the Q footprint on velocity reconstruction is minimal. However, it is evident that the reconstructed Q model is significantly influenced by the velocity footprint, with notable variations observed.

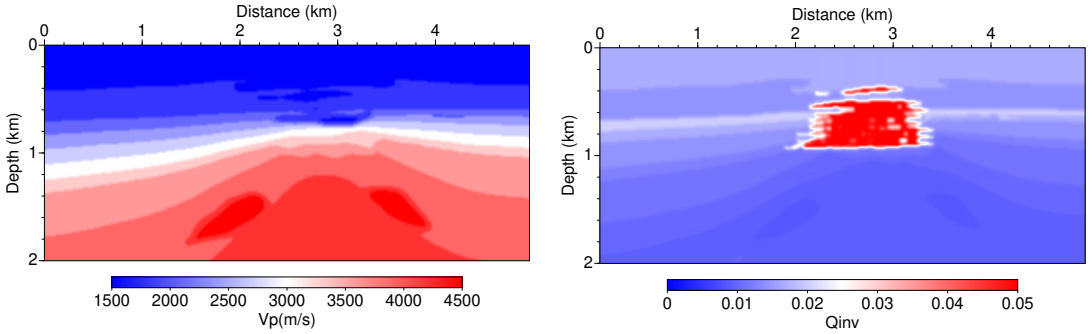
Figure 7 (c-d) showcases the reconstructed velocity and $1/Q$ at the second stage of simultaneous inversion. It is noteworthy that the update of the velocity model is minor in this stage. Moreover, a crucial observation is that the strong velocity footprint can be significantly mitigated as the starting velocity model becomes more accurate. This test underscores the notion that the expected Q model derived from FWI should be much smoother compared to the velocity model. This finding aligns with the understanding that the influence of Q on wave propagation consists of a weak interaction, accumulating along the wave path.

Figure 8 exhibits the outcomes of hierarchical inversion. A comparison with simultaneous inversion reveals that the Q model obtained through hierarchical inversion within 30 iterations is closer to the true Q model. Additionally, the disparity between the two

reconstructed velocity models is considerably reduced. To elucidate this phenomenon, we present the evolution of data and model misfits (Figure 9). In simultaneous inversion, velocity and Q parameters are employed to address the data discrepancy between the observed and calculated data. However, when the Q model restarts with the initial one at the onset of the second stage, a jump in data misfit occurs due to the incompatible combinations of velocity and Q models. In contrast, hierarchical inversion only updates the velocity in the first stage, resulting in velocity and Q models that are compatible with respect to accounting for data discrepancy. This incompatibility issue in simultaneous inversion somewhat hinders the convergence rate.

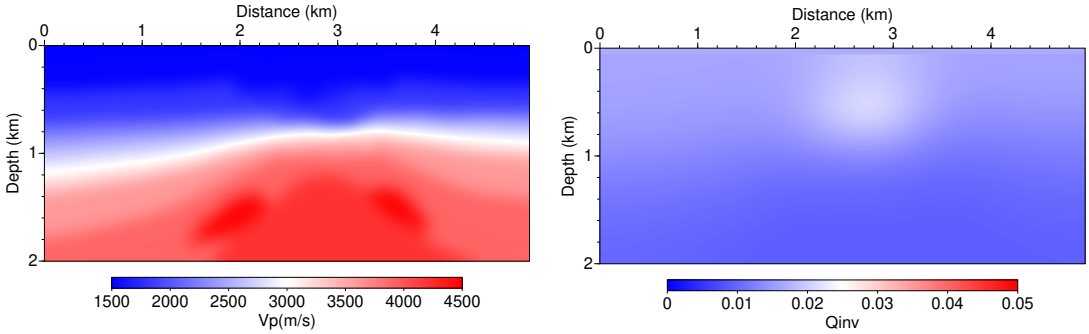
To provide a more intuitive illustration of the effect of the incompatibility issue, we examine the gradients and preconditioned gradients of the first iteration at the second stage. Figure 10 showcases the preconditioned gradients, revealing that the data discrepancy resulting from the change in the Q model is addressed by the velocity and Q parameters, with a noticeable leakage from the Q gradient to the velocity gradient. Furthermore, Figure 11 (c-d) demonstrates that the preconditioned gradients more accurately represent the differences between the true models and the given starting models. As a result, a faster convergence rate can be achieved.

This comparison between gradients and preconditioned gradients once again emphasizes the effectiveness of the developed preconditioner in mitigating significant disparities in the gradients of the multi-parameter FWI. Figure 11 (c-d) shows that the gradient value of velocity is at least one order of magnitude larger than that of $1/Q$, even though the primary data discrepancy is caused by the Q model. Interestingly, the value of the preconditioned gradient of velocity becomes smaller than that of $1/Q$. Hence, this indicates that the preconditioner effectively adjusts the contribution of different parameter classes in interpreting



(a) (b)

Figure 4: True models (part of the BP model): (a) P-wave velocity, (b) $1/Q$ model.



(a) (b)

Figure 5: Smooth starting models: (a) P-wave velocity, (b) $1/Q$ model.

data discrepancy.

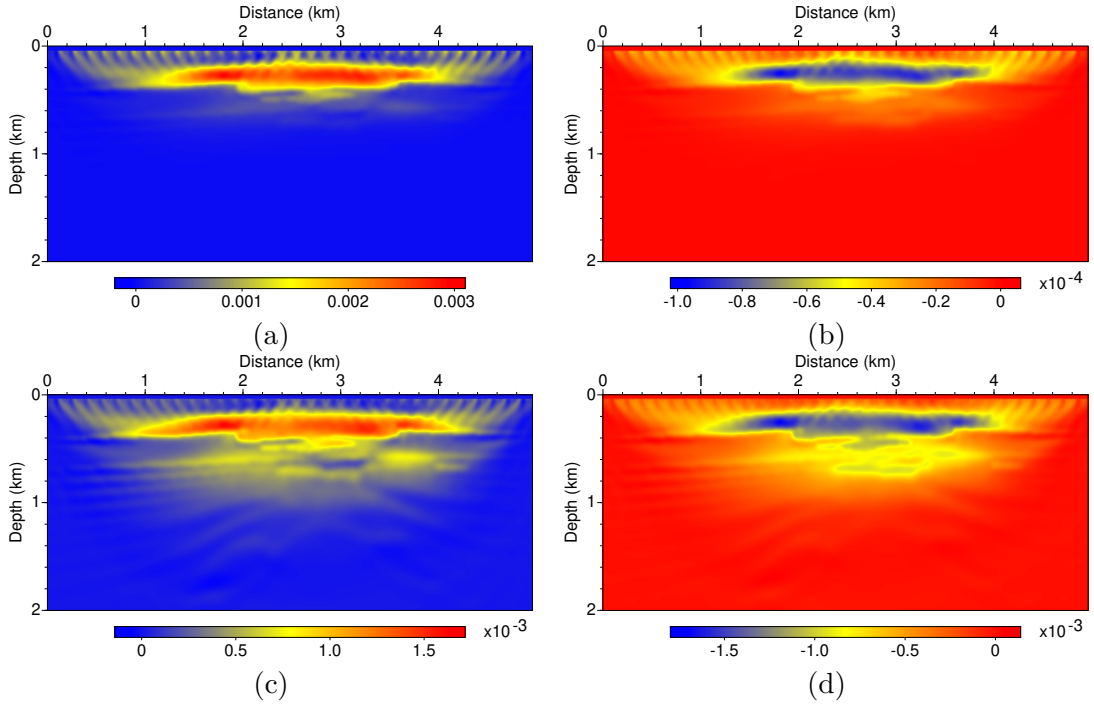


Figure 6: The developed preconditioner can align the gradients of different parameter classes to a similar order of magnitude: (a) gradient of P-wave velocity, (b) gradient of $1/Q$, (c) preconditioned gradient of P-wave velocity, (d) preconditioned gradient of $1/Q$.

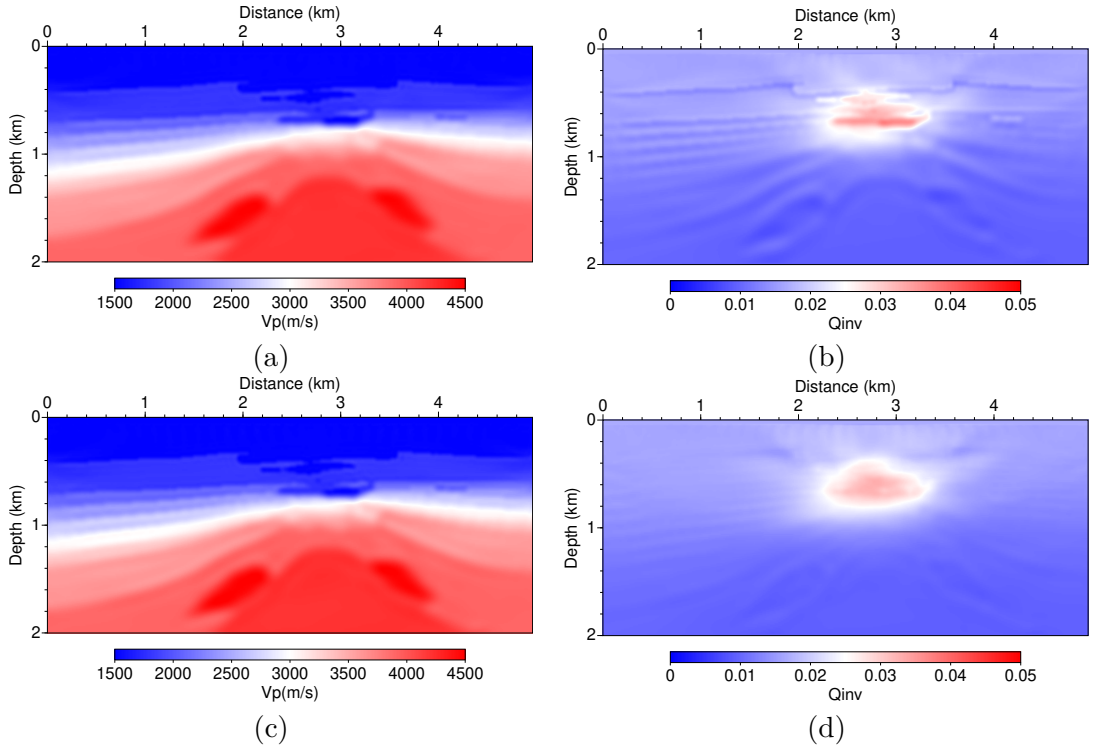


Figure 7: Simultaneous inversion of multiple parameters: Top row presents (a) Vp and (b) $1/Q$ results of stage 1 that starts with the smooth models shown in Figure 5. The bottom row displays the final (c) Vp and (d) $1/Q$ models at stage 2 that starts with the obtained Vp model and smooth $1/Q$ model (Figure 5 (b)).

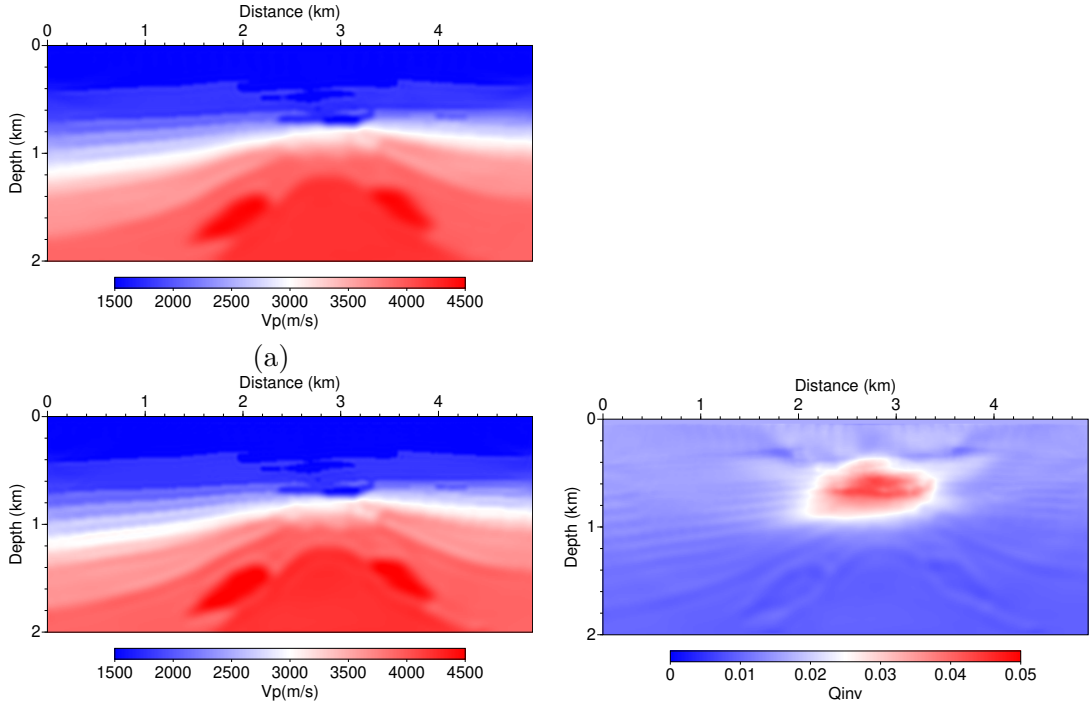


Figure 8: Hierarchical inversion of multiple parameters: the P-wave velocity is reconstructed at stage 1, while Vp and $1/Q$ models are jointly updated at the sequential stage. The final (a) Vp model obtained with 30 iterations at stage 1, and the final (b) Vp and (c) $1/Q$ models reconstructed from stage 2.

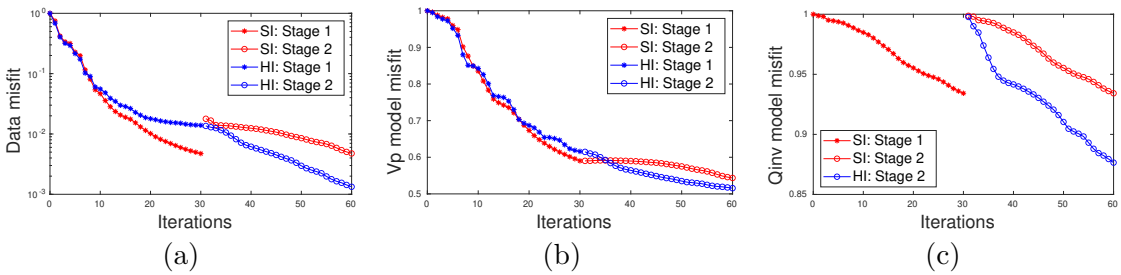


Figure 9: Comparison of convergence rate between two two-stage sequential strategies: (a) data misfit, (b) Vp model misfit, and (c) $1/Q$ model misfit. Please note the jump of the data misfit at the beginning of stage 2 in the simultaneous inversion.

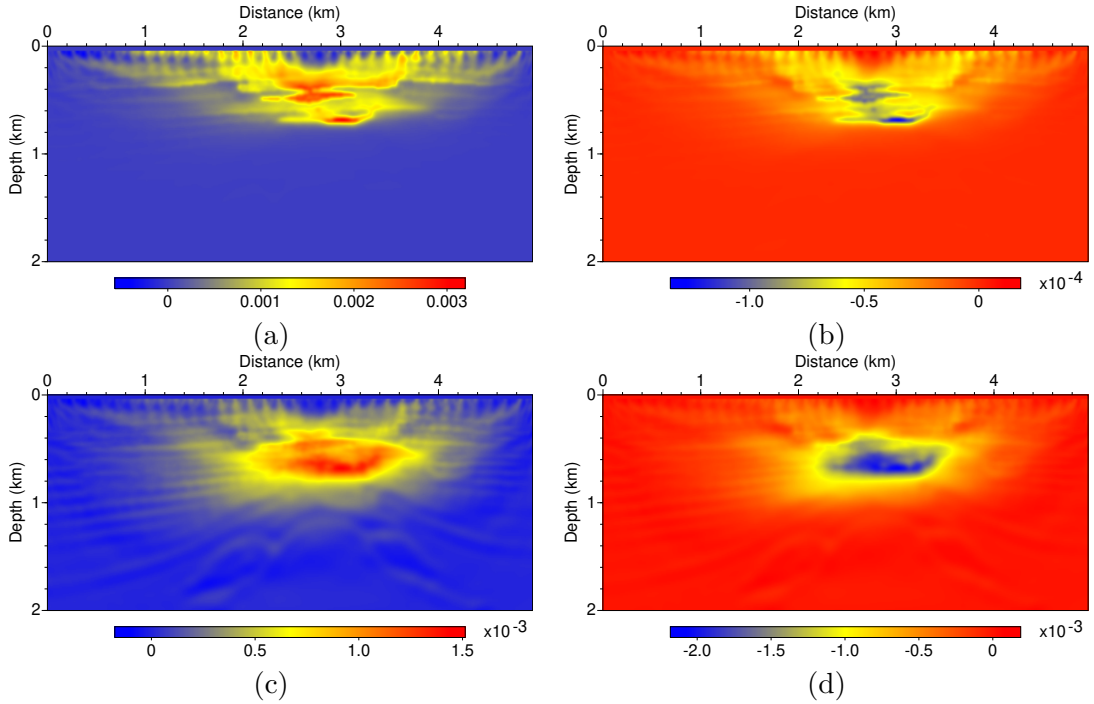


Figure 10: Comparison of the gradients of simultaneous inversion in the first iteration of stage 2: gradients of (a) Vp and (b) $1/Q$ and preconditioned gradients of (c) Vp and (d) $1/Q$. Please note the significant cross-talk between Vp and $1/Q$ caused by the restarting of the $1/Q$ model.

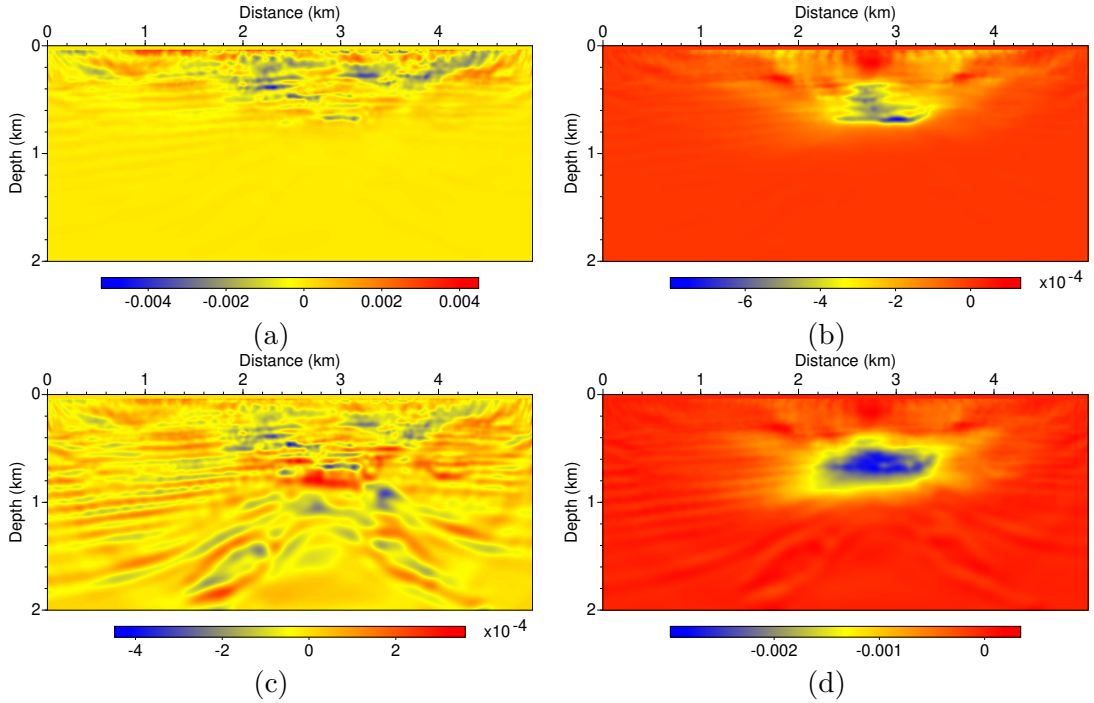


Figure 11: Comparison of the gradients of hierarchical inversion in the first iteration of stage 2: gradients of (a) Vp and (b) $1/Q$ and preconditioned gradients of (c) Vp and (d) $1/Q$. The preconditioned gradients exhibit strong agreement with the discrepancies between the true models and the starting models.

Valhall model

We consider a more realistic case where there is limited prior information on the Q model. The true and initial Valhall models, presented in Figure 12, represent the shallow water environment of an oil field in the north sea (Sirgue et al., 2010). The initial velocity model is obtained by applying a triangle smoother with a radius of 250 m to the true one, while the initial Q model is a two-layer model consisting of $Q = 1000$ in the water column and $Q = 200$ in the sediments. The true and initial density models are obtained from the corresponding velocity models, following Gardner's law $\rho = 309.6V^{0.25}$, and the density model remains unchanged throughout the inversion procedure. In this numerical study, we will employ the two-stage hierarchical inversion strategy based on the result of the previous experiment. During the first stage, we utilize the localized adaptive waveform inversion (LAWI) method (Yong et al., 2023b,a) to achieve a robust and efficient velocity reconstruction, effectively avoiding any potential cycle-skipping problems (Virieux and Operto, 2009). In the second stage, we will investigate the effect of different misfit measurements, namely waveform, ICF and FWA, on Q estimation. Prior to this investigation, we show the importance of long-offset acquisition and broadband data for robust Q inversion.

In our study, we use a fixed-spread acquisition, employing 32 equally spaced sources and 351 equally spaced receivers with a 25-meter interval on the surface. To simulate long-offset acquisition, we introduce two vertical lines, each comprising 69 receivers spaced at 25-meter intervals, positioned near the left and right model boundaries (▼ in Figure 12(a)). The top side of the model utilizes a free-surface boundary condition, while the other three sides employ an absorbing boundary condition. This combination is implemented to mitigate artificial reflections and ensure more accurate simulations of wave propagation. Figure 13

presents the seismogram of the first shot with clockwise rotation. Figure 14 illustrates the recordings of the 320th receiver with various settings of Q models and source wavelets. It is evident that the primary impact of viscosity on wave propagation is amplitude decay, which becomes more pronounced at higher frequencies, as observed from the spectra. Indeed, it is important to mention that in this shallow water environment, the spectra of recordings are significantly different from the spectrum of the wavelet, mainly due to the presence of strong ghosts (Aytun, 1999). These ghost reflections can introduce additional complexities and challenges in the analysis and interpretation of the recorded data (Provenzano et al., 2020).

The ℓ-BFGS method with $\ell = 5$ is employed to update the models, and each stage of the inversion uses 30 iterations. In Figure 15, the final reconstructed velocity models at the first stage are presented. Across four settings, which include or exclude the two vertical receiver lines and use 4 Hz or 8 Hz Ricker wavelets, the LAWI method consistently provides reliable velocity models of comparable quality. Figure 16 displays the final $1/Q$ models at the second stage using classical FWI, which minimizes the waveform difference under the L^2 norm. When utilizing reflection acquisition, only the shallow part of the Q model can be updated, while the middle area, corresponding to the gas cloud, is challenging to update effectively, as well as the backgroud. However, with the aid of two vertical receiver lines, the shallow part and the gas-cloud area get better updated, as well as the background. Figure 16 also illustrates that broadband data plays a crucial role in improving the quality of Q estimation. It leads to more accurate estimations in strongly attenuative areas and reduces artifacts at the boundary. This aligns with the findings in Keating and Innanen (2019) that the frequency-band used in frequency-domain FWI is a crucial factor in enhancing Q estimation.

Figure 17 displays the final $1/Q$ results of the ICF and FWA measurements using an 8 Hz dataset, including receivers from two vertical lines. Consistent with the analysis shown in Figure 3, the ICF measurement does not improve the Q estimation because the frequency band (0-20 Hz) does not cause a notable centroid frequency shift, and slight time shifts due to velocity errors significantly influence the misfit function. On the other hand, the FWA measurement effectively captures amplitude decay, which is the main influence of Q on wave propagation, providing a more reliable Q estimation for the shallow part and the middle gas cloud area. It is important to note that seismic amplitudes are easily contaminated by ambient noise, source and receiver coupling problems, source radiation patterns, and elastic effects.

In a more realistic scenario where the observed data contains strong low-frequency Gaussian noise, the L^2 norm of waveform difference remains robust to the noise (Bube and Langan, 1997). However, we observe that the robustness is lost when we try to minimize the difference of the frequency-weighted amplitude (FWA) measurement, the (weighted) absolute values of waveforms in the time-frequency domain. Figure 19 (a) shows the $1/Q$ result of the FWA measurement using noisy data (see Figure 18). Within the FWA measurement, the final result yields a large Q value due to the added noise, which increases the energy of the observed seismogram, leading to a less attenuative Q model. Additionally, it is important to note that in the presence of strong noise in the observed data, the positive-only FWA measurement can significantly degrade velocity inversion (Figure 20(a)). This emphasizes the challenges posed by noise and highlights the need for careful consideration when employing positive-only measurements in noisy environments.

Inspired by the double-difference concept, commonly used to construct differential measurements between stations to reduce systematic errors (Zhang and Thurber, 2003), we

apply a time differentiation to the FWA measurement to alleviate the common occurrence of low-frequency noise in realistic seismic data. The new misfit under the L^2 norm of the differential FWA measurement can be expressed as follows:

$$\mathbb{J}_{DFWA} = \frac{1}{2}\int_{\mathbb{R}} (D\mathcal{A}_u(t) - D\mathcal{A}_d(t))^2 \mathrm{d}t. \tag{26}$$

The adjoint source for FWA with differential methodology is written as:

$$r_{DFWA} = \mathcal{G}^{-1}\left[D^T(D\mathcal{A}_u(t) - D\mathcal{A}_d(t))|\omega|\frac{\hat{u}(\omega,t)}{A_u(\omega,t)}\right]. \tag{27}$$

Here, D and D^T denote first-order finite difference forward and backward in time, respectively. This relative FWA measurement provides an enhanced result of similar quality to the FWA measurement using clean data, as shown in Figure 19. From the adjoint sources (Figures 21-24), it is evident that strong noise in the observed data can significantly modify the adjoint source of the frequency-weighted amplitude (FWA) measurement: Noise tends to boost the contribution of weak-amplitude events in the predicted data. Thanks to the differential operation, the impact of low-frequency noise on the adjoint source can be greatly reduced, making the inversion robust to noise.

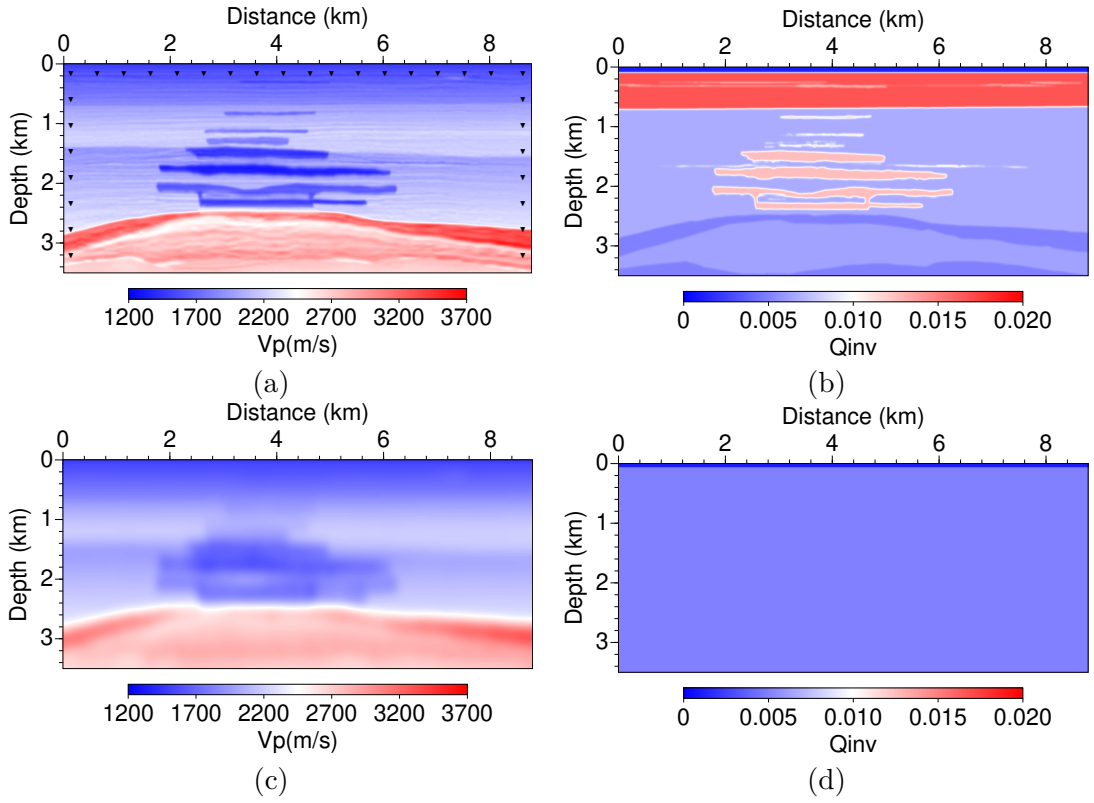


Figure 12: Valhall models: the first row presents the true (a) P-wave velocity and (b) $1/Q$ models. The second row presents the initial (a) P-wave velocity and (b) $1/Q$ models. The true and initial density models are derived from the corresponding velocity models according to Gardner's law, with the density model remaining unaltered during the inversion process.

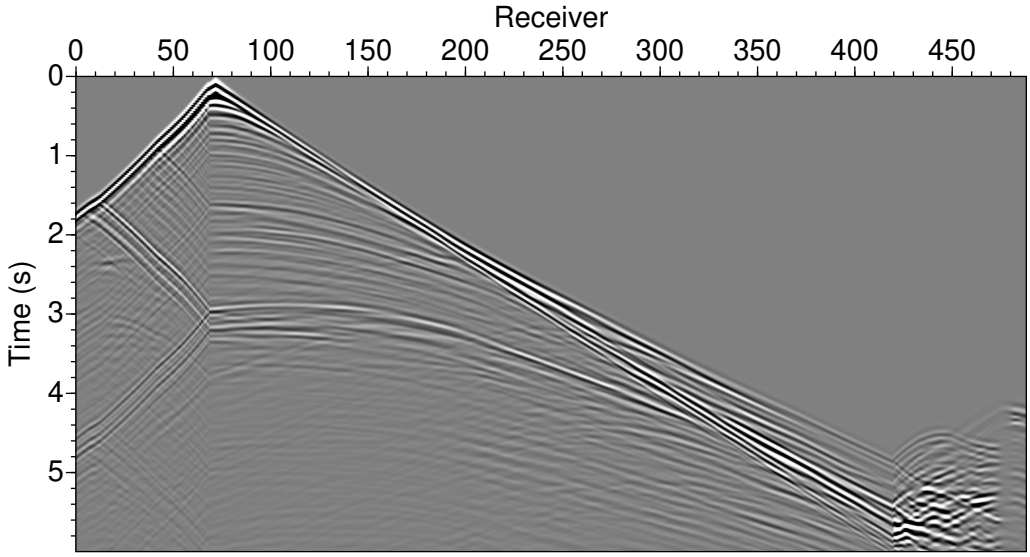


Figure 13: The recordings of the first shot with clockwise rotation: A fixed-spread acquisition is employed, consisting of 351 receivers evenly distributed at 25 m intervals on the surface. In addition, there are two vertical lines, each comprising 69 receivers, located near the left and right boundaries of the model.

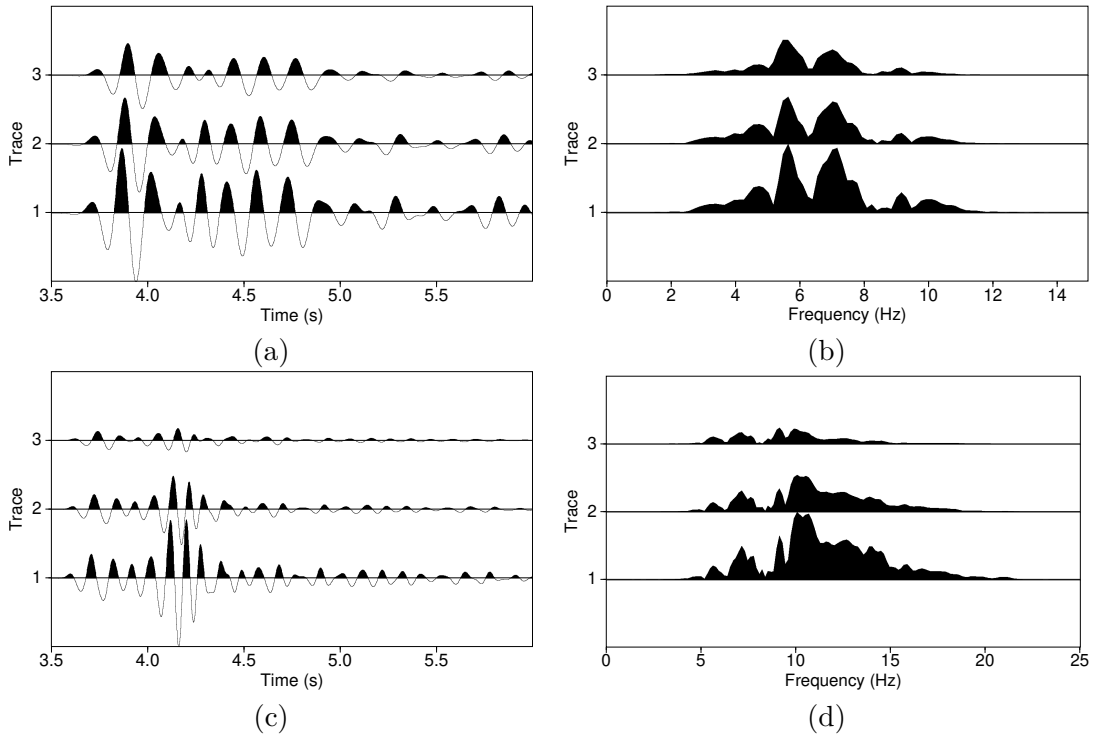


Figure 14: Data generated by a (a) 4Hz and (c) 8 Hz Ricker wavelet at the 320th receiver (left), along with its spectra (right) using different Q models: Q $=\infty$, Q = 200, true Q model, arranged from bottom to top.

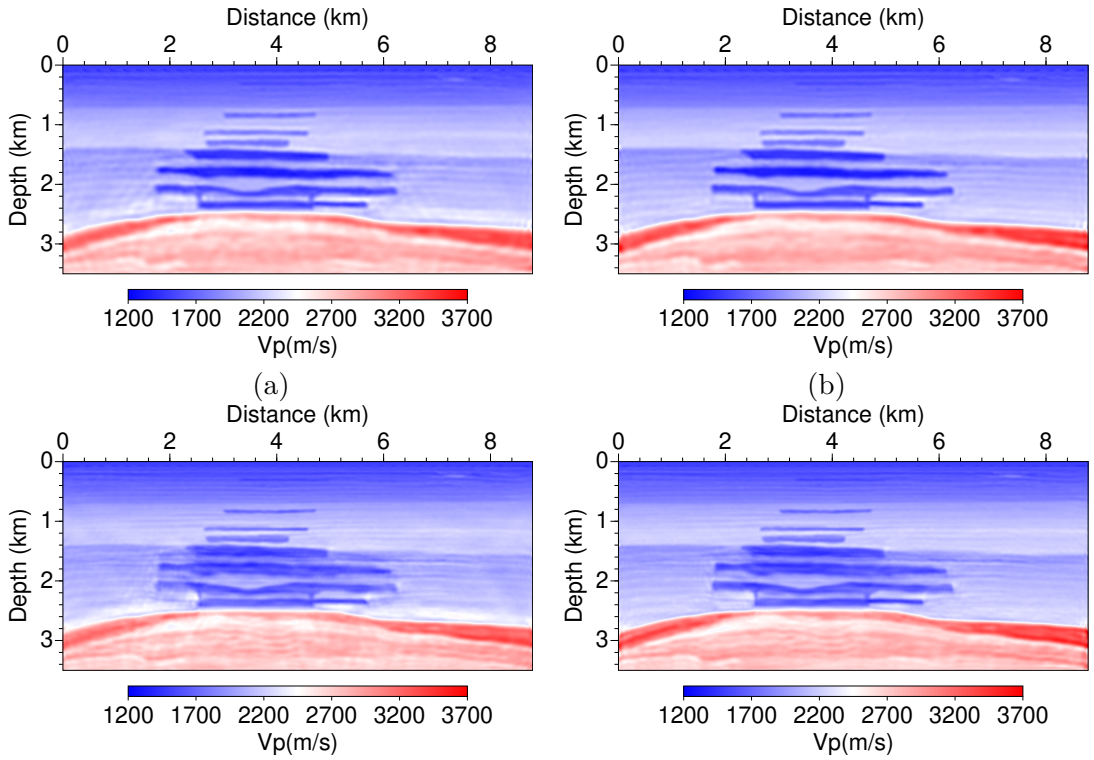


Figure 15: The final reconstructed velocity models at stage 1 with a hierarchical inversion strategy: The first row presents the results using a 4 Hz data set with (a) excluding and (b) including receivers from the two vertical lines. The second row displays the results using an 8 Hz data set with (c) excluding and (d) including receivers from the two vertical lines.

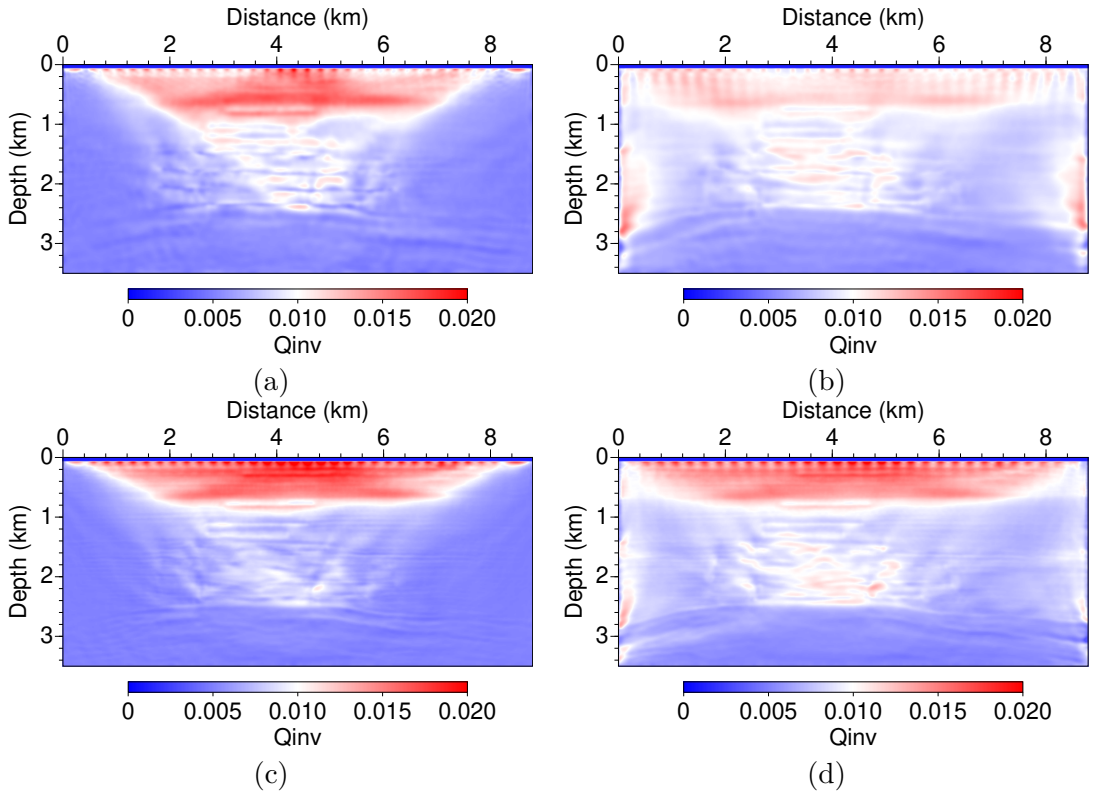


Figure 16: The final reconstructed $1/Q$ models at stage 2 with a hierarchical inversion strategy: The first row presents the results using a 4 Hz data set with (a) excluding and (b) including receivers from the two vertical lines. The second row displays the results using an 8 Hz data set with (c) excluding and (d) including receivers from the two vertical lines. Long-offset acquisition and broadband data significantly influence the quality of the reconstructed $1/Q$ model.

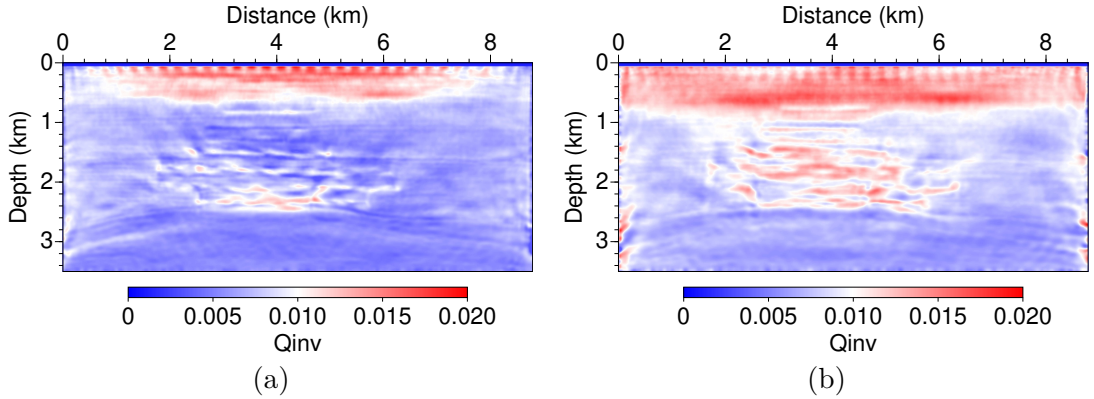


Figure 17: The final reconstructed $1/Q$ model with different misfit functions: (a) the instantaneous centroid frequency measurement and (b) the frequency-weighted amplitude measurement. The inversion process utilizes a data set generated by an 8 Hz Ricker wavelet, including receivers from the two vertical lines.

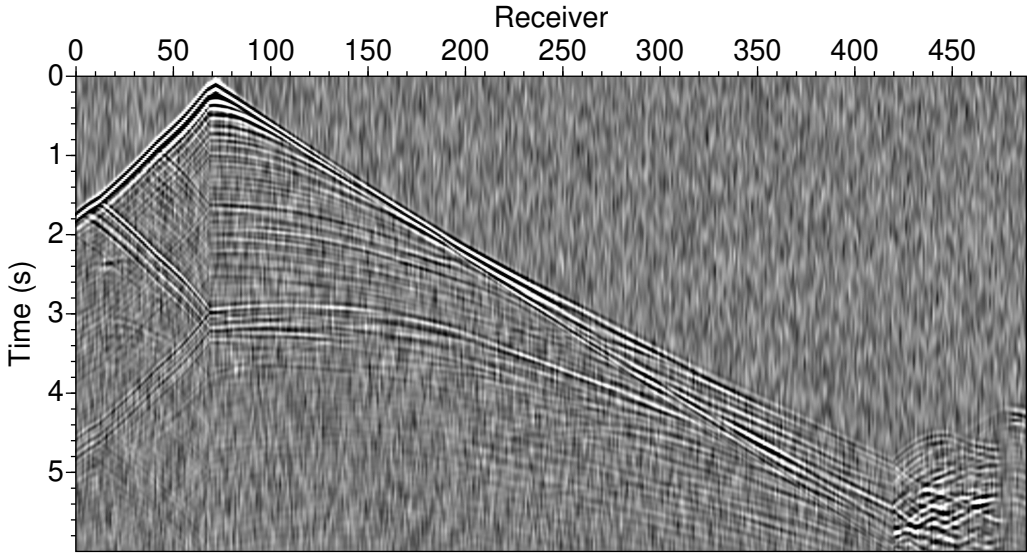
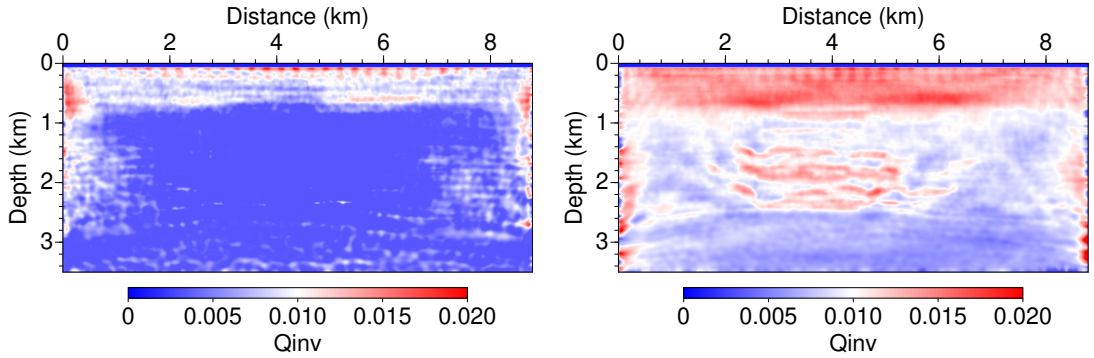
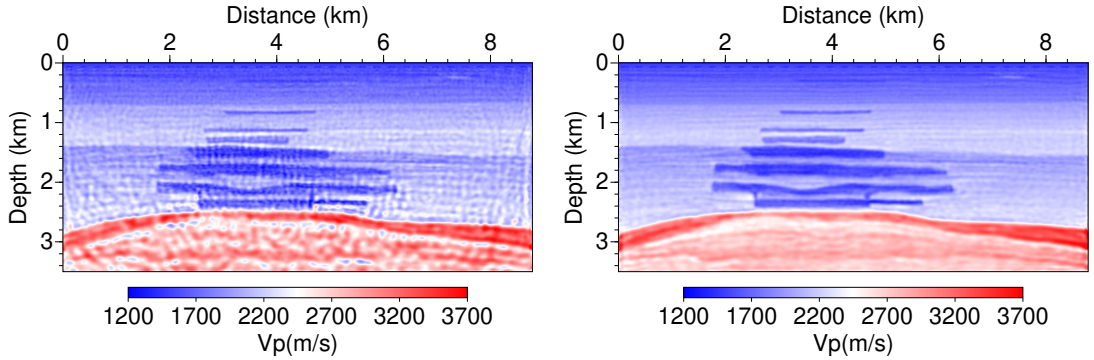


Figure 18: The recordings of the first shot with strong noise in the frequency band of 1-5 Hz.



(a) (b)

Figure 19: The final reconstructed $1/Q$ models with noisy data: (a) frequency-weighted amplitude measurement, (b) differential frequency-weighted amplitude measurement.



(a) (b)

Figure 20: The final reconstructed velocity model with noisy data: (a) frequency-weighted amplitude measurement, (b) differential frequency-weighted amplitude measurement.

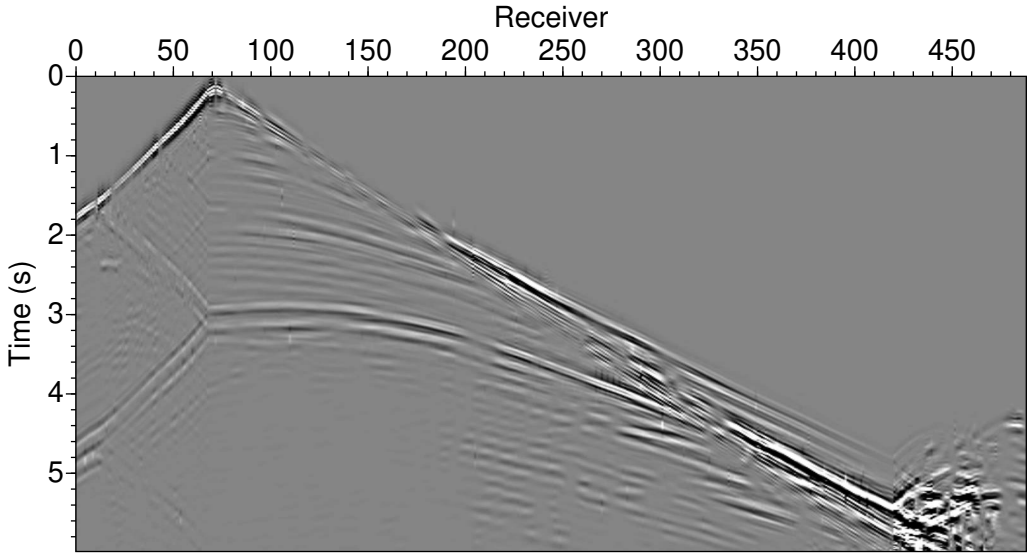


Figure 21: The adjoint source of the frequency-weighted amplitude measurement for noiseless observed data.

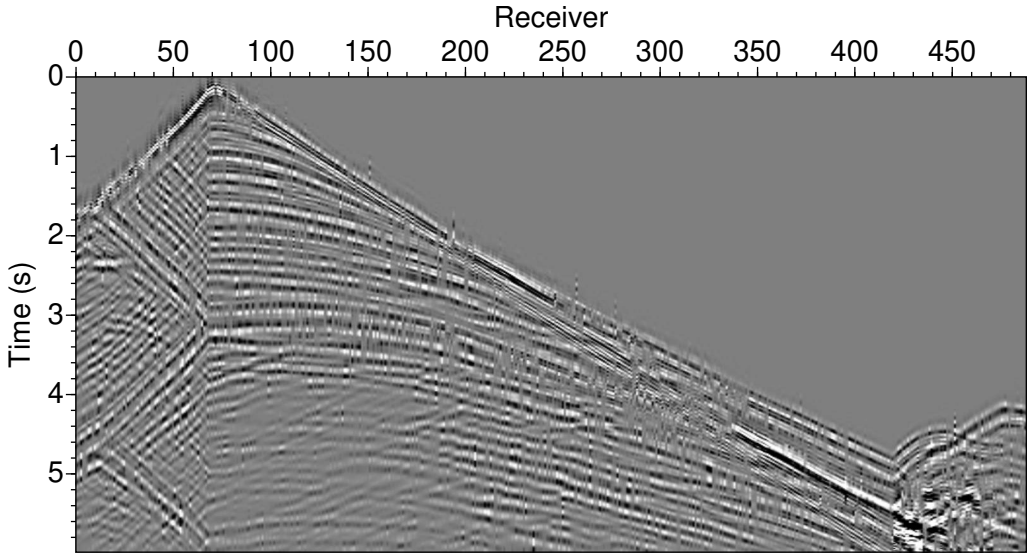


Figure 22: The adjoint source of the frequency-weighted amplitude measurement for noisy observed data.

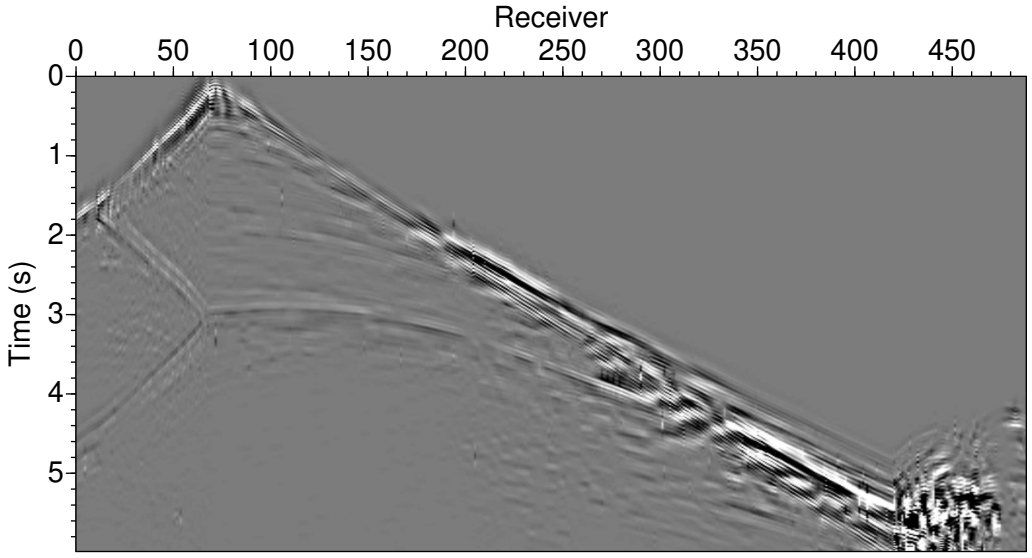


Figure 23: The adjoint source of the differential frequency-weighted amplitude measurement for noiseless observed data.

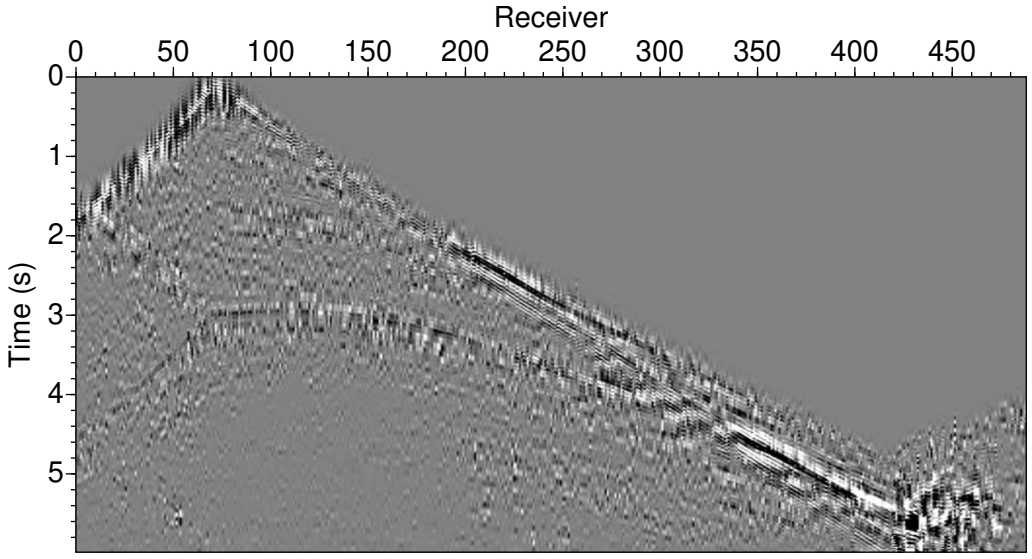


Figure 24: The adjoint source of the differential frequency-weighted amplitude measurement for noisy observed data.

DISCUSSION

The constant Q model is commonly used to characterize the propagation of seismic waves through viscous media. Figures 1 and 14 reveal that the primary effect of viscosity on seismic wave propagation is amplitude decay, which linearly increases with frequency. In comparison to the first-order seismic sensitivity parameter, velocity, the phase shift induced by viscosity is minor. To achieve robust Q estimation, broadband data and long-offset acquisition are essential, as the Q effect on wave propagation involves weak interactions and accumulates along the wave path. Besides, due to the absence of strong scattering at Q interfaces, the expected Q model resulting from waveform inversion is anticipated to have lower resolution compared to the velocity model obtained through waveform inversion.

The first-order seismic sensitivity parameter, velocity, can effectively address most data discrepancies. The two-stage workflow prioritizes resolving primary contradictions in the problem. Once the velocity model is well-resolved, the second-order seismic sensitivity parameter, like Q, can be integrated into the inversion process. Note that involving velocity in the second stage is crucial since building a perfect velocity model is nearly impossible. This inclusion can significantly reduce the risk of Q being over-updated due to data discrepancies caused by velocity variations. The developed preconditioner is capable of auto-tuning parameters of different classes, eliminating the need for special tuning of the scaling parameter.

Enhancing amplitude information in waveform inversion captures the primary Q effect, providing a more reliable Q estimation. Although the ICF measurement better represents peculiar features of Q, it requires a wide frequency band, making it susceptible to over-updating due to errors in kinematics. In contrast, the FWA measurement effectively mag-

nifies the Q effect within the frequency band commonly considered in waveform inversion, enabling a more robust Q estimation. However, transforming a signed signal to a positive-only measurement for inversion, based on minimizing the L2 norm error, compromises robustness to Gaussian noise. Nevertheless, the differentiation strategy can significantly mitigate low-frequency noise for the FWA measurement. Other methods that minimize least-squares differences between transformed positive measurements, such as envelope inversion (Wu et al., 2014) and normalized integration (Donno et al., 2013), may also suffer from noise. Our time-deviration approach to handle the noise issue may provide insights also for these measurements.

CONCLUSION

Achieving robust Q estimation through waveform inversion demands broadband data and long-offset acquisition due to weak interactions and accumulation along wave paths. The incorporation of two-stage sequential strategies, along with the developed preconditioner, proves to be effective in minimizing the velocity footprint on Q reconstruction. Additionally, enhancing amplitude information can improve Q estimation to some extent, but careful consideration is essential when handling frequency band and noise effects, especially when converting a signed signal to a positive-only measurement for inversion. In this context, the differential strategy stands out as a valuable technique for mitigating low-frequency noise and enhancing the inversion process. Numerical examples demonstrate that integrating these approaches enhances the accuracy and reliability of Q estimation in waveform inversion, underlining the potential for more precise Q characterization in realistic environments through further investigation.

ACKNOWLEDGMENTS

This study was partially funded by the SEISCOPE consortium (http://seiscope2.osug.fr), sponsored by AKERBP, CGG, DUG, EXXONMOBIL, GEOLINKS, JGI, PETROBRAS, SHELL, SINOPEC and TOTALENERGIES. This study was granted access to the HPC resources provided by the GRICAD infrastructure (https://gricad.univ-grenoble-alpes.fr), which is supported by Grenoble research communities, the HPC resources of Cray Marketing Partner Network (https://partners.cray.com), and those of IDRIS/TGCC under the allocation 046091 made by GENCI. The authors wish to thank Jean Virieux and Giuseppe Provenzano for insightful discussions.

APPENDIX A

PRECONITIONER FROM PSEUDO HESSIAN APPROXIMATION

An appropriate preconditioner is crucial to accelerate the convergence of iterations in FWI and correctly scale the gradients of different parameter classes. A preconditioner usually approximates the Hessian matrix. In practice, the diagonal of the Gauss-Newton Hessian matrice is extracted as the preconditioner (Pratt et al., 1998; Yang et al., 2018; Yong et al., 2022).

$$\begin{aligned} H_{GN}(\mathbf{m}) &= \int_0^T \mathrm{d}t \left(\frac{\partial w}{\partial \mathbf{m}}\right)^\dagger R^\dagger R \frac{\partial w}{\partial \mathbf{m}} \\ &= \int_0^T \mathrm{d}t \left(R\left(A(\mathbf{m})\right)^{-1} \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}} w\right)^\dagger \left(R\left(A(\mathbf{m})\right)^{-1} \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}} w\right) \end{aligned} \tag{A-1}$$

Using the Green's function $G(\mathbf{x}, \mathbf{x}_s, t)$ defined by $A(\mathbf{m})G(\mathbf{x}, \mathbf{x}_s, t) = \delta(\mathbf{x} - \mathbf{x_s}, t)$, we have

$$\begin{aligned} H_{GN}|_{i,j} = \int_0^T \int_{\mathbf{x}_r} \mathrm{d}t\mathrm{d}r \left(G(\mathbf{x}_r, \mathbf{x}_i, t) * \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(\mathbf{x}_i, t)\right) \\ \times \left(G(\mathbf{x}_r, \mathbf{x}_j, t) * \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_j} w(\mathbf{x}_j, t)\right), \end{aligned} \tag{A-2}$$

where $*$ is a temporal convolution operator. $G(\mathbf{x}_r, \mathbf{x}_i, t) * \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(\mathbf{x}_i, t)$ can be regarded as the wavefield at the receiver location $\mathbf{x}_r$ with a source term $\frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(\mathbf{x}_i, t)$ at the location of $\mathbf{x}_i$. Taking the advantage of spatial reciprocity of Green's function, the Gauss-Newton Hessian can be rewritten as

$$\begin{aligned} H_{GN}|_{i,j} = \int_{\mathbf{x}_r} \int_0^T \mathrm{d}r\mathrm{d}t \left(G(\mathbf{x}_i, \mathbf{x}_r, t) * \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(\mathbf{x}_i, t)\right) \\ \times \left(G(\mathbf{x}_j, \mathbf{x}_r, t) * \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_j} w(\mathbf{x}_i, t)\right). \end{aligned} \tag{A-3}$$

Now, the computation of $G(\mathbf{x}_i, \mathbf{x}_r, t) * \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(\mathbf{x}_i, t)$ involves convolving $\frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(\mathbf{x}_i, t)$ with the Green's function $G(\mathbf{x}_i, \mathbf{x}_r, t)$ originating from the receiver location $\mathbf{x}_r$ towards the location $\mathbf{x}_i$. However, equation (A-3) is not used to compute Gauss-Newton Hessian in practice because it is difficult to compute numerical Green's functions at each receiver in the time-domain FWI.

Under the assumption of zero-offset geometry, along with replacing $G(\mathbf{x}_i, \mathbf{x}_r, t)$ with $w(t)$, the computation of the diagonal of Gauss-Newton (GN) Hessian can be approximated by

$$H_{GN}(\mathbf{m})|_{i,i} \approx \int_0^T \mathrm{d}t \left(w(t) * \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(t) \right)^2 . \tag{A-4}$$

In the frequency-domain FWI, the convolution operator can be easily implemented (Choi and Shin, 2008). However, the convolution computation becomes challenging for the time-domain FWI because it is not feasible to store all time-dependent wavefields for large-scale 3D applications. Yang et al. (2018) used zero-lag cross-correlation instead of convolution to reduce computational cost

$$H_{GN}(\mathbf{m})|_{i,i} \approx \int_0^T \mathrm{d}t \left(w(t) \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(t) \right)^2 . \tag{A-5}$$

This further simplified preconditioner, to some extent, can be seen as a form of energy illumination for various parameters. For multi-parameter FWI, the values of the gradients of different parameter classes may be not of the same order of magnitude. From numerical experiments, we have observed that the preconditioner given by equation (A-5) may lead to over-compensation of unbalanced illumination for different parameter classes. Specifically, we have noticed that the values of the gradient for a parameter that is one order of magnitude smaller can become one order of magnitude larger after applying this preconditioner.

A meticulous tunable scaling strategy (Kamei and Pratt, 2013) is adopted to balance physical sensitivity of different parameters (Yang et al., 2018). In order to automatically align the values of preconditioned gradients for different parameter classes to the same (or similar) order of magnitude, we propose taking the square root of the approximated diagonal Hessian. Through numerical tests, we have discovered that this approach effectively aligns the preconditioned gradients, ensuring they are in the same (or similar) order of magnitude.

$$P_i = \sqrt{\int_0^T \mathrm{d}t \left(w(t) \frac{\partial A(\mathbf{m})}{\partial \mathbf{m}_i} w(t) \right)^2}. \tag{A-6}$$

It is clear that this preconditioner is positive definite. Note that we only approximate the diagonal of Gauss-Newton Hessian in this paper. One may consider the diagonal of off-diagonal block in Gauss-Newton Hessian to account for the trade-off between different parameters (Yang et al., 2018). However, this may make the preconditioner become indefinite from numerical experiments.

APPENDIX B

ADJOINT SOURCE IN THE TIME-FREQUENCY DOMAIN

We demonstrate the derivation of the adjoint source for misfits defined in the time-frequency domain. To illustrate the essential steps involved in this process, we will utilize the FWA measurement as an example. In order to obtain the adjoint source of the FWA misfit, let us introduce a significant identity

$$\frac{\partial |\hat{u}(t,\omega)|}{\partial \hat{u}(t,\omega)} = \frac{\hat{u}^\dagger(t,\omega)}{|\hat{u}(t,\omega)|}, \tag{B-1}$$

where † represents complex conjugation. This identity can be derived from the following equation:

$$\begin{aligned}\frac{\partial|\tilde{u}(t,\omega)|^2}{\partial\hat{u}(t,\omega)} &= 2|\hat{u}(t,\omega)|\frac{\partial|\hat{u}(t,\omega)|}{\partial\hat{u}(t,\omega)} \\ &= \frac{\partial\left[\hat{u}^\dagger(t,\omega)\hat{u}(t,\omega)\right]}{\partial\hat{u}(t,\omega)} \\ &= \frac{\partial\left[\hat{u}^\dagger(t,\omega)\hat{u}_1(t,\omega)\right]}{\partial\hat{u}_1(t,\omega)} + \frac{\partial\left[\hat{u}_2^\dagger(t,\omega)\hat{u}(t,\omega)\right]}{\partial\hat{u}_2(t,\omega)} \\ &= \hat{u}^\dagger(t,\omega) + \frac{\partial\left[\hat{u}_2^\dagger(t,\omega)\hat{u}(t,\omega)\right]^\dagger}{\partial\hat{u}_2(t,\omega)} \\ &= 2\hat{u}^\dagger(t,\omega). \end{aligned} \tag{B-2}$$

By combining the definition of the objective function and the Gabor transform pair, along with the aforementioned identity and the chain rule, we obtain the following expressions

$$\begin{aligned}\frac{\partial\mathbb{J}_{FWA}}{\partial u(\xi)} &= \frac{\partial\mathbb{J}_{FWA}}{\partial\mathcal{A}_u(t)} \times \frac{\partial\mathcal{A}_u(t)}{\partial A_u(t,\omega)} \times \frac{\partial A_u(t,\omega)}{\partial\hat{u}(t,\omega)} \times \frac{\partial\hat{u}(t,\omega)}{\partial u(\xi)} \\ &= \int_{\mathbb{R}} \mathcal{A}_u(t) - \mathcal{A}_d(t)\mathrm{d}t \times \int_{\mathbb{R}} |\omega|\mathrm{d}\omega \times \frac{\hat{u}^\dagger(t,\omega)}{|\hat{u}(t,\omega)|} \times \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} h_\sigma^\dagger(t-\xi)e^{-i\omega\xi}\mathrm{d}\xi \\ &= \left\{\int_{\mathbb{R}^2} (\mathcal{A}_u(t) - \mathcal{A}_d(t))|\omega|\frac{\hat{u}(\omega,t)}{A_u(\omega,t)}\mathrm{d}t\mathrm{d}\omega \times \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} h_\sigma(t-\xi)e^{i\omega\xi}\mathrm{d}\xi\right\}^\dagger \\ &= \left\{\int_{\mathbb{R}} \mathcal{G}^{-1}\left[(\mathcal{A}_u(t) - \mathcal{A}_d(t))|\omega|\frac{\hat{u}(\omega,t)}{A_u(\omega,t)}\right]\mathrm{d}\xi\right\}^\dagger \end{aligned} \tag{B-3}$$

Since the adjoint source is a real-valued signal, we can write it as follows

$$r_{FWA} = \mathcal{G}^{-1}\left[(\mathcal{A}_u(t) - \mathcal{A}_d(t))|\omega|\frac{\hat{u}(\omega,t)}{A_u(\omega,t)}\right]. \tag{B-4}$$

Following a similar procedure, the adjoint source of the other misfits defined in the time-frequency domain can be readily obtained.

REFERENCES

Alkhalifah, T. and R. Plessix, 2014, A recipe for practical full-waveform inversion in anisotropic media: An analytical parameter resolution study: Geophysics, 79, R91–R101.

Aytun, K., 1999, The footsteps of the receiver ghost in the fk domain: Geophysics, 64, 1618–1626.

Bube, K. P. and R. T. Langan, 1997, Hybrid l_1/l_2 minimization with applications to tomography: Geophysics, 62, 1183–1195.

Carcione, J. M., 2015, Wave fields in real media, wave propagation in anisotropic, anelastic, porous and electromagnetic media: Elsevier, third edition edition.

Choi, Y. and C. Shin, 2008, Frequency-Domain Elastic Full Waveform Inversion Using the New Pseudo-Hessian Matrix: Experience Of Elastic Marmousi 2 Synthetic Data: Bulletin of the Seismological Society of America, 98, 2402–2415.

da Silva, N. V., G. Yao, and M. Warner, 2019a, Semiglobal viscoacoustic full-waveform inversionsemiglobal viscoacoustic inversion: Geophysics, 84, R271–R293.

———, 2019b, Wave modeling in viscoacoustic media with transverse isotropy: Geophysics, 84, C41–C56.

Donno, D., H. Chauris, and H. Calandra, 2013, Estimating the background velocity model with the normalized integration method: EAGE Technical Program Expanded Abstracts 2013, Tu0704.

Dutta, G. and G. T. Schuster, 2016, Wave-equation q tomography: Geophysics, 81, R471–R484.

Fichtner, A., B. L. N. Kennett, H. Igel, and H. P. Bunge, 2008, Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain: Geophysical Journal International, 175, 665–685.

Gabor, D., 1946, Theory of communication. part 1: The analysis of information: Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering, 93, 429–441.

Gao, L., Y. Pan, A. Rieder, and T. Bohlen, 2021, Multiparameter viscoelastic full-waveform inversion of shallow-seismic surface waves with a pre-conditioned truncated newton method: Geophysical Journal International, 227, 2044–2057.

Hak, B. and W. A. Mulder, 2011, Seismic attenuation imaging with causality: Geophysical Journal International, 184, 439–451.

He, W. and R. Plessix, 2017, Analysis of different parameterisations of waveform inversion of compressional body waves in an elastic transverse isotropic earth with a vertical axis of symmetry: Geophysical Prospecting, 65, 1004–1024.

Jiang, H., 2019, Seismic imaging: strategies for visco-acoustic full waveform inversion: PhD thesis, Université Paris sciences et lettres.

Kamath, N., R. Brossier, L. Métivier, A. Pladys, and P. Yang, 2021, Multiparameter full-waveform inversion of 3D ocean-bottom cable data from the Valhall field: Geophysics, 86, B15–B35.

Kamath, N. and I. Tsvankin, 2016, Elastic full-waveform inversion for vti media: Methodology and sensitivity analysis: Geophysics, 81(2), C53–C68.

Kamei, R. and R. G. Pratt, 2013, Inversion strategies for visco-acoustic waveform inversion: Geophysical Journal International, 194, 859–894.

Keating, S. and K. A. Innanen, 2019, Parameter crosstalk and modeling errors in viscoacoustic seismic full-waveform inversion: Geophysics, 84, R641–R653.

Malinowski, M., S. Operto, and A. Ribodetti, 2011, High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full wave-

form inversion: Geophysical Journal International, 186, 1179–1204.

Métivier, L. and R. Brossier, 2016, The seiscope optimization toolbox: A large-scale nonlinear optimization library based on reverse communication: Geophysics, 81, F11–F25.

Métivier, L., R. Brossier, S. Operto, and J. Virieux, 2015, Acoustic multi-parameter FWI for the reconstruction of P-wave velocity, density and attenuation: preconditioned truncated Newton approach: 85^th Annual Meeting-New Orleans, Expanded Abstracts, 1198–1203, SEG.

Moczo, P. and J. Kristek, 2005, On the rheological models used for time-domain methods of seismic wave propagation: Geophysical Research Letters, 32.

Mulder, W. A. and B. Hak, 2009, An ambiguity in attenuation scattering imaging: Geophysical Journal International, 178, 1614–1624.

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

Operto, S. and A. Miniussi, 2018, On the role of density and attenuation in 3D multiparameter visco-acoustic VTI frequency-domain FWI: an OBC case study from the North Sea: Geophysical Journal International, 213, 2037–2059.

Pan, W., K. A. Innanen, and W. Liao, 2017, Accelerating hessian-free gauss-newton full-waveform inversion via l-bfgs preconditioned conjugate-gradient algorithm: Geophysics, 82, R49–R64.

Pan, W., K. A. Innanen, and Y. Wang, 2023a, Adjoint q tomography with central-frequency measurements in viscoelastic medium: Geophysical Journal International, 233, 1144–1165.

Pan, W., L. Qu, K. A. Innanen, J. Dettmer, M. Macquet, D. Lawton, and Y. Wang, 2023b,

Imaging near-surface s-wave velocity and attenuation models by full-waveform inversion with distributed acoustic sensing-recorded surface waves: Geophysics, 88, R65–R78.

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

Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, 133, 341–362.

Prieux, V., R. Brossier, S. Operto, and J. Virieux, 2013, Multiparameter full waveform inversion of multicomponent OBC data from Valhall. Part 1: imaging compressional wavespeed, density and attenuation: Geophysical Journal International, 194, 1640–1664.

Provenzano, G., T. J. Henstock, J. M. Bull, and G. Bayrakci, 2020, Attenuation of receiver ghosts in variable-depth streamer high-resolution seismic reflection data: Marine Geophysical Research, 41, 1–15.

Quan, Y. and J. M. Harris, 1997, Seismic attenuation tomography using the frequency shift method: Geophysics, 62, 895–905.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: the next leap forward in imaging at Valhall: First Break, 28, 65–70.

Strang, G. and T. Nguyen, 1996, Wavelets and filter banks: SIAM.

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

Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An introduction to Full Waveform Inversion, in Grechka, V. and K. Wapenaar, eds., Ency-

clopedia of Exploration Geophysics, R1–1–R1–40. Society of Exploration Geophysics.

Virieux, J. and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, 74, WCC1–WCC26.

Wang, Y., 2009, Seismic inverse Q filtering: John Wiley & Sons.

Wu, F. V., D. Borisov, F. J. Simons, and P. Williamson, 2021, Waveform inversion for shear velocity and attenuation via the spectral-element adjoint method: First International Meeting for Applied Geoscience & Energy, 697–701.

Wu, R.-S., J. Luo, and B. Wu, 2014, Seismic envelope inversion and modulation signal model: Geophysics, 79, WA13–WA24.

Xie, Y., C. A. Rychert, and N. Harmon, 2023, Elastic and anelastic adjoint tomography with and full hessian kernels: Geophysical Journal International, 234, 1205–1235.

Xing, G., 2022, Deciphering seismic attenuation: fractal description, wavefield modeling and q full waveform inversion: PhD thesis, Pennsylvania State University.

Yang, J., H. Zhu, X. Li, L. Ren, and S. Zhang, 2020, Estimating p wave velocity and attenuation structures using full waveform inversion based on a time domain complex-valued viscoacoustic wave equation: The method: Journal of Geophysical Research: Solid Earth, 125, e2019JB019129.

Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016, A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium: Geophysical Journal International, 207, 129–149.

Yang, P., R. Brossier, L. Métivier, J. Virieux, and W. Zhou, 2018, A Time-Domain Preconditioned Truncated Newton Approach to Multiparameter Visco-acoustic Full Waveform Inversion: SIAM Journal on Scientific Computing, 40, B1101–B1130.

Yong, P., R. Brossier, and L. Métivier, 2022, Parsimonious truncated newton method for

time-domain full-waveform inversion based on the fourier-domain full-scattered-field approximation: Geophysics, 87, R123–R146.

Yong, P., R. Brossier, L. Métivier, and J. Virieux, 2021, Q estimation by full-waveform inversion: analysis and misfit functions comparison: First International Meeting for Applied Geoscience & Energy, 827–831.

———, 2023a, Localized adaptive waveform inversion: theory and numerical verification: Geophysical Journal International, 233, 1055–1080.

———, 2023b, Localized adaptive waveform inversion: regularizations for gabor deconvolution and 3d field data application: Geophysical Journal International, 235, 448–467.

Zhang, H. and C. H. Thurber, 2003, Double difference tomography: The method and its application to the Hayward fault, California: Bulletin of the Seismological Society of America, 93, 1875–1889.

Zhu, T., J. Harris, and B. Biondi, 2014, Q-compensated reverse-time migration: Geophysics, 79, S77–S87.