Geophysical Journal International

Geophys. J. Int. (2023) **234,** 1205–1235
Advance Access publication 2023 March 17
GJI Seismology

https://doi.org/10.1093/gji/ggad114

Elastic and anelastic adjoint tomography with and full Hessian kernels

Yujiang Xie [1], Catherine A. Rychert [1,2] and Nicholas Harmon [1,2]

[1] *Ocean and Earth Science, University of Southampton, UK. E-mail: Yujiang.Xie@soton.ac.uk*
[2] *Geology and Geophysics, Woods Hole Oceanographic Institution, USA*

Accepted 2023 March 11. Received 2023 January 6; in original form 2022 May 26

SUMMARY

The elastic and anelastic structures of the Earth offer fundamental constraints for understanding its physical and chemical properties. Deciphering small variations in the velocity and amplitude of seismic waves can be challenging. Advanced approaches such as full-waveform inversion (FWI) can be useful. We rewrite the anelastic Fréchet kernel expression of Fichtner & van Driel using the displacement–stress formulation. We then derive the full Hessian kernel expression for viscoelastic properties. In these formulations, the anelastic Fréchet kernels are computed by the forward strain and a shift of the adjoint strain. This is complementary to the quality factor Q (i.e., inverse attenuation) Fréchet kernel expressions of Fichtner & van Driel that are explicit for the velocity–stress formulation. To reduce disk space and I/O requirements for computing the full Hessian kernels, the elastic full Hessian kernels are computed on the fly, while the full Hessian kernels for Q are computed by a combination of the on-the-fly approach with the parsimonious storage method. Applications of the Fréchet and full Hessian kernels for adjoint tomography are presented for two synthetic 2-D models, including an idealized model with rectangular anomalies and a model that approximates a subduction zone, and one synthetic 3-D model with an idealized geometry. The calculation of the full Hessian kernel approximately doubles the computationally cost per iteration of the inversion; however, the reduced number of iterations and fewer frequency stages required to achieve the same level of convergence make it overall computationally less expensive than the classical Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) FWI for the 2-D elastic tested models. We find that the use of full Hessian kernels provides comparable results to the L-BFGS inversion using the improved anelastic Fréchet kernels for the 2-D anelastic models tested for the frequency stage up to 0.5 Hz. Given the computational expense of the Q full Hessian kernel calculation, it is not advantageous to use it in Q inversions at this time until further improvements are made. For the 3-D elastic inversion of the tested model, the full Hessian kernel provides similar image quality to the L-BFGS inversion for the frequency stage up to 0.1 Hz. We observe an improved convergence rate for the full Hessian kernel inversion in comparison to L-BFGS at a higher frequency stage, 0.1–0.2 Hz, and we speculate that at higher frequency stages the use of full Hessian kernels may be more computationally advantageous than the classical L-BFGS for the tested models. Finally, we perform 3-D elastic and Q L-BFGS inversions simultaneously using the rederived Q kernels, which can reduce the computational cost of the inversion by about 1/3 when compared to the classical anelastic adjoint tomography using the additionally defined adjoint source. The recovered Q model is smeared when compared to the recovered elastic model at the investigation frequencies up to 0.5 Hz. Q inversion remains challenging and requires further work. The 2-D and 3-D full Hessian kernels may be used for other purposes for instance resolution analysis in addition to the inversions.

Key words: Inverse theory; Computational seismology; Seismic tomography; Theoretical seismology; Wave propagation; Wave scattering and diffraction.

1 INTRODUCTION

1.1 The anelastic Earth and full-waveform inversion

Imaging of the seismic structure of the Earth has improved continuously over the past four decades with rapid advances in high-performance computing, fast expansions in global and regional seismic networks, and continuous developments in waveform inversion approaches (e.g. Bamberger *et al.* 1977, 1982; Lailly 1983; Tarantola 1984; Gauthier *et al.* 1986; Tarantola 1988; Igel *et al.* 1996; Pratt *et al.* 1998; Tape *et al.* 2009; Fichtner *et al.* 2009; Virieux & Operto 2009; Liu & Gu 2012; Zhu *et al.* 2012; French & Romanowicz 2015; Bozdağ *et al.* 2016; Tromp 2020; Harmon *et al.* 2020; Rychert *et al.* 2021; Agius *et al.* 2021). Although regional arrays provide high resolution imaging, full waveform approaches can further increase resolution, imaging structure smaller than a seismic wavelength (e.g. Chen *et al.* 2007a; Tape *et al.* 2009; Fichtner *et al.* 2009; Zhu *et al.* 2012; Zhu & Tromp 2013; Fichtner & Villaseñor 2015; Gao 2018).

The elastic and anelastic seismic structures of the Earth are important and fundamental constraints. The two are invaluable in determining the properties of the Earth, although anelasticity is more rarely investigated. At a global scale in the mantle, the observed attenuation structures from surface waves highlight regions of relatively high and low temperatures, with low Q observed beneath most of the Earth's ridges and rifts and high Q beneath the ancient stable continental interiors (Dalton *et al.* 2008). Regional and global surface wave studies also distinguish low attenuation lithosphere over high attenuation asthenosphere beneath the oceans (Yang *et al.* 2007; Dalton *et al.* 2008; Ruan *et al.* 2018; Saikia *et al.* 2021). Q anomaly observations in subduction zone mantle wedges have been used to infer the locations of water and partial melt (Myers *et al.* 1998; Takanami *et al.* 2000; Tsumura *et al.* 2000; Schurr *et al.* 2003; Stachnik *et al.* 2004; Pozgay *et al.* 2009; Ko *et al.* 2012; Wei *et al.* 2015; Eberhart-Phillips *et al.* 2018). Finally, shallow Q may be used to estimate seismic hazard (Cornell 1968).

Anelasticity is particularly important because it provides a complementary constraint to typical velocity tomography. Laboratory experiments predict unique relationships between attenuation and velocity depending on the physical and chemical properties of the Earth (Jackson & Faul 2010; McCarthy *et al.* 2011; Yamauchi & Takei 2016; Havlin *et al.* 2021). Joint consideration of velocity and attenuation observations in subduction zone settings can more tightly constrain the properties of the Earth including the locations and pathways of water and melt through the mantle wedge (Rychert *et al.* 2008; Syracuse *et al.* 2008; Abers *et al.* 2014; Wei *et al.* 2015; Wei & Wiens 2018). In addition, experimental constraints suggest that the frequency dependence of attenuation may be different in the lithosphere in comparison to the asthenosphere, possibly because of different physical properties such as the presence of partial melt (Faul & Jackson 2015; Jackson 2015). This phenomenon has been observed via comparison of oceanic attenuation studies that used waveforms at different periods and interpreted in terms of either partial melt and/or pre-melt conditions (Yamauchi & Takei 2016; Takeuchi *et al.* 2020). Therefore, attenuation constraints play a key role in determining the factors that control subduction zone dynamics and associated hazards and also understanding the physical and chemical properties that define the plate, a fundamental question for plate tectonic theory (Rychert *et al.* 2005, 2007, 2010, 2012, 2014, 2018a, 2018b, 2018c, 2020, 2021; Rychert & Harmon 2017; Rychert and Shearer 2009, 2011; Fischer *et al.* 2010, 2020; Tharimena *et al.* 2016, 2017a, 2017b; Chichester *et al.* 2018; Lavayssiere *et al.* 2018; Harmon *et al.* 2018, 2020, 2021a, 2021b; Wang *et al.* 2020; Possee *et al.* 2021; Saikia *et al.* 2021).

Full-waveform inversion (FWI) has the potential to improve constraints on the velocity and attenuation structures of the Earth given that it uses less approximations to the wave simulation than typical seismic techniques (e.g. Virieux & Operto 2009; Fichtner 2010; Tromp 2020). FWI uses sensitivity kernels which indicate the sensitivity of seismograms to model parameters. Typically, Fréchet kernels, the first order derivatives of the seismological data functionals (e.g. Dahlen *et al.* 2000; Tromp *et al.* 2005) are used. Full Hessian kernels, the second order partial derivatives of the data functionals applied to an arbitrary model perturbation (e.g. Fichtner & Trampert 2011), have the potential to increase convergence rate and/or mitigate inter-parameter trade-offs, but full Hessian kernels are not frequently used owing to their higher computational cost and the large disk space and I/O requirements to save and transmit the required wavefields (see e.g. Xie *et al.* 2021). Specifically, computation of Fréchet derivatives can be accomplished with two simulations per event, in which the forward fields may be recomputed on the fly (Liu & Tromp 2006) or adaptively stored or compressed and stored to the disk space (Fichtner *et al.* 2009; Boehm *et al.* 2016). In contrast, the full Hessian kernel requires four simulations (Fichtner & Trampert 2011; Xie *et al.* 2021). Full Hessian kernel calculations using the classical storage approach require large amounts of disk space to store multiple wavefields with increased associated I/O expense. The approach generally saves each kth ($k \geq 1$) time step of the required wavefields during the forward simulations and then reads each kth time step of the stored wavefields for the kernel constructions. To reduce the disk space and I/O requirements, Boehm & Ulbrich (2015) applied the lossy compression methods of Boehm *et al.* (2016) to the required wavefield for the elastic Hessian kernel construction, resulting in a trade-off between the accuracy of the Hessian kernels and the compression factors. The lossy compression and decompression approach is promising for the Fréchet kernel calculation. For the full Hessian kernel calculation, errors may be propagated into the two small perturbed wavefields because the decompressed wavefields also appear as a distributed source term in the two auxiliary wave equations (Boehm & Ulbrich 2015). To reduce computational expense, Epanomeritakis *et al.* (2008) computed approximate elastic Hessian kernels with lower storage requirements, using checkpointing to recompute the forward fields during the adjoint simulation. The checkpointing was used to compute each part of the approximate Hessian kernels. Pakravan *et al.* (2016) computed the Gauss–Newton Hessian vector product for material profile reconstruction in viscoelastic semi-infinite solid media. Luo *et al.* (2014) rewrote the Hessian kernel formulas and then used the diagonal terms of the Hessian to construct four pre-conditioners for FWI and resolution analysis. The finite-difference (FD) approximation of the elastic Hessian kernel is sometimes computed for resolution analysis in FWIs (e.g. Zhu *et al.* 2012, 2015; Bozdağ *et al.* 2016; Tao *et al.* 2018; Gao *et al.* 2021), which is usually constructed by two gradients under the linear assumption: $\mathbf{H}\Delta\mathbf{m} \approx \mathbf{g}(\mathbf{m} + \Delta\mathbf{m}) - \mathbf{g}(\mathbf{m})$, where the

gradient **g** in the classical sense is given by the projection of the sensitivity Fréchet kernel onto the model basis functions. More recently, the storage and I/O requirements of calculation of the full Hessian kernels were reduced by the introduction of an on-the-fly approach, otherwise known as QuadSEM (Xie *et al.* 2021). This approach is accurate since the forward and adjoint fields are computed exactly and used on the fly with two additional forward simulations (Xie *et al.* 2021).

Hessian kernels may be advantageous for determining anelastic properties as well, although this also doubles the storage and I/O requirements in comparison to the use of Fréchet kernels alone. Previous work has developed viscoacoustic Hessian kernels for use in active source applications, where *S* waves are not often measured or used. Yang *et al.* (2018) computed the viscoacoustic Hessian kernels in the time domain via checkpointing implementation (Yang *et al.* 2016a, b) for truncated Newton FWIs, achieving promising results in comparison to the classical L-BFGS FWIs. Yong *et al.* (2020) reduced the computational effort and storage requirement for constructing the viscoacoustic Hessian kernels by compressing the wavefields approximately in the Fourier domain. As previously discussed, wavefield compression methods may involve trade-offs between accuracy and the level of compression or approximation (Boehm & Ulbrich 2015; Boehm *et al.* 2016). Xing & Zhu (2020) computed the viscoacoustic Hessian kernels via a fractional viscoacoustic wave equation (Xing & Zhu 2019). In earthquake seismology, fully elastic and anelastic waveforms and consideration of all components of the wavefield including *P* waves, *S* waves, surface waves, etc. are required and are the focus of the work here. Although using full Hessian kernels theoretically improve resolution and convergence rates over approaches using Fréchet kernels, the improvement achieved by the on-the-fly approach has yet to be demonstrated or quantified.

Given the irreversible nature of the anelastic wave equation and the large disk space required to save the set of forward fields, Komatitsch *et al.* (2016) developed the parsimonious storage method, also known as checkpointing, for anelastic Fréchet kernel calculations. The method saves the forward fields at some discrete points, known as checkpoints, and then uses the fields at these checkpoints to recompute the necessary time steps of the forward field for the kernel calculation during the adjoint simulation. The method reduces the storage requirement by down-sampling the number of time steps. It is numerically accurate using the exactly computed fields, but there is a trade-off with computational costs, which are increased by recalculating wavefields in between the checkpoints. Because of its accuracy, we extend the parsimonious storage method of Komatitsch *et al.* (2016) to calculate the full Hessian kernels for viscoelastic media, which we refer to as the QuadSEM-Q. The QuadSEM-Q is complementary to the on-the-fly QuadSEM approach (Xie *et al.* 2021). The QuadSEM-Q simultaneously computes two forward fields and saves them at the checkpoints during the forward simulation, then recomputes the two forward fields while computing two adjoint fields during the adjoint simulation.

In the following, we demonstrate the approximation between the frequency-independent *Q* Fréchet kernels of Fichtner & van Driel (2014) and the frequency-independent *Q* Fréchet kernels of Tromp *et al.* (2005) (Section 2.1). The frequency-independent *Q* Fréchet kernels of Fichtner & van Driel (2014) are used with a slight extension for displacement–stress formulation and are computed and implemented based on the public SPECFEM2D package (Tromp *et al*. 2008). In Section 2.2, we derive the full Hessian kernels for frequency-(in)dependent *Q* based on the work of Fichtner & Trampert (2011) and Fichtner & van Driel (2014). For the full Hessian kernel calculations for viscoelastic properties, we combine the on-the-fly approach (Xie *et al.* 2021) with the parsimonious storage method (Komatitsch *et al.* 2016) to reduce the disk space and I/O requirements. We apply the kernels using the adjoint tomography workflow (e.g. Zhu *et al.* 2015, Section 3) for the elastic and viscoelastic inversion. We consider isotropic elastic and anelastic parameters and their trade offs here, but we neglect anisotropy. We use two 2-D models for the demonstration of the elastic and anelastic tests, one with an idealized source–receiver geometry example with a few box-shaped anomalies, and the other with a realistic source–receiver geometry example with the structure of the Nicaragua subduction model (Section 4). We then demonstrate the Hessian kernel recovery of elastic and anelastic structure using the idealized 2-D synthetic model and the more realistic Nicaragua subduction model.

Two 3-D inversions are considered in this work. The 3-D model is an extension of the 2-D idealized block model. One is the 3-D elastic FWI using the full Hessian kernels on the fly. The 3-D elastic inversion is compared with the classical elastic FWI that uses the Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) algorithm. The purpose of this comparison is not to compare the L-BFGS and the Newton-CG algorithms but to see the performance of full Hessian kernel for FWI while using the L-BFGS FWI as a reference. The second 3-D inversion is a simultaneous recovery of the 3-D elastic and attenuated properties with the *Q* Fréchet kernels computed by checkpointing using the improved *Q* Fréchet kernels rederived in this work based on the L-BFGS algorithm. The 3-D *Q* inversion that uses the *Q* full Hessian kernels have not been performed because (1) It is computationally prohibitive at present and (2) It has comparable performance in comparison to the L-BFGS inversion as the 2-D case for the frequency up to 0.5 Hz. In this work, the 3-D elastic full Hessian kernels are computed on the fly with QuadSEM3D (https://github.com/yujiangxie/QuadSEM3D).The QuadSEM3D is an extension of the QuadSEM (Xie *et al.* 2021) based on the open-source SPECFEM3D Cartesian package (Komatitsch *et al.* 1999; Peter *et al.* 2011).

2 FRÉCHET AND FULL HESSIAN KERNELS FOR VISCOELASTIC MEDIA

Full Hessian kernels for viscoelastic media may be computed in a similar way as the Fréchet kernels for viscoelastic media (Tromp *et al.* 2005; Fichtner & van Driel 2014). One approach the calculation relies on defining an additional adjoint source for the Q Fréchet kernel calculations (Bozdağ *et al.* 2011; Zhu *et al.* 2015). The other approach uses the forward strain and the adjoint memory variable (Fichtner & van Driel 2014). The latter is more computationally efficient since the adjoint memory variable is a by-product of the adjoint simulation for the viscoelastic properties. The *Q* Fréchet kernels can be computed while computing the Fréchet kernels for elastic properties. The full

Hessian kernels computed for elastic properties are constructed by the expressions of Appendix A. In this section, starting from the link between the work of Tromp *et al.* (2005) and Fichtner & van Driel (2014) for the Q Fréchet kernel calculations, we derive the full Hessian kernels for Q_μ and Q_κ, based the method of Fichtner & Trampert (2011) and the method of Fichtner & van Driel (2014).

In the adjoint method for viscoelastic media, the variation in $\delta\chi$ of the measurement functional may be defined as (e.g. Fichtner & van Driel 2014)

$$\delta\chi = -\int_{-\infty}^{\infty}\int_{V}\delta\rho\,\dot{s}_i^{\dagger}(t)\dot{s}_i(t)\,d^3\mathbf{x}\,dt + \int_{-\infty}^{\infty}\int_{V}\left[\int_{-\infty}^{\infty}\epsilon_{ij}^{\dagger}(t)\,\delta\dot{C}_{ijkl}(t-t')\epsilon_{kl}(t')\,dt'\right]d^3\mathbf{x}\,dt, \tag{1}$$

where ρ is the mass density, ϵ_{kl} and $\epsilon_{ij}^{\dagger}$ are the forward and adjoint strain, respectively. The $\dot{C}_{ijkl}$ represents the time derivative of the elastic tensor C_{ijkl} that can be simplified for an isotropic medium as $C_{ijkl} = (\kappa - \frac{2\mu}{3})\delta_{ij}\delta_{kl} + \mu\delta_{ik}\delta_{jl} + \mu\delta_{il}\delta_{jk}$, where μ and κ are the shear and bulk moduli. For the shear modulus variation, eq. (1) reduces to

$$\delta\chi = 2\int_{-\infty}^{\infty}\int_{V}\int_{-\infty}^{\infty}\tilde{\epsilon}_{ij}^{\dagger}(t)\,\delta\dot{C}_{\mu}(t-t')\,\tilde{\epsilon}_{ij}(t')\,dt'\,d^3\mathbf{x}\,dt, \tag{2}$$

where $\tilde{\epsilon}_{ij}$ and $\tilde{\epsilon}_{ij}^{\dagger}$ are the deviators of ϵ_{ij} and $\epsilon_{ij}^{\dagger}$, respectively. We denote C_μ here for the convenience of deriving the Q kernels in the following sections. The integral interval depends on the simulation system.

2.1 Fréchet kernels from frequency and time space

The shear modulus $\mu(\omega)$ may be written in the form (Kolsky 1965; Liu *et al.* 1976)

$$\mu(\omega) = \mu(\omega_0)\left[1 + \left(\frac{2}{\pi Q_\mu}\right)\ln\left(\frac{|\omega|}{\omega_0}\right) - i\,sgn(\omega)\frac{1}{Q_\mu}\right], \tag{3}$$

where $\mu(\omega_0)$ indicates the shear modulus at the reference frequency, $\omega_0 = 2\pi f_0$, which can be set accordingly. The ω and Q_μ are the angular frequency and quality factor, respectively. Considering the change in the shear modulus $\delta\mu(\omega)$ due to perturbation in shear attenuation δQ_μ^{-1}, one may write

$$\delta\mu(\omega) = \mu(\omega_0)\left[\left(\frac{2}{\pi}\right)\ln\left(\frac{|\omega|}{\omega_0}\right) - i\,sgn(\omega)\right]\delta Q_\mu^{-1}, \tag{4}$$

where $sgn(\omega)$ denotes the sign of ω. With one Fourier transform and one inverse Fourier transform, eq. (2) may be equivalent to

$$\delta\chi = 2\mu\int_{V}\left[\int_{-\infty}^{\infty}\tilde{\epsilon}_{ij}^{\dagger}\,\tilde{\bar{\epsilon}}_{ij}\,dt\right]\delta Q_\mu^{-1}\,d^3\mathbf{x} = \int_{V}K_{Q_\mu^{-1}}\,\delta Q_\mu^{-1}\,d^3\mathbf{x} = \int_{V}K_{Q_\mu}\frac{\delta Q_\mu}{Q}\,d^3\mathbf{x}, \tag{5}$$

where $K_{Q_\mu} = -K_{Q_\mu^{-1}}/Q_\mu$ and the shifted forward strain, $\bar{\epsilon}_{ij}$, may be computed by the shifted forward field (Tromp *et al.* 2005),

$$\bar{\mathbf{s}}(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\left[\left(\frac{2}{\pi}\right)\ln\left(\frac{|\omega|}{\omega_0}\right) - i\,sgn(\omega)\right]\mathbf{s}(\omega)\exp(i\omega t)\,d\omega. \tag{6}$$

The $\tilde{\bar{\epsilon}}_{ij}$ is the deviator of $\bar{\epsilon}_{ij}$. The $\mathbf{s}(\omega)$ is a Fourier transform of the regular forward field without the shift. The shift operator may be applied in an equivalent way to the adjoint field, where the Q adjoint field may be computed by an additionally defined adjoint source (Bozdağ *et al.* 2011; Zhu *et al.* 2015).

Following exactly the steps described for the formulation of $K_{Q_\mu^{-1}}$ in eq. (5), the $K_{Q_\kappa^{-1}}$ can be similarly derived by replacing the μ with κ and omitting the number '2'. If the reference frequency is selected to be the central frequency of the simulation band and Q is constant over a frequency range, eq. (3) may approximate the complex modulus $M(\omega)$ (e.g. Emmerich & Kornm 1987; Carcione 1988; Moczo & Kristek 2005) in terms of μ within the simulated frequency band. See eq. (17) to eq. (38) of Liu *et al.* (1976) for a proof. Note that the work of Liu *et al.* (1976) uses the creep function, an inverse of the relaxation function, and the $1/N$ difference has been discussed in the work of Zhang *et al.* (2016). As an equivalent expression between the Zener and Maxwell body, the complex modulus may be written as (e.g. Moczo & Kristek 2005)

$$M(\omega) = \frac{M^r}{N}\sum_{l=1}^{N}\frac{1 + iw\tau_\epsilon^{(l)}}{1 + iw\tau_\sigma^{(l)}} = M^r\left[1 + \frac{1}{N}\sum_{l=1}^{N}\frac{\omega^2\tau_\sigma^{(l)}\left(\tau_\epsilon^{(l)} - \tau_\sigma^{(l)}\right)}{1 + \omega^2\left(\tau_\sigma^{(l)}\right)^2} + i\frac{1}{N}\sum_{l=1}^{N}\frac{\omega\left(\tau_\epsilon^{(l)} - \tau_\sigma^{(l)}\right)}{1 + \omega^2\left(\tau_\sigma^{(l)}\right)^2}\right], \tag{7}$$

where M^r denotes the relaxed modulus, $\tau_\epsilon^{(l)}$ and $\tau_\sigma^{(l)}$ are the relaxation times and N is the maximum number of relaxation mechanisms. Eq. (7) shows the shift operator which includes the real and imaginary parts to shift the relaxed modulus M^r. Invoking the convolution and Fourier

transform properties, the misfit function of eq. (2) may be written as

$$\delta\chi = 2\int_{-\infty}^{\infty}\int_{V}\tilde{\epsilon}_{ij}^{\dagger}(t)\left[\int_{-\infty}^{\infty}\tilde{\epsilon}_{ij}(t')\delta\dot{C}_{\mu}(t-t')\,dt'\right]d^3\mathbf{x}\,dt = 2\int_{-\infty}^{\infty}\int_{V}\tilde{\epsilon}_{ij}^{\dagger}(t)\mathcal{F}^{-1}\left[\tilde{\epsilon}_{ij}(\omega)\delta M_{\mu}(\omega)\right]d^3\mathbf{x}\,dt, \tag{8}$$

where $M_{\mu}(\omega)$ represents the frequency domain of the time derivative of the elastic tensor $C_{\mu}(t-t^{'})$. Eq. (8) is equivalent to eq. (5) and shows that the kernels may also be computed by using a shift operator $\delta M_{\mu}(\omega)$ to the adjoint strain in the frequency domain and back to the time domain. The shift operator may be used in the time domain. The time domain modulus, $M(t)$, of the complex modulus $M(\omega)$ may be computed from the relaxation function (e.g. Liu *et al.* 1976; Emmerich & Kornm 1987; Carcione 2001; Moczo & Kristek 2005; Fichtner & van Driel 2014)

$$C(t) = C^{r}\left[1+\frac{1}{N}\sum_{l=1}^{N}\left(\frac{\tau_{\epsilon}^{(l)}}{\tau_{\sigma}^{(l)}}-1\right)e^{-t/\tau_{\sigma}^{(l)}}\right]H(t) = C^{r}\left[1+\tau\sum_{l=1}^{N}D^{(l)}e^{-t/\tau_{\sigma}^{(l)}}\right]H(t) \tag{9}$$

where C^{r} denotes the relaxed modulus, τ controls the strength of viscoelastic attenuation, and $D^{(l)}$ represents the weights of the relaxation mechanisms. By setting a constant $\tau = 1/Q$, the change of $D^{(l)}$ and $\tau_{\sigma}^{(l)}$ could be enough to approximate the input Q in a wide frequency band (Fichtner & van Driel 2014). Differentiating the right hand side of eq. (9) with respect to time t yields

$$M(t) = M^{r}\left[1+\frac{1}{Q}\sum_{l=1}^{N}D^{(l)}e^{-t/\tau_{\sigma}^{(l)}}\right]\delta(t) - M^{r}\frac{1}{Q}\sum_{l=1}^{N}\frac{D^{(l)}}{\tau_{\sigma}^{(l)}}e^{-t/\tau_{\sigma}^{(l)}}\,H(t), \tag{10}$$

where $\delta(t)$ is the delta function. The time domain modulus of eq. (10) corresponds to the frequency domain modulus of eq. (7). Differentiating eq. (10) with respect to Q^{-1} leads to

$$\frac{\partial M(t)}{\partial Q^{-1}} = M^{r}\sum_{l=1}^{N}D^{(l)}e^{-t/\tau_{\sigma}^{(l)}}\delta(t) - M^{r}\sum_{l=1}^{N}\frac{D^{(l)}}{\tau_{\sigma}^{(l)}}e^{-t/\tau_{\sigma}^{(l)}}\,H(t). \tag{11}$$

By the chain rule $\delta M = (\partial M/\partial Q^{-1})\delta Q^{-1}$ and substituting eq. (11) into eq. (2), the misfit function may be written as

$$\delta\chi = 2M_{\mu}^{r}\int_{V}\int_{-\infty}^{\infty}\left[\left(\sum_{l=1}^{N}D^{(l)}\tilde{\epsilon}_{ij}^{\dagger}+Q_{\mu}\tilde{R}_{ij,\mu}^{\dagger}\right)\right]\tilde{\epsilon}_{ij}(t')dt'\delta Q_{\mu}^{-1}\,d^3\mathbf{x}. \tag{12}$$

The auxiliary term, $\tilde{R}_{ij,\mu}^{\dagger}$, that relates to the N summation of the deviator of the adjoint strain, may be defined by

$$\tilde{R}_{ij,\mu}^{\dagger} = -\frac{1}{Q_{\mu}}\sum_{l=1}^{N}\frac{D^{(l)}}{\tau_{\sigma}^{(l)}}\int_{-\infty}^{\infty}e^{-(t'-t)/\tau_{\sigma}^{(l)}}H(t'-t)\,\tilde{\epsilon}_{ij}^{\dagger}\,dt'. \tag{13}$$

The $\tilde{R}_{ij,\mu}^{\dagger}$ may be called the adjoint memory variable although there are many alternative expressions that exist in the literature. Based on eq. (12), we obtain the Fréchet Q kernels for Q_{μ}^{-1},

$$K_{Q_{\mu}^{-1}} = 2M_{\mu}^{r}\int_{-\infty}^{\infty}\left[\sum_{l=1}^{N}D^{(l)}\tilde{\epsilon}_{ij}^{\dagger}+Q_{\mu}\tilde{R}_{ij,\mu}^{\dagger}\right]\tilde{\epsilon}_{ij}\,dt. \tag{14}$$

Eqs (13) and (14) indicate that the Q kernel can be computed by the forward strain deviator ($\tilde{\epsilon}_{ij}$) and a shift of the adjoint strain deviator ($\tilde{\epsilon}_{ij}^{\dagger}$) with the shift operator given by $[\sum_{l=1}^{N}D^{(l)} - \sum_{l=1}^{N}\frac{D^{(l)}}{\tau_{\sigma}^{(l)}}\int_{-\infty}^{\infty}e^{-(t'-t)/\tau_{\sigma}^{(l)}}H(t'-t)\,dt']$. The second term in the bracket relates to the adjoint memory variable depending on how the memory variable is parameterized. Following exactly the same steps as derived for eq. (14), the Fréchet kernel for Q_{κ}^{-1} may be written as

$$K_{Q_{\kappa}^{-1}} = M_{\kappa}^{r}\int_{-\infty}^{\infty}\left[\sum_{l=1}^{N}D^{(l)}\epsilon_{kk}^{\dagger}+Q_{\kappa}R_{\kappa}^{\dagger}\right]\epsilon_{kk}\,dt, \tag{15}$$

where ϵ_{kk} and $\epsilon_{kk}^{\dagger}$ are the traces of the strain for the forward and adjoint field, respectively. The summed adjoint memory variable is

$$R_{\kappa}^{\dagger} = -\frac{1}{Q_{\kappa}}\sum_{l=1}^{N}\frac{D^{(l)}}{\tau_{\sigma}^{(l)}}\int_{-\infty}^{\infty}e^{-(t'-t)/\tau_{\sigma}^{(l)}}H(t'-t)\,\epsilon_{kk}^{\dagger}\,dt'. \tag{16}$$

The Q Fréchet kernels can be similarly derived for Q_{μ} and Q_{κ} using the derivative of $1/Q$, that is $K_{Q_{\mu}} = -K_{Q_{\mu}^{-1}}/Q_{\mu}$ and $K_{Q_{\kappa}} = -K_{Q_{\kappa}^{-1}}/Q_{\kappa}$, which is derived for $\delta Q/Q$ instead of δQ^{-1}. The $K_{Q_{\mu}}$ and $K_{Q_{\kappa}}$ here are equivalent (or approximately equivalent) to eqs (35) and (40) of Fichtner & van Driel (2014) when the perturbation of the time derivative of $C(t\text{-}t')$ is equal (or approximately equal) to the time derivative of

the perturbed $C(t\text{-}t')$. Note that these Q kernels derived in this work are computed by the displacement–stress formulation without the time derivative of the adjoint strain and its associated memory variables. Either the relaxed or unrelaxed modulus can be used to compute the Q kernels since the relaxation times are already computed. A short note of how the Q kernels are computed by the public SPECFEM2D and SPECFEM3D Cartesian packages is given in Appendix C.

Eqs (14) and (15) are frequency-independent kernels that assume a constant Q within a frequency range, similar to that of the frequency-independent Fréchet kernels presented in the work of Fichtner & van Driel (2014). Keeping the Q_0 for a specific targeted frequency range fixed, the Q can be considered as a function of α by (e.g. Müller 1983; Fichtner & van Driel 2014)

$$Q_{target}(\omega) = Q_0\left(\frac{\omega}{\omega_0}\right)^{\alpha} \approx Q_{simulated}(Q_0, \omega, D^{(l)}, \tau_\sigma^{(l)}). \tag{17}$$

Based on the chain rule, $\delta M/\delta\alpha = (\partial M/\partial Q)(\partial Q/\partial\alpha)$, the misfit function of eq. (2) leads to

$$\delta\chi = 2\int_V \left[\int_{-\infty}^{\infty} \alpha\frac{\partial Q_\mu}{\partial\alpha}\left(\frac{-M_\mu^r}{Q_\mu^2}\sum_{l=1}^{N} D^{(l)}\tilde{\epsilon}_{ij}^{\dagger} - \frac{M_\mu^r}{Q_\mu}\tilde{R}_{ij,\mu}^{\dagger}\right)\tilde{\epsilon}_{ij}(t')\,dt'\right]\frac{\delta\alpha}{\alpha}d^3\mathbf{x}, \tag{18}$$

which forms the frequency-dependent $K_{Q_{\mu\alpha}}$ due to the perturbation of $\delta\alpha$ as

$$K_{Q_{\mu\alpha}} = -2M_\mu^r\alpha\frac{\partial Q_\mu}{\partial\alpha}\int_{-\infty}^{\infty}\left(\frac{1}{Q_\mu^2}\sum_{l=1}^{N} D^{(l)}\tilde{\epsilon}_{ij}^{\dagger} + \frac{1}{Q_\mu}\tilde{R}_{ij,\mu}^{\dagger}\right)\tilde{\epsilon}_{ij}\,dt, \tag{19}$$

where the $\frac{\partial Q}{\partial\alpha}$ may be computed by a finite-difference approximation similar to that used in eq. (47) in the work of Fichtner & van Driel (2014). The $D^{(l)}$ and $\tau_\sigma^{(l)}$ are recomputed for a slightly perturbed α, which is computationally cheap in the meshing stage in comparison to the subsequent wave simulations. Fréchet kernels for fractional perturbations in shear α are orders of magnitude smaller than kernels for fractional perturbations in shear Q, which means that the frequency-dependence of Q in the Earth is difficult to constrain (Fichtner & van Driel 2014). Similar derivations form the frequency-dependent $K_{Q_{\kappa\alpha}}$,

$$K_{Q_{\kappa\alpha}} = -M_\kappa^r\alpha\frac{\partial Q_\kappa}{\partial\alpha}\int_{-\infty}^{\infty}\left(\frac{1}{Q_\kappa^2}\sum_{l=1}^{N} D^{(l)}\epsilon_{kk}^{\dagger} + \frac{1}{Q_\kappa}R_\kappa^{\dagger}\right)\epsilon_{kk}\,dt. \tag{20}$$

2.2 Hessian kernels for frequency-(in)dependent Q

Based on previous work (Fichtner & Trampert 2011; Fichtner & van Driel 2014; Xie *et al.* 2021), the full Hessian kernel with respect to Q_μ^{-1} may be written as

$$H_{Q_\mu^{-1}} = 2M_\mu^r\int_{-\infty}^{\infty}\left[\left(\sum_{l=1}^{N} D^{(l)}\tilde{\epsilon}_{ij}^{\dagger} + Q_\mu\tilde{R}_{ij,\mu}^{\dagger}\right)\delta_1\tilde{\epsilon}_{ij} + \left(\sum_{l=1}^{N} D^{(l)}\delta_1\tilde{\epsilon}_{ij}^{\dagger} + Q_\mu\delta_1\tilde{R}_{ij,\mu}^{\dagger}\right)\tilde{\epsilon}_{ij}\right]dt, \tag{21}$$

where $\delta_1\tilde{R}_{ij,\mu}^{\dagger}$ denotes the perturbation of the deviator of the adjoint memory variable, which can be computed by a similar implementation of eq. (13) by substituting $\tilde{\epsilon}_{ij}^{\dagger}$ with $\delta_1\tilde{\epsilon}_{ij}^{\dagger}$. It is worth noting that there are two sets of relaxation times for two models used in the calculation of the perturbation. By using a similar derivation as used for eq. (21), we have $H_{Q_\mu} = -H_{Q_\mu^{-1}}/Q_\mu$. Implementation of eq. (21) requires access to these strains and variables simultaneously. It can be done by combining the on-the-fly approach (Xie *et al.* 2021) with the checkpointing parsimonious storage approach (Komatitsch *et al.* 2016) for recomputing the viscoelastic forward fields during the adjoint simulation. Similarly, the full Hessian kernels for the Q_κ^{-1} may be expressed as

$$H_{Q_\kappa^{-1}} = M_\kappa^r\int_{-\infty}^{\infty}\left[\left(\sum_{l=1}^{N} D^{(l)}\epsilon_{kk}^{\dagger} + Q_\kappa R_\kappa^{\dagger}\right)\delta_1\epsilon_{kk} + \left(\sum_{l=1}^{N} D^{(l)}\delta_1\epsilon_{kk}^{\dagger} + Q_\kappa\delta_1 R_\kappa^{\dagger}\right)\epsilon_{kk}\right]dt. \tag{22}$$

A similar derivation for eq. (22) can be used for $H_{Q_\kappa} = -H_{Q_\kappa^{-1}}/Q_\kappa$. Frequency-dependent Hessian kernels, that is those related to perturbations in α, are not likely to be well constrained in the Earth. Frequency-dependent Hessian kernel $H_{Q_{\mu\alpha}}$ and $H_{Q_{\kappa\alpha}}$ may be similarity constructed by

$$H_{Q_{\mu\alpha}} = -2M_\mu^r\alpha\frac{\partial Q_\mu}{\partial\alpha}\int_{-\infty}^{\infty}\left[\left(\frac{1}{Q_\mu^2}\sum_{l=1}^{N} D^{(l)}\tilde{\epsilon}_{ij}^{\dagger} + \frac{1}{Q_\mu}\tilde{R}_{ij,\mu}^{\dagger}\right)\delta_1\tilde{\epsilon}_{ij} + \left(\frac{1}{Q_\mu^2}\sum_{l=1}^{N} D^{(l)}\delta_1\tilde{\epsilon}_{ij}^{\dagger} + \frac{1}{Q_\mu}\delta_1\tilde{R}_{ij,\mu}^{\dagger}\right)\tilde{\epsilon}_{ij}\right]dt, \tag{23}$$

$$H_{Q_{\kappa\alpha}} = -M_\kappa^r\alpha\frac{\partial Q_\kappa}{\partial\alpha}\int_{-\infty}^{\infty}\left[\left(\frac{1}{Q_\kappa^2}\sum_{l=1}^{N} D^{(l)}\epsilon_{kk}^{\dagger} + \frac{1}{Q_\kappa}R_\kappa^{\dagger}\right)\delta_1\epsilon_{kk} + \left(\frac{1}{Q_\kappa^2}\sum_{l=1}^{N} D^{(l)}\delta_1\epsilon_{kk}^{\dagger} + \frac{1}{Q_\kappa}\delta_1 R_\kappa^{\dagger}\right)\epsilon_{kk}\right]dt. \tag{24}$$

The two frequency-dependent full Hessian kernels can be computed by perturbing α. The $\partial Q/\partial\alpha$ here may be computed by eq. (47) in the work of Fichtner & van Driel (2014). Note the difference between the Fréchet and Hessian kernels in terms of the power-law exponent α as indicated by their expressions, where the latter takes the perturbed forward and perturbed adjoint fields into account.

The above Fréchet and Hessian kernels can be similarly derived for the Generalized Maxwell Body, GMB-EK, defined in the work of Emmerich & Kornm (1987). The work of Moczo & Kristek (2005) presents an equivalent expression between the Zener and Maxwell body. For the Q kernel calculation, it requires enforcing Q into eq.(5) of Moczo & Kristek (2005), which ensures the input Q is approximated closely for the simulations. The associated stress and memory variables can be computed similarly for the wavefield and kernel calculations.

3 ADJOINT TOMOGRAPHY USING KERNELS

The total misfit for an isotropic anelastic inversion may be defined as (e.g. Zhu *et al.* 2015)

$$\chi = \omega_\phi \chi_\phi + \omega_A \chi_A, \tag{25}$$

where ω_ϕ and ω_A are the weighting factors for phase and amplitude misfit. The initial two weighting factors may be computed by the phase and amplitude normalised misfit with respect to model parameters. Multiple misfit functions and associated adjoint sources may be used (e.g. Bozdağ *et al.* 2011; Karaoğlu & Romanowicz 2018; Pan & Wang 2020; Pan *et al.* 2021). The perturbation in the misfit may be expressed as

$$\delta\chi = \int_V \left[K_C \frac{\delta C}{C} + K_Q \frac{\delta Q}{Q} \right] d^3\mathbf{x} = \int_V \left[K_C \frac{\delta C}{C} + K_{Q^{-1}} \delta Q^{-1} \right] d^3\mathbf{x}, \tag{26}$$

where K_C and K_Q (or $K_{Q^{-1}}$) denote the elastic and Q Fréchet kernels, and $\delta C/C$ and $\delta Q/Q$ represent the relative model perturbations in terms of the elastic and Q properties. Optimization algorithms are commonly used for solving the full-waveform inverse problem, which are typically designed for solving the Newton linear system:

$$\mathbf{H}(m_k)\Delta\mathbf{m}_k = -\nabla_m \chi(m_k), \tag{27}$$

where $\mathbf{H}$ and $\nabla_m \chi$ denote the Hessian matrix and gradient (Fréchet kernels), respectively. The k represents the iteration index. Different optimization algorithms differ in the search for the model update $\Delta\mathbf{m}$. In each inversion step, we update the model by

$$\mathbf{m}_{k+1} = \mathbf{m}_k + a_k \Delta\mathbf{m}_k, \tag{28}$$

where a_k is a scalar parameter computed through a line-search or a trust-region procedure (e.g. Bonnans *et al.* 2006; Nocedal & Wright 2006). For different optimization and search algorithms, we refer to the work of Nocedal & Wright (2006). We use the Newton-CG algorithm (see Nocedal & Wright 2006, p. 112 and p. 119) for the Hessian kernel inversion and its results are primarily compared with that of the L-BFGS FWIs implemented in the SeisFlows package (Modrak *et al.* 2018). We call this the L-BFGS inversion throughout this work as a reference since there are many possible different L-BFGS implementations with possible differences in performance in the literature. For the Newton-CG implementation, for example, for stop criterion and threshed update, one can refer to the codes of Métivier & Brossier (2016). We use the bracket line search to find the step length. In the implementation of the line search, the step length for velocity and Q might vary due to the distinct scales of velocity and Q, where the step length is a vector. A comparison of the L-BFGS and the Newton-CG algorithm and the use of different step search methods are not the purpose of this work. We use the L-BFGS inversion as a reference for comparison to see the performance of the full Hessian kernel inversion. For simplicity in the following sections, inversions that use the full Hessian kernels that utilize the Newton-CG algorithm are called full Hessian kernel inversions or Hessian kernel inversions, while the approximate Hessian inversions that use the gradient-based L-BFGS algorithm are called L-BFGS inversions. Here we demonstrate the approach with synthetics as an initial step. Although other factors like anisotropy are also likely important in the real Earth, we leave this as a topic for future work.

4 MODELS

4.1 2-D models

We initially examine two 2-D synthetic examples (Figs 1 and 2) to explore the effectiveness of the full Hessian kernel inversion schemes. The first model has an idealized setup designed to demonstrate the potential of the inversion schemes (Fig. 1). It has sources at regularly spaced intervals near the bottom of the model and regularly spaced receivers at the top. The model has a uniform background shear wave velocity of 3900 km s^{-1}, compressional wave velocity of 6800 km s^{-1}, Q_μ of 125, Q_κ of 3000, a density of 2600 kg m^{-3} at shallower than 50 km depth and a density of 2900 kg m^{-3} at deeper depths. Within the model there are five rectangular low velocity zones (5, 10 and 3.3 per cent low for Vp, Vs and ρ, respectively) and three rectangular Q_μ anomalies ($Q_\mu = 30$) that are not co-located with the velocity anomalies. The difference in the Q_μ and velocity anomaly locations allows us to examine parameter trade-off. The second model is based on an attenuation study from the Central American subduction zone (Rychert *et al.* 2008), where a temporary seismic network was deployed for 2 yr by the TUCAN (Tomography and other things Under Costa Rica And Nicaragua) experiment. The deployment consisted of two dense arc, perpendicular lines, one in Nicaracua and one in Costa Rica, and two more broadly spaced arc parallel lines, one in the forearc, and one in the backarc. We use stations in the dense arc perpendicular Nicaragua line here. The goal of testing using this model is to examine the ability of the method to recover structure with realistic source–receiver geometry and approximately real Earth structures. The source locations are taken from real source locations in the subducting slab and overriding plate used in the study, as well as the receiver locations from the deployment. The

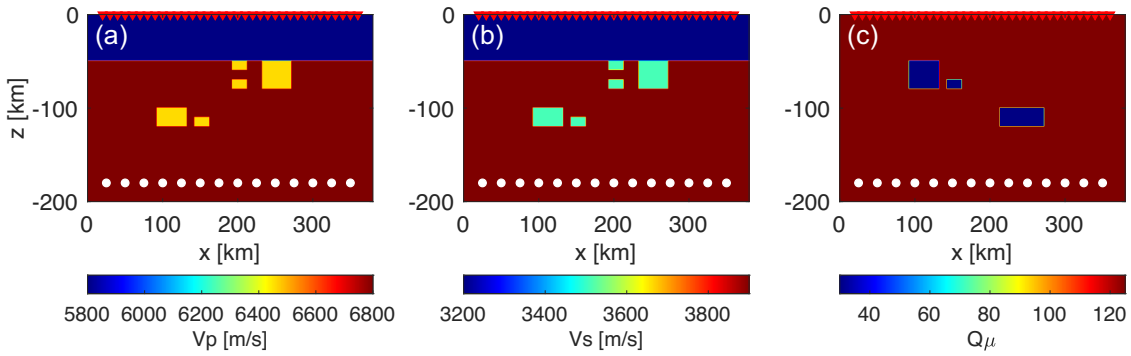

Figure 1. The idealized synthetic model is shown for the true *Vp* (a), *Vs* (b) and Q_μ (c). The models are constructed using background models in which rectangular anomalies are imbedded. We use a density of 2600 kg m^{-3} at depths shallower than $z = -50$ km and a density of 2900 kg m^{-3} at depths deeper than $z = -50$ km. The background models are used as the initial elastic models. The Q background models are set to $Q_\kappa = 3000$ and $Q_\mu = 125$ and used as the initial anelastic models. The white dots in each subfigure denote the source locations, while the red inverted triangles on the top of the model represent the receivers.

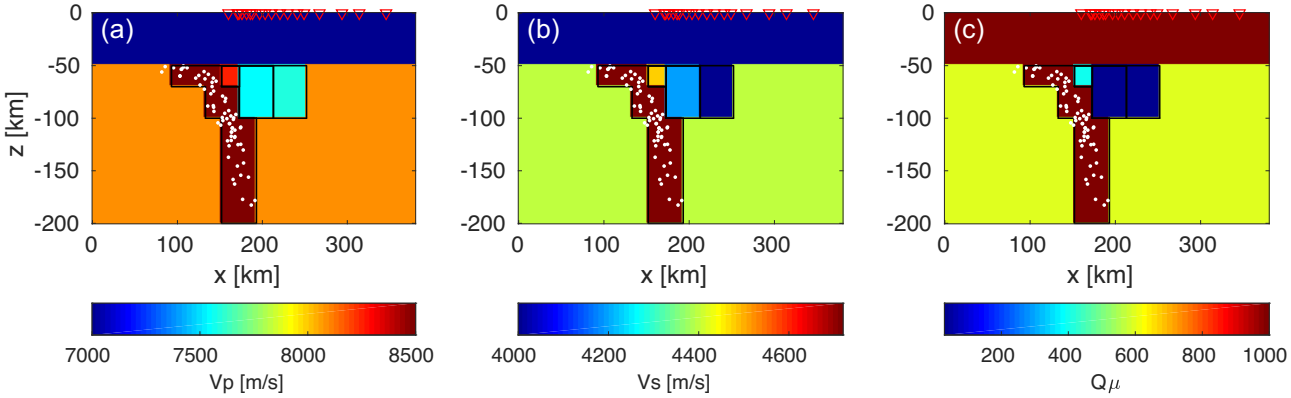

Figure 2. The more realistic Nicaragua subduction model is shown for the true *Vp* (a), *Vs* (b) and Q_μ (c). We use a constant Q_κ model of 3000, and a two-layer density model with its upper part set to 2800 km m^{-3} and bottom part set to 3300 km m^{-3}, separated at $z = -50$ km. The white dots within the dark red slab denote the earthquakes and the red inverted triangles represent the stations.

model has a Q structure taken from the minimum parameterization inversion results from the work of Rychert *et al.* (2008; Fig. 2). The model has a *Vs* structure with the same anomaly shapes as those in the Q model, but with velocities based on the shear wave velocities derived from a regional surface wave model (Harmon *et al.* 2013). The velocity and attenuation structures generally reflect the pertinent features of a subduction zone, including a slab, overriding plate, cold nose and arc and backarc anomalies.

The two 2-D synthetic models have the same dimensions of 380 km × 200 km in the horizontal and vertical directions, respectively (Figs 1 and 2). Before the forward or adjoint simulation, we generate the database and mesher using the QuadSEM-Q package (https://github.com/yujiangxie/QuadSEM). The package was developed for computing the 2-D full Hessian kernels on the fly (elastic) or partially on the fly (anelastic, via checkpointing) based on the public SPECFEM2D package (Tromp *et al.* 2008). We use 5 × 5 Gauss-Lobatto-Legendre (GLL) points for each element and place 190 elements in the horizontal direction and 200 elements in the vertical direction, resulting in approximately 500 m and 250 m grid spacings in the horizontal and vertical directions, respectively. We use 10 000-time steps and a time interval of $dt = 0.01$ s for the forward and adjoint simulations. For the viscoelastic wave simulations, we use 100 time steps for each chunk in order to recompute the forward fields for a chunk between two snapshots. Please see the work of Komatitsch *et al.* (2016) for an explanation of the chunk. This leads to 100 checkpoints for one forward simulation. For the first box model (Fig. 1), 14 events based on the same moment tensor solution are used. They are located at $z = -180$ km, and every 25 km in the x-direction between 25 km and 350 km. The three components of the moment tensor solution are, $Mxx = 8.14e + 12$, $Mzz = 3.93e + 13$ and $Mxz = 1.96e + 13$. The same triangle source wavelet with the period of 1 s is used for all events. We use thirty five evenly-spaced receivers at the top of the model, separated by 10 km, ranging from $x = 20$ to 360 km (see Fig. 1). To simulate the anelastic properties, we use three (i.e., $N = 3$) standard linear solids. The reference frequency for shifting the velocities in the viscoelastic simulation is set to the central frequency of the simulation frequency band based on the 2-D implementation, instead of the infinite frequency that corresponds to the unrelaxed modulus. For the subduction model, all required starting setups for the simulations are set as the same as the box model, except the number of events is increased to 69 to simulate the earthquakes distributed along the slab. We use eighteen receivers on the top of the model (Fig. 2).

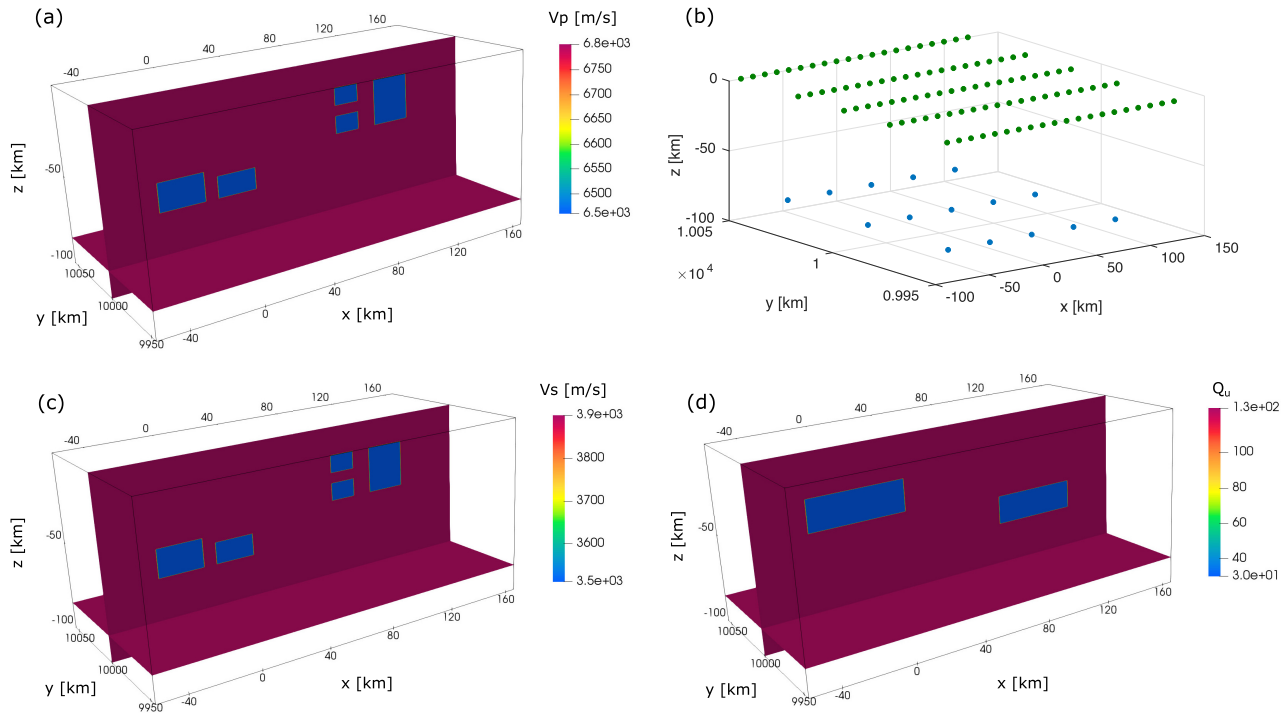

Figure 3. 3-D synthetic model. (a) the compressional wave speed, Vp (m s^{-1}), with –5 per cent perturbation relative to the background model setting for the five scatterers. (b) the source and receiver geometry used for the inversion. The blue dots indicate the source locations buried at $z = -85$ km. The source interval is set to 38.85 km in both the x- and y-directions. The green dots indicate the station locations at an interval that is set to 11.1 km in the x-direction and 27.75 km in the y-direction. (c) the shear wave speed, Vs (m s^{-1}), with –10 per cent perturbation relative to the background model. (d) the attenuation quality factor, Q_μ, where the background is set to $Q_\mu = 125$ and the anomaly is set to $Q_\mu = 30$. The perturbation to density, ρ, is set by the following relationship: $\delta p/p = 0.33\,\delta Vs/Vs$. The perturbation of Q_κ shares the same shape and location as that of the Q_μ but with the background set to $Q_\kappa = 3000$ and the anomaly set to $Q_\kappa = 102$.

4.2 3-D models

The 3-D model has the same model properties and perturbations as the 2-D idealized model except the dimension and size of the model and perturbations are different (Fig. 3). We use 15 events and 100 receivers for the inversion. We use a moment tensor source for all events and the six components of the moment tensor are set to $Mrr = -4.100000e{+}22$, $Mtt = 1.990000e{+}23$, $Mpp = -1.580000e{+}23$, $Mrt = -7.170000e{+}22$, $Mrp = 1.220000e{+}23$, $Mtp = -2.910000e{+}23$. An external triangle source time function with the period of 1.2 s is used. The time step, dt, is set to 0.02 s based on an estimation of wave speeds and grid spacing of mesher. The total number time step is of $NT = 5000$, which leads to a 100 s seismogram record. The model is set to 222 km in the x-direction and 111 km in the y-direction witha maximum depth of 100 km in the vertical direction. We use 60 elements in the x-direction, 30 elements in the y direction, and 50 elements in the depth direction. Each element covers 3.7 km of model space horizontally and 2 km vertically (Fig. 3). We use $5 \times 5 \times 5$ GLL points for each element. We first invert for an elastic model and then simultaneously invert for an elastic and Q model. The reference frequency for shifting the velocities in the 3-D viscoelastic simulation is also set to the central frequency of the simulation frequency band although it can be set to other frequencies. We use 50 checkpoints ($50 \times 100 = 5000$ time steps) for the 3-D viscoelastic simulations.

5 RESULTS AND DISCUSSIONS

5.1 2-D example I: the idealized scatterer model

5.1.1 Q_μ and Vs approximation

We check the Q and Vs approximation before the anelastic wave simulations to ensure the relaxation times are reasonably determined. Fig. 4 shows the Q and velocity approximations in terms of angular frequency, ω. The Q is approximated by eq. (10) of Fichtner & van Driel (2014), which is equivalent to solving an optimization problem using the objective function of eq. (12) of Fichtner & van Driel (2014). We solve the optimization problem by the SolvOpt optimization (Kappel & Kuntsevich 2000) as developed by Blanc *et al.* (2016) in the SPECFEM2D package, enforcing $D^{(l)}/Q = \tau^{(l)}/N$. It can be seen from Fig. 4 that the Q is approximated well within the frequency band while using three standard linear solids. The shear wave velocity (Vs; Fig. 4b) is computed in terms of ω by the frequency-dependent viscoelastic modulus $M(\omega)$. The relaxed, unrelaxed and reference modulus can be transformed among each other once the $\tau_\sigma^{(l)}$ and $D^{(l)}$ are determined. We use a

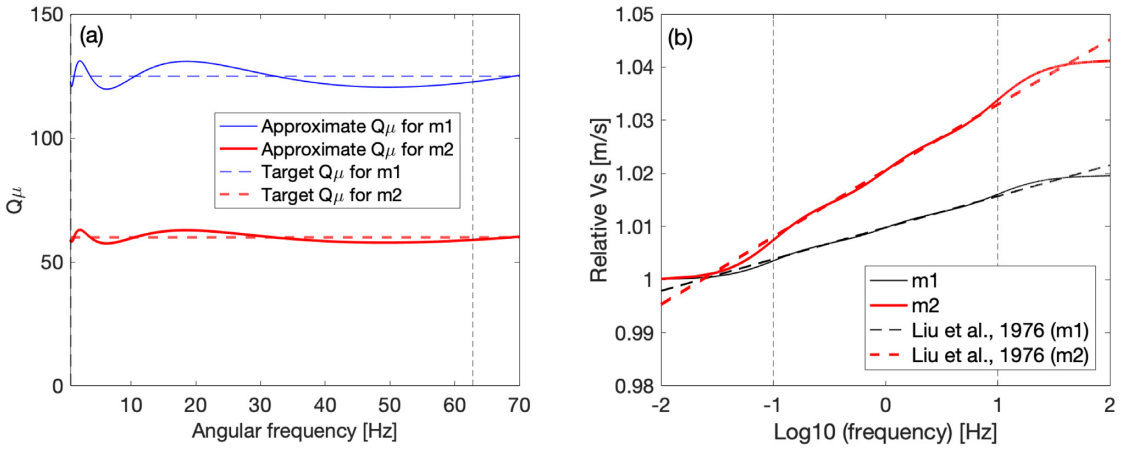

Figure 4. Approximation of the Q_μ and Vs in terms of angular frequency ω. (a) Q as a function of angular frequency as approximated with three standard linear solids for two models. (b) The viscoelastic shear wave velocity $Vs(\omega)$ versus the shear wave velocity of Liu *et al.* (1976) for two models, where the $Vs(\omega)$ is computed by the viscoelastic modulus $M(\omega)$ of eq. (7). The vertical dash lines indicate the simulated frequency boundary. At lower frequencies the velocities of m1 and m2 are relatively constant and correspond to the fully relaxed velocity. At higher frequencies the m1 and m2 velocities are similarly constant, corresponding to the unrelaxed velocity.

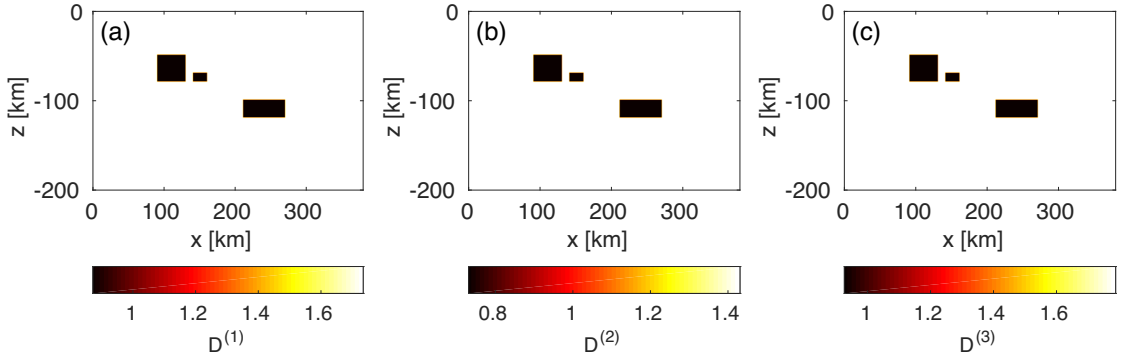

Figure 5. Optimal weights $D^{(l)}$ for approximating Q_μ of Fig. 1(c), where the three optimal relaxation times are $\tau_\sigma^{(1)} = 1.5915$, $\tau_\sigma^{(2)} = 0.1592$ and $\tau_\sigma^{(3)} = 0.0159$. The minimum and maximum frequency search range is set from 0.01 to 10 Hz, and $\omega_0 = 2\pi$. The weights $D^{(l)}$, $l = 1, 2, 3$, are used for the three standard linear solids.

reference frequency of $f_0 = 1$ Hz and the associated reference velocity is $V(\omega_0)$, where $\omega_0 = 2\pi f_0$ (Fig. 4b). The relaxed modulus is computed from the relaxed velocities from the input that can be prepared before the simulation. One key point here is to ensure the two lines are matched within the simulated frequency band as shown in Fig. 4(b). The optimal weights $D^{(l)}$ and the relaxation times for approximating the Q_μ of Fig. 1(c) are shown in Fig. 5. These are consistent with the input Q_μ model, and this indicates that the Q is approximated accurately.

5.1.2 Fréchet and full Hessian kernels for Q_μ^{-1}

In Section 2 we presented a way to calculate the Fréchet and full Hessian kernels for variations in Q using a shifted adjoint wavefield. The calculation is also presented for using a shifted adjoint source in Appendix A. Using a shifted adjoint wavefield reduces the number of simulations required for the calculation of the Fréchet kernel by one or two in comparison to the unshifted approach. For the full Hessian kernel calculation, the number of simulations is reduced by two or four.

We now proceed to compare and contrast the kernels calculated using the shifted adjoint source with the kernels calculated using the shifted adjoint wavefield to quantify the differences that result from two different calculations. We also compare the full Hessian kernels with the Hessian kernels calculated by a finite-difference approximation.

Fig. 6 shows the waveforms and adjoint sources for the Q Fréchet kernel calculations for one source-receiver geometry using the model of Fig. 1, where the source is located at $x = 125$ km and $z = -180$ km, and the receiver is located at $x = 130$ km and $z = 0$ km. We use a standard Ricker wavelet with the dominant frequency of 1 Hz to compute the kernels. The regular adjoint sources in Fig. 6(b) are computed by eq. (67) of Tromp *et al.* (2005) and the shifted adjoint sources in Fig. 6(b) are computed by eq. (52) of Bozdağ *et al.* (2011) once the regular adjoint sources are computed. Fig. 7 shows the Q Fréchet kernels, where Fig. 7(a) shows the Q Fréchet kernel computed using the shifted adjoint source [see eq. (5) for the $K_{Q_\mu^{-1}}$ kernel], which is approximately numerically equivalent to Fig. 7(b) computed by eq. (14) using the shifted adjoint wavefield. The two kernels are slightly different in amplitude. The kernel computed by eq. (14) is a factor of –1.02 and 1.51 larger than the kernel computed by eq. (5) when comparing the minimum and maximum kernel values. This may be due to the lower

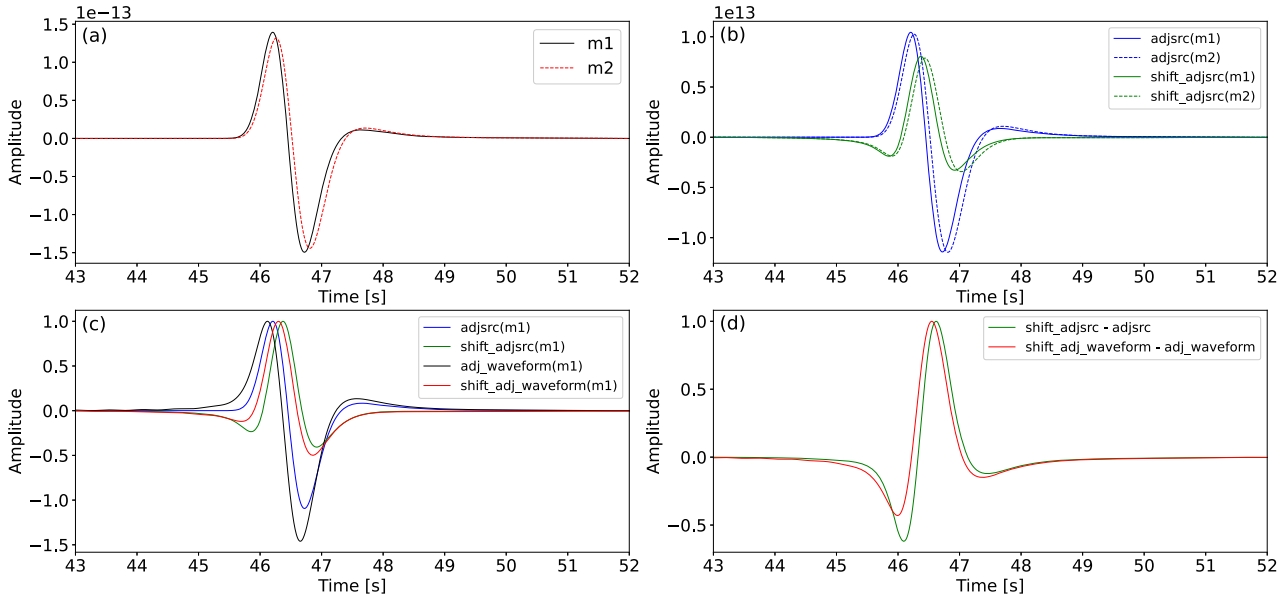

Figure 6. Waveforms and adjoint sources are shown for the *x*-components. (a) Displacement waveforms computed for m1 and m2. (b) Regular and shifted adjoint source due to attenuation anomalies for the two models. (c) Regular adjoint sources and regular adjoint waveforms and their shifts for m1. The waveforms have been normalized for comparison. (d) The difference between the shifted and regular adjoint sources for m1 (green) and the difference between the shifted and regular adjoint waveform for m1 (red). This demonstrates that if the shift operator is applied to the adjoint source it leads to a deviation in comparison to applying it to the adjoint wavefield (eqs 6 and 8). The adjoint waveform here represents the adjoint field recorded at the receiver.

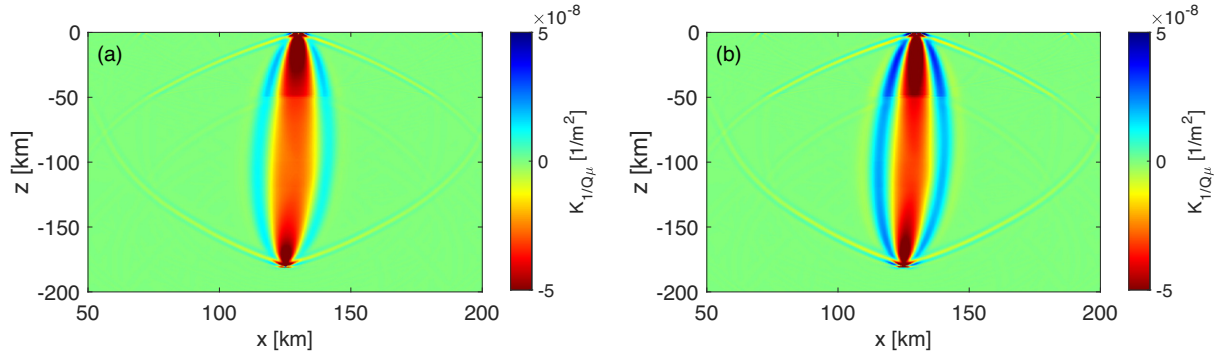

Figure 7. Comparison of Q Fréchet kernels for two shift operators. (a) The kernel calculated using the shifted adjoint source. (b) The kernel calculated by eq. (14) using the shifted wavefield. The original kernel values range from -4.09×10^{-7} to 1.53×10^{-7} and -4.01×10^{-7} to 2.31×10^{-7}, respectively. The first Fresnel zone (negative values) of the two kernels are similar but the second Fresnel zone (positive values) are much different, with panel (b) being about 1.5 times than that in (a) due to the different accuracies of the shift operators. The colour bar is clipped for visualization purposes.

accuracy that is achieved by applying the shift to the adjoint source (Fig. 7a) in comparison to applying it to the adjoint field (Fig. 7b). In practice, small differences in the kernel amplitudes may affect the convergence rate, but may not necessarily the final result.

The full Hessian kernel is composed of three components (Fichtner & Trampert 2011; Xie *et al.* 2021). Fig. 8 shows the three components of the full Hessian kernel computed for a viscoelastic media. The first row shows the three components computed by the full Hessian kernel expression in Appendix A using the shifted adjoint source with Q_μ perturbation set from $Q_\mu = 125$ to $Q_\mu = 60$ within the two boxes (Fig. 1c). The second row of Fig. 8 shows the same as the first row, but computed using eq. (21) using the shifted adjoint wavefield, where the second term of eq. (21) sums the kernels shown in Figs 8(e) and (f). These strains and perturbations are simultaneously computed to reduce disk space and I/O requirements based on the QuadSEM-Q implementation, which is a combination of the on-the-fly approach (Xie *et al.* 2021) incorporated with the parsimonious storage method (Komatitsch *et al.* 2016). The components of the full Hessian kernels are about one order of magnitude smaller than the Fréchet kernels. The size of the difference is related to the magnitude of the perturbation. For more details about these components and their physical meaning, we refer the reader to previous papers (Fichtner & Trampert 2011; Xie *et al.* 2021).

A comparison of the full Hessian kernels with the finite-difference (FD) approximate Hessian kernels for both the shifted adjoint source and the shifted adjoint wavefield is shown in Fig. 9. The similarity in the shape and magnitude of the full Hessian kernel calculations and the FD calculations verifies the Hessian approach. The positive kernel lobes are nearly the same, with only a small difference in amplitude between the two (Figs 9a vs. c or b vs. d). The magnitudes of the negative portions of the kernels within the box anomaly are similar with some small differences in localized areas. The exact size of the differences depends on the size of the Q anomaly and step size of the FD. The

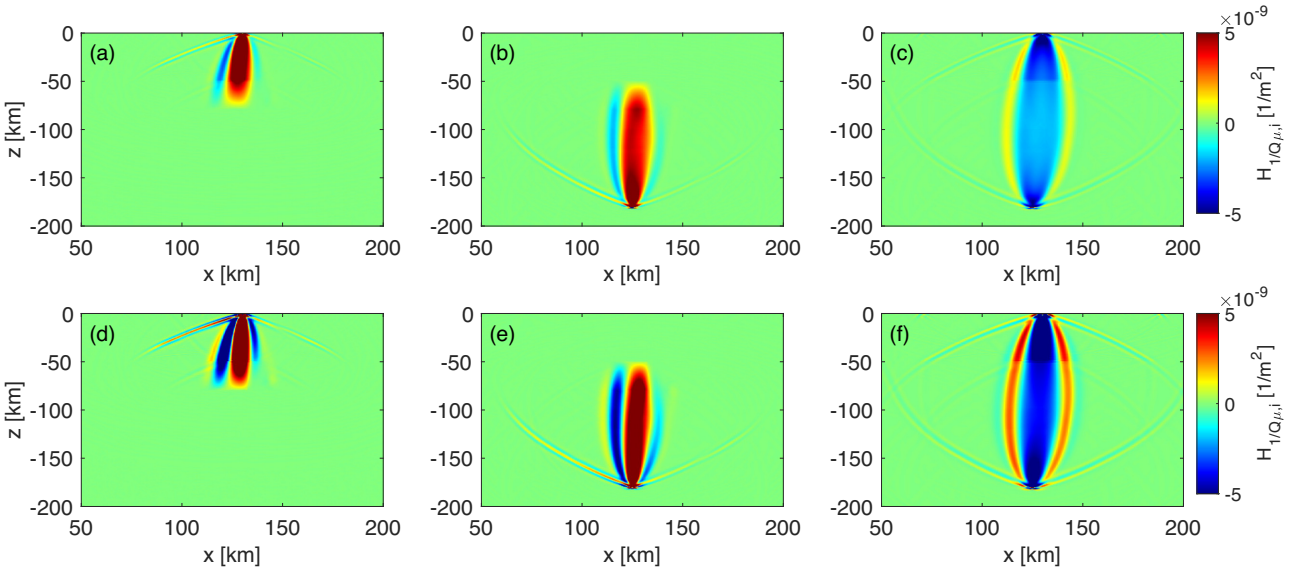

Figure 8. Three components of the Q full Hessian kernels. (a) The first component that is owing to the perturbation of the forward field. (b) The second component that is owing to the perturbation of the adjoint field because of the model perturbation in the adjoint simulation. (c) The last component, which is owing to the perturbation of the adjoint source because of the model perturbation in the forward simulation. The first row is computed using eq. (A3) of Appendix A, which uses the shifted adjoint source. The second row is same as the first row but computed by eq. (21) using the shifted wavefield. Each row shares the same colour bar, and the subscript 'i' indicates the Hessian kernel component.

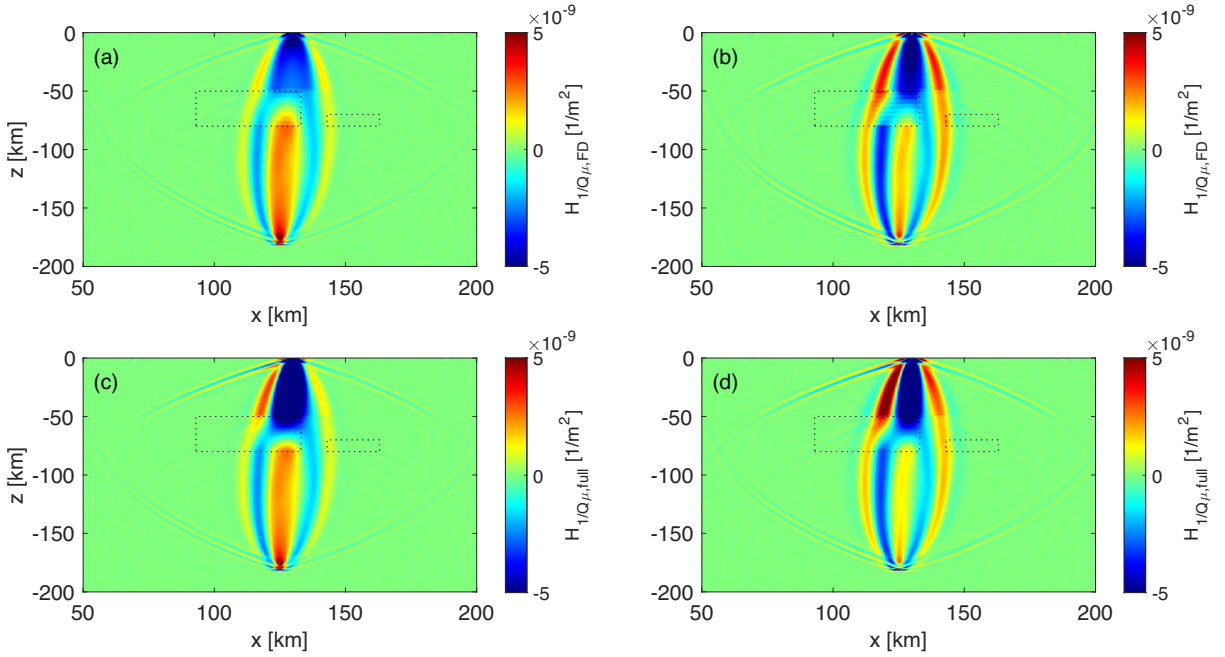

Figure 9. A comparison of the approximate Hessian kernels computed by the finite-difference (FD) approximation with the full Hessian kernels computed here. The Fréchet kernels used in (a) are computed by the kernel expression of eq. (5) using the shifted adjoint source. The Fréchet kernels in (b) are computed by the kernel expression of eq. (14) using the shifted adjoint wavefield. The full Hessian kernels in (c) are computed by the elastic Hessian kernel formulation (Appendix A) using the attenuated forward and adjoint fields, in which the adjoint field is computed by the shifted adjoint source. The full Hessian kernels in (d) are computed by eq. (21) using the shifted adjoint wavefield. The dash box areas indicate the location of the perturbation that the Hessian kernel will use for the model update.

kernels calculated using the shifted adjoint wavefield (Figs 9b and d) are larger than those calculated using the shifted adjoint source (Figs 9a and c). This is because of the greater accuracy of the shifted adjoint wavefield kernels.

5.1.3 Search direction, inverted elastic and Q_μ models

Fig. 10 shows the gradient descent direction **p** computed from the Fréchet and Hessian kernels for the first iteration of the inversion. The

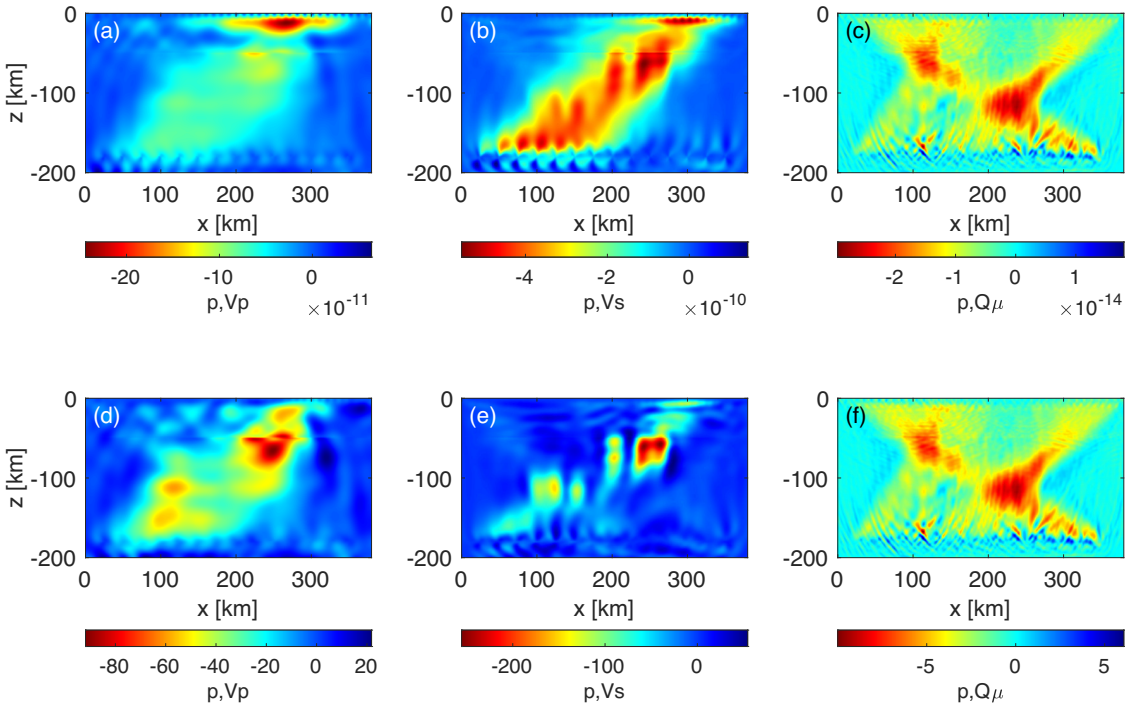

Figure 10. A comparison of the model update direction **p** computed from the Fréchet kernels (L-BFGS, top row), and from the Fréchet and full Hessian kernels (bottom row). They are obtained from the first iteration of the inversion. Note the different colour scale of the top and bottom rows.

full Hessian kernel shows an improved search direction in comparison to L-BFGS at the first iteration . The Q update in the first iteration is large, but it is updated in very small increments in subsequent iterations, compared to the elastic update (see the inversion result below). The computation of the Fréchet kernels in terms of Vp and Vs is the same as that of Tromp *et al.* (2005). For the computation of full Hessian kernels in terms of Vp and Vs, one can refer to the work of Fichtner & Trampert (2011) and Xie *et al.* (2021). The Fréchet and full Hessian kernels in terms of Q_μ^{-1} are computed by eqs (14) and (21) for each event. We do not include the frequency dependence of Q, that is the power-law exponent α, in the computation of the Fréchet and Hessian kernels since in reality this value would not likely be resolvable in the Earth (see e.g. Fichtner & van Driel 2014).

Fig. 11 shows the FWI recoveries and Fig. 12 shows the difference between the inversion recoveries and the original models. Figs 11(a)–(c) shows the L-BFGS inversion recovery of Vp, Vs and ρ for the second frequency stage (0.1–0.2 Hz) after 12 iterations. The Vs model is constructed reasonably well based on the L-BFGS optimization, but the Vp model is poorly recovered, in particular, for the bottom rectangle scatterer (Fig. 11a). Figs 11(d)–(f) shows the Hessian kernel inversion for the first frequency stage of 0.01–0.1 Hz, after 6 iterations, which has better structural model recovery in comparison to the L-BFGS inversion for the second frequency stage (0.1–0.2 Hz) after 12 iterations (Figs 11a–c). Figs 11(g)–(i) shows the Hessian kernel inversion after 2 iterations for the frequency stage of 0.01–0.2 Hz, which has the same misfit reduction as that of the L-BFGS inversion shown in the first row of Fig. 11 (intersections with the black line of Fig. 13b, bottom). The Hessian shows better structural recovery than the L-BFGS for this example. Figs 11(j)–(l) shows the Hessian kernel inversion after 4 iterations for the frequency stage of 0.01–0.2 Hz. It takes the same computational cost as the L-BFGS inversion (Figs 11 a–c) but has a lower misfit (Fig. 13b, bottom). The same phenomena is observed when 50 per cent coherent Gaussian noise is added to each trace of the simulated true data (Figs 14 and 15). The test demonstrates that the elastic inversion either with or without using the full Hessian kernels can still recover the structure of this model even with up to 50 per cent coherent Gaussian noise.

The Hessian kernel inversion increases the convergence rate. For each specific iteration (i.e. the external iteration of the Newton-CG algorithm), the Hessian kernel inversion via the Newton-CG algorithm is more expensive than the gradient-based L-BFGS algorithm. This is because the inner CG iteration requires the computation of the full Hessian kernel where additional forward and adjoint simulation are required. In this example, each iteration of the L-BFGS inversion takes about 8 simulations including 3 simulations for computing the wavefields and Fréchet kernels and about another 5 simulations to compute the step length in the line search. The Hessian kernel inversion requires about 18 simulations since 3 inner CG iterations are used for this example, where each CG iteration takes 6 simulations to compute one full Hessian kernel. The number of simulations required also depends on the implementation strategies related to the way in which the line search is used in either the L-BFGS or the Newton-CG algorithm. In practice, a balance of the number of inner and external iterations in the Newton-CG algorithm may be required. We use 3 inner CG iterations, similar to the one previous FWI example (Métivier *et al.* 2014). However, this also means that more external iterations are needed to solve the Newton equation. The use of more inner CG iterations (N) will increase the computational cost of full Hessian kernel calculation by the factor of $6N$, but the number of external iterations can be reduced due to the increased number of inner CG iterations to improve the Newton direction. The convergence rate also depends on how

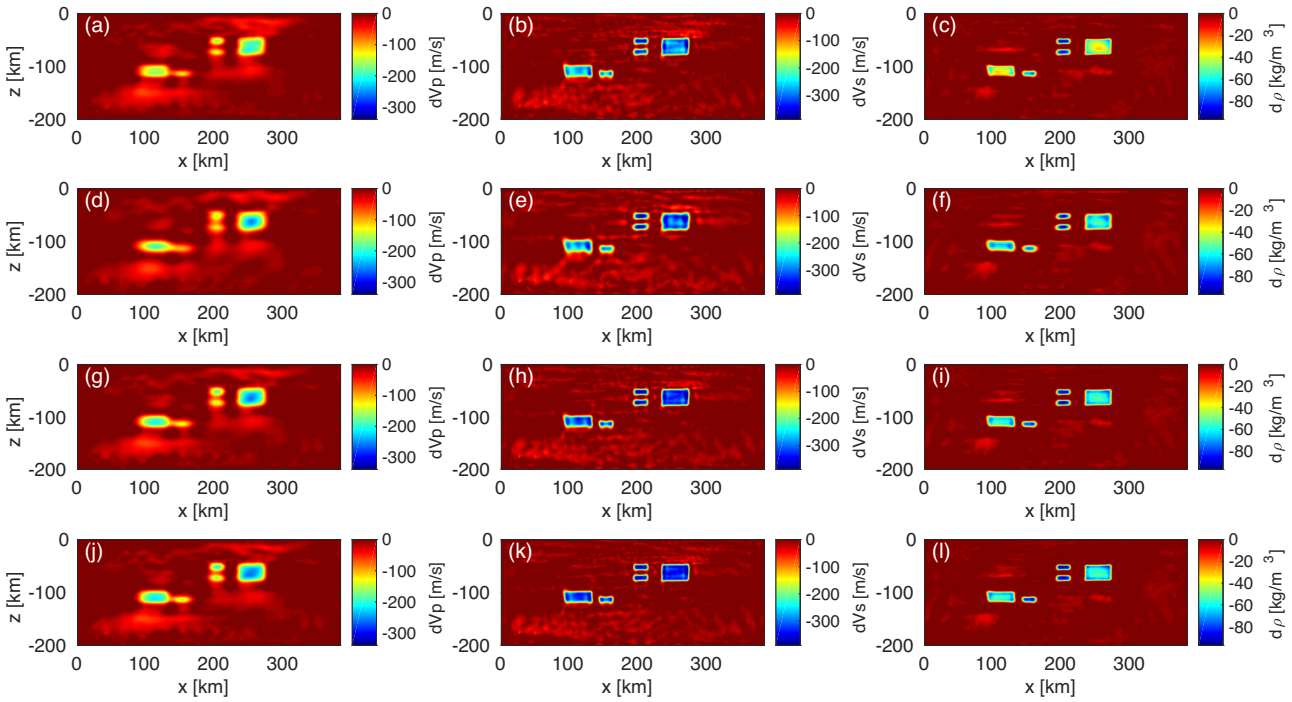

Figure 11. A comparison of full-waveform inversions using gradients only (L-BFGS) and using both gradients and full Hessian kernels (Hessian). The first row (a)–(c) shows the L-BFGS inversion for the frequency from 0.01 to 0.2 H after 12 iterations, where the first frequency stage was 0.01–0.1 Hz with 12 iterations. Each subfigure shows the relative model variation in terms of velocity and density with respective to the initial model. The second row (d)–(f) shows the Hessian inversion for the frequency stage from 0.01 to 0.1 Hz after 6 iterations, which has the same computational cost as the first stage of the L-BFGS inversion after 12 iterations. The third row (g)–(i) show the Hessian inversion for the frequency stage from 0.01 to 0.2 Hz after 2 iterations, which has the same misfit as the L-BFGS inversion of the two stages (see Fig. 13). The bottom row (j)–(l) shows the Hessian inversion recovery for the frequency stage from 0.01 to 0.2 Hz at the 4th iteration, which has the same computational cost as used by the L-BFGS inversion after the second requency stage shown in the first row.

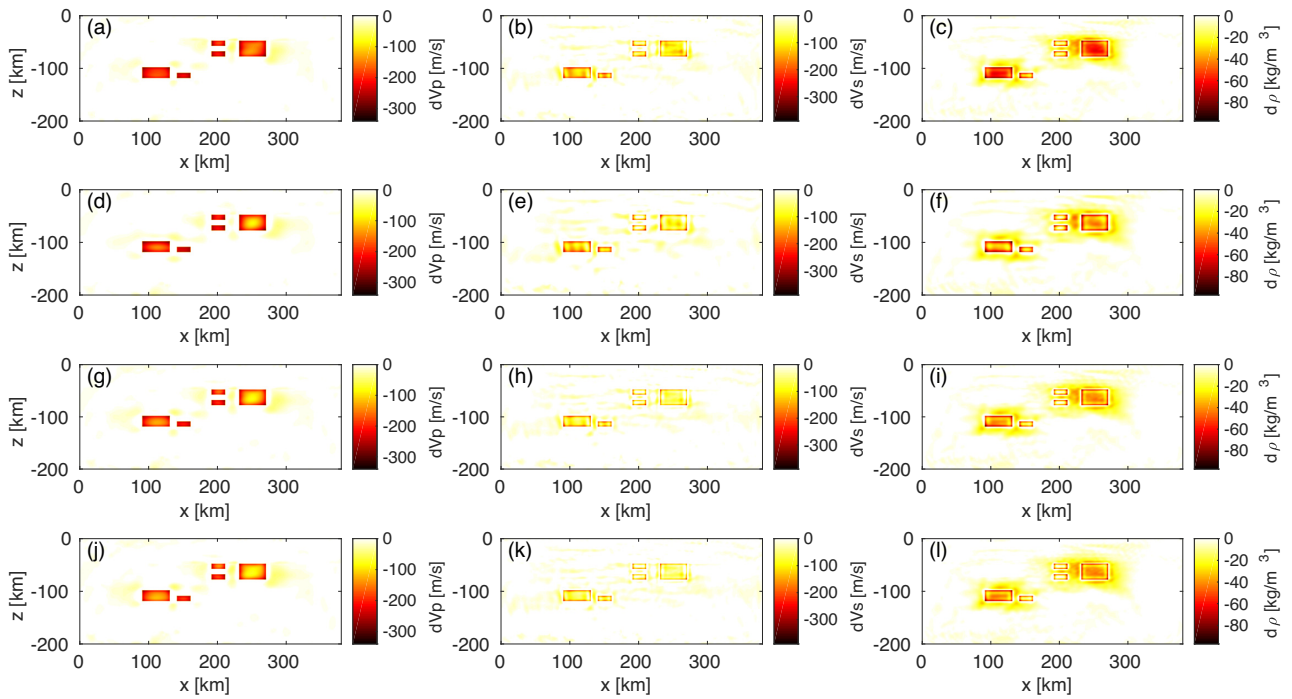

Figure 12. A comparison of the true model with the FWI-recovered models shown in Fig 11. The first row (a)–(c) shows the model difference for the L-BFGS inversion for the frequency stage of 0.01–0.2 Hz after 12 iterations. The second row (d)–(f) shows the model difference for the Hessian inversion for the frequency range of 0.01–0.1 Hz after 6 iterations. The third row (g)–(i) shows the model difference for the Hessian inversion for the frequency range of 0.01 to 0.2 Hz after 2 iterations (same misfit as first row). The bottom row (j)–(l) shows the model difference for the Hessian inversion for the frequency range of 0.01–0.2 Hz after 4 iterations (same computational cost as first row).

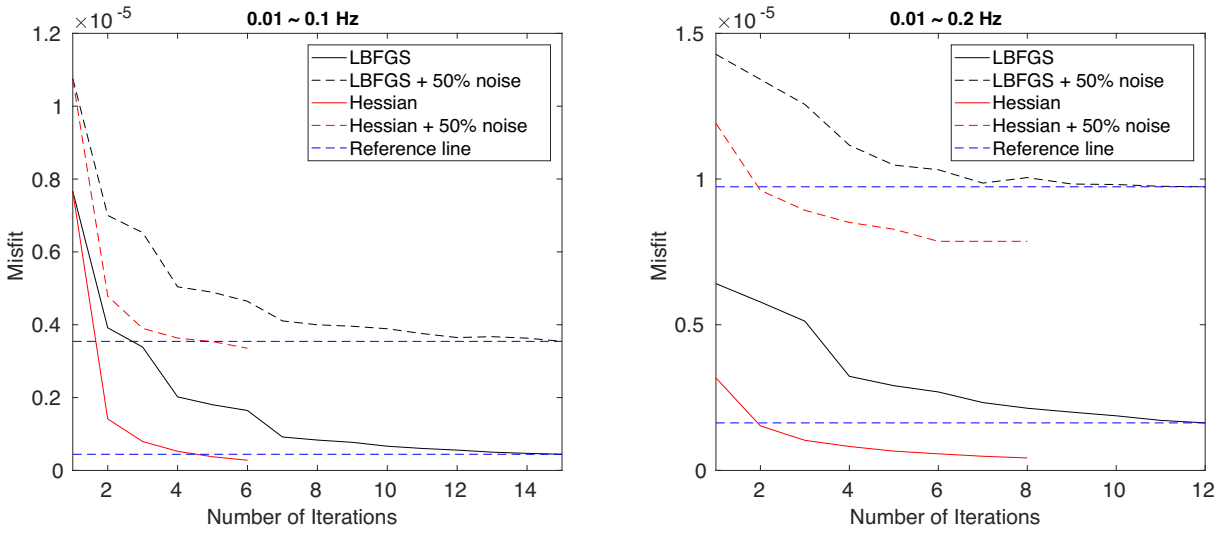

Figure 13. Waveform misfit versus iterations for the two methods both with and without 50 per cent coherent Gaussian noise contamination randomly added to each trace. Left-hand panel: a comparison between L-BFGS and full Hessian at the frequency stage of 0.01–0.1 Hz . The 'LBFGS' indicates the classical L-BFGS inversion. Right-hand panel: the same as left but for the frequency stage of 0.01–0.2 Hz. The reference lines illuminate iterations where the two methods have the same misfit.

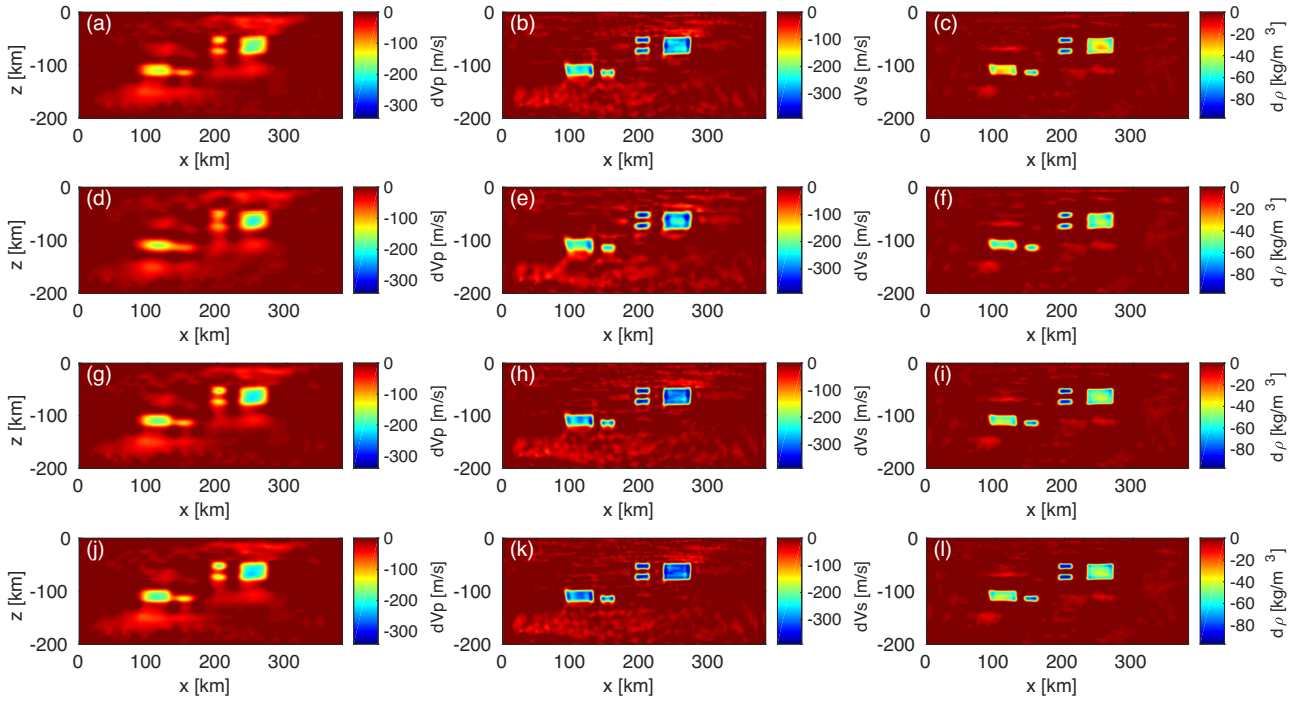

Figure 14. The same as Fig 11 but with 50 per cent coherent Gaussian noise randomly added to each trace of the synthetic data that are generated from the true model.

each iteration is implemented. There is a trade-off between the number of inner and external iterations in practical usage. Regardless, we are most interested in the total computational cost and best convergence. In our implementation, the Newton-CG algorithm we used degenerates to be approximately equal to the non-linear conjugate gradient method when the number of inner CG N set to 1. The classical Newton-CG algorithm returns a negative gradient if there is a negative curvature. We avoid negative gradients by using the non-linear conjugate gradient method for one step movement of the search direction. For positive curvature cases, the classical newton-CG algorithm is used. Future work will focus on the use of negative curvature information for the search direction update.

We calculate the relative computational cost in this work based on the number of simulations instead of the absolute clock time since each run could be affected by the node I/O due to multiple users on the clusters, where pending job submissions are difficult to account for. Fig. 13 shows the convergence rate curve between the two inversions for this example. The L-BFGS inversion takes a total of 96 (8×12) simulations while the Hessian kernel inversion takes 90 (18×5) simulations. Therefore, the full Hessian inversion provides an improved

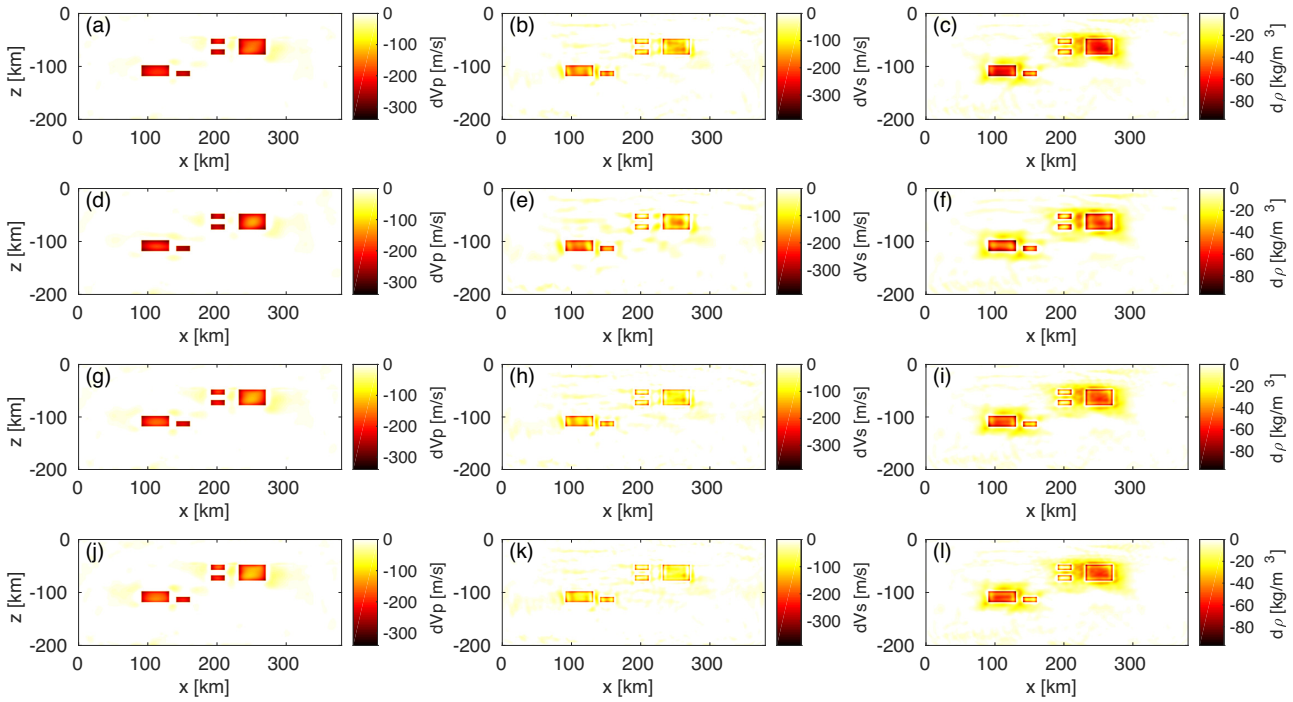

Figure 15. The same as Fig 12 but with 50 per cent coherent Gaussian noise randomly added to each trace of the data generated from the true model.

structural recovery over the L-BFGS inversion with a smaller computational cost for this example. In other words, the use of full Hessian kernel improves the convergence rate, regardless of the high computational cost of each iteration. For the comparison here we use a dense grid for all frequency stages. However, since the Hessian converges closer to true model at an earlier frequency stage than L-BFGS, additional computational savings could be achieved, given that increasingly dense grids are required for higher frequency stages. In FWI multiple stages of inversion are used, typically, with an initial low frequency stage to get the broad scale structure. Then subsequent stages incorporating higher frequency information are used to resolve finer scale structure. In spectral element methods, the computational costs increase exponentially ($f^4 - f^5$) with increased frequency ranges.

This conclusion only applies to this model. However, we hypothesize that the use of full Hessian kernel is likely similarly beneficial for other models with small-scale scatterers or strong heterogeneity, since the full Hessian kernel aids in accounting for second-order scattering wavefields (Fichtner & Trampert 2011).

Fig. 16 shows the simultaneous elastic and Q inversion for the idealized model recovered by the L-BFGS inversion and the full Hessian kernel inversion based on the Newton-CG algorithm. The third iteration of the full Hessian kernel inversion and the seventh iteration of the L-BFGS inversion are compared as these correspond to similar computational costs. The L-BFGS and full Hessian kernel inversions generally recover the magnitude of the Q perturbations, achieving Q values as low as the input value of 30. However, the shape of the Q anomalies are less well-constrained compared to the elastic anomalies. The full Hessian kernel inversion shows similar patterns to the L-BFGS inversion, both characterized by smearing. This agrees with previous FWI work that previously reported the difficulty in recovering Q anomalies (e.g. Askan *et al.* 2007; Fabien-Ouellet *et al.* 2017; Pan & Wang 2020; Yang *et al.* 2020) in comparison to recovering elastic anomalies. The full Hessian kernel inversion shows slightly lower misfit values in comparison to L-BFGS because the result of the elastic full Hessian kernel inversion is used as the start model, which is more accurate than the result of the elastic L-BFGS inversion as the start model (Fig. 17). Q recovery would likely improve at higher frequencies. However, here we have only used frequencies up to 0.5 Hz owing to the high computational cost of using higher frequencies. Doubling the investigated frequency f will exponentially increase the computational cost (at least f^3, e.g. doubling the mesh in multiple directions and half the time step) in the spectral-element method inversion.

5.2 2-D example II: Nicaragua model

Fig. 18 shows the results for the L-BFGS and full Hessian inversions (on-the-fly QuadSEM) for the elastic parameters for 0.01–0.2 Hz and 15 iterations. The starting model is a smooth version of the true model (Figs 18a and d), similar to what might be the expected output from classical tomography approaches (Syracuse *et al.* 2008; Harmon *et al.* 2013). The inversion result for the L-BFGS has good agreement with the true model (Figs 18b,e versus Figs 2a,b) for the *Vp* and *Vs* models. The edges of the slab are well defined, although the deeper parts of the slab retain vestiges of the smooth starting model. The cold nose anomaly is not visible in the *Vp* model but is visible and close to the true value in the *Vs* model. The arc and backarc anomalies are well recovered by the L-BFGS, with the sharp edges recovered well. In the recovered L-BFGS model there are artefacts visible due to smearing into the upper plate and the surrounding mantle wedge that are visible.

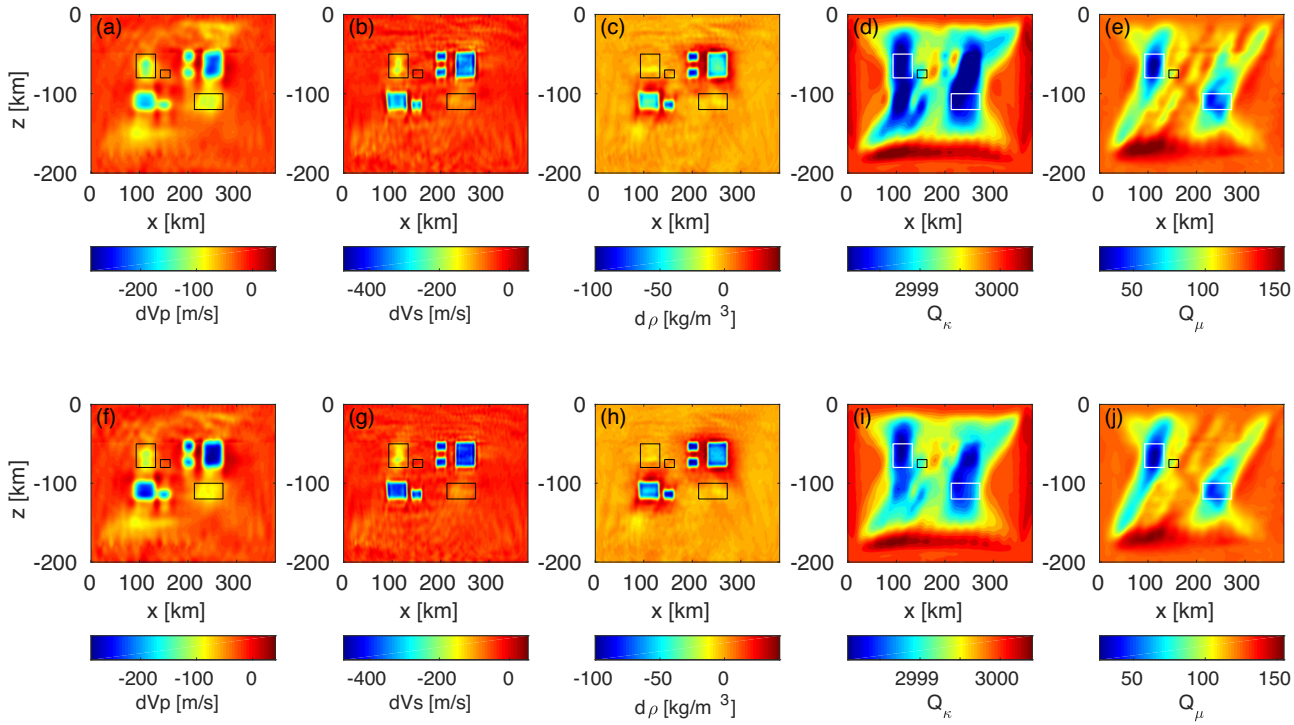

Figure 16. Simultaneous inversion of the elastic parameters (*Vp*, *Vs* and ρ) and the two attenuation quality factors Q_κ and Q_μ using the L-BFGS inversion (top row) and the full Hessian kernel inversion (bottom row). The black frames indicate the location of the attenuation anomalies. For visualization purposes, relative elastic parameters and the absolute Q parameters are shown. The Q parameters trade-off with the elastic parameters. Less trade-off issues are observed for the case in which we compute the elastic and Q kernels simultaneously while using the updated kernel formulation. The Q recoveries are less concentrated around the anomaly locations in comparison to the elastic parameter recoveries due to the different sensitivity of the parameters at this frequency stages.

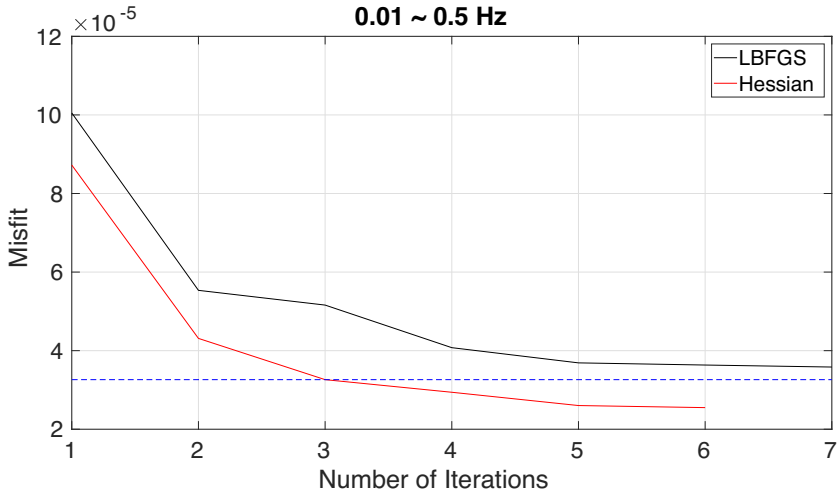

Figure 17. Misfit curves of the anelastic inversion of Fig 16. The initial misfit difference between the two curves is because of the initial input model. For the inversion with full Hessian kernel, the starting model is the result of the elastic full Hessian kernel inversion. We use the third iteration of the full Hessian kernel inversion which has a similar misfit to the L-BFGS.

The full Hessian kernel recovery also has good agreement with the true model. The slab is well recovered, with sharper edges visible, and the slab anomaly is also recovered better at greater depths than recovered by L-BFGS. The cold nose anomaly is visible in both the recovered *Vp* and *Vs* models. The arc and backarc anomalies are also well recovered both in terms of their shapes and magnitudes. The full Hessian kernel inversion also has artefacts visible in the upper plate and underlying mantle. There are some subtle differences between the L-BFGS and the full Hessian kernel inversion. The full Hessian kernel inversion does better at recovering some of the finer details such as the cold nose and deeper and shallower parts of the slab (Fig. 19). In general both methods do better at recovering the *Vs* structure than the *Vp* structure. Similar convergence behaviour is observed for this example to what was observed in the ideal source–receiver geometry examples, with the full Hessian kernel inversion converging faster than the L-BFGS (Fig. 20).

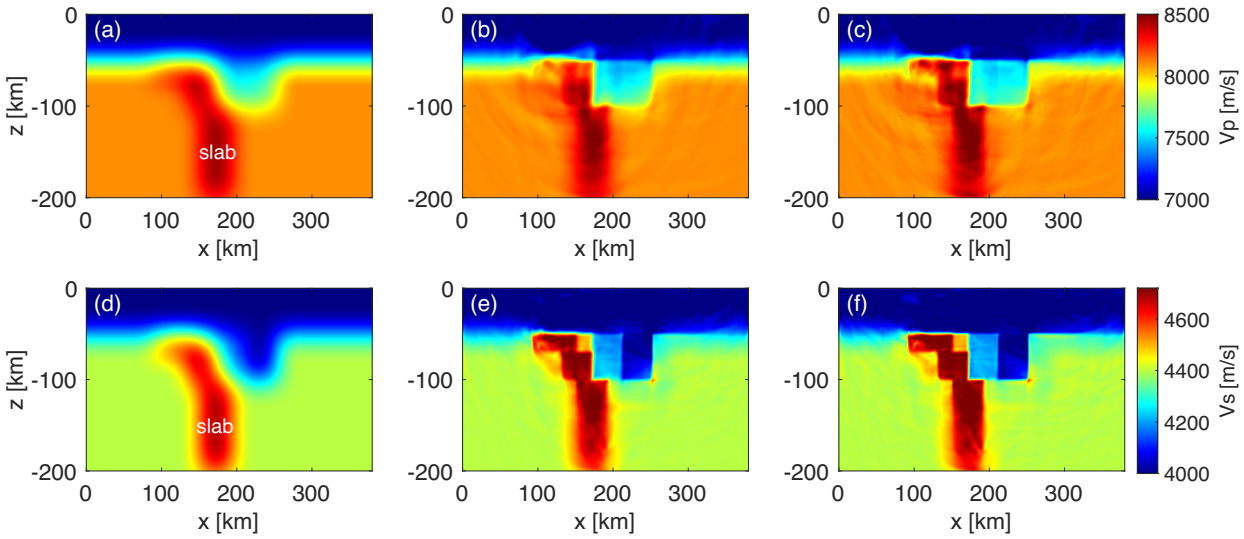

Figure 18. Elastic full-waveform inversion for the Nicaragua subduction models for the frequency range of 0.01–0.2 Hz. The top row shows the inverted *Vp* models and the bottom row shows the inverted *Vs* models. The left-hand column (a and d) shows the smooth start models. The second column (b and e) shows recovery using the classical L-BFGS inversion, and the last column (c and f) shows recovery using full Hessian kernel inversion.

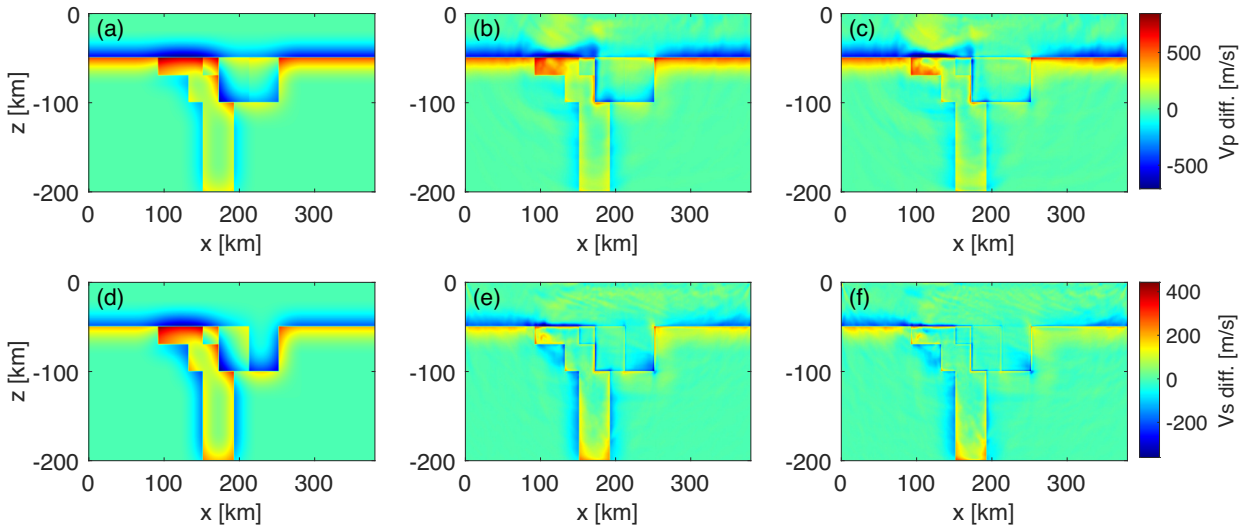

Figure 19. Differences between the recovered models and the true models shown in Figs 2 and 18. Panel (a) shows the difference between the true model and the smoothed *Vp* model. Panel (b) shows the difference between the true *Vp* model and the L-BFGS inversion recovery. Panel (c) shows the difference between the true *Vp* model and the Hessian inversion recovery. Panels (d)–(f) are the same as (a)–(c) but for the *Vs* model differences. The use of full Hessian kernels improves the recovery of the edges.

The inversion recoveries for the anelastic parameters using L-BFGS and the full Hessian kernel are shown in Fig. 21. The magnitudes of the subarc and backarc Q_μ anomalies (40 and 30, respectively) are recovered by L-BFGS and by the full Hessian, but with a range between 10 and 110. The Q_μ magnitude of the upper plate (1000) is recovered by L-BFGS and the full Hessian as 900–1000. The shapes of the subarc anomalies are not well-recovered, in particular the deepest corner towards the backarc is not recovered likely owing to a lack of ray paths. The sharp boundary between the upper plate Q and the subarc anomalies is well recovered by L-BFGS and the full Hessian. The slab is not very well recovered by either L-BFGS or the full Hessian. The cold nose Q_μ anomaly is not well-recovered by L-BFGS or the full Hessian. The anelastic recovery may be imperfect for several reasons. For example, the high Q values in the slab may be difficult for the inversion to recover because the difference between the Q of 1000 in the slab relative to the 600 in the background might not produce enough variation in the amplitude of the waveform to be recoverable. The low Q structures are also poorly resolved, although the inversions do recover the anomalies in a broad sense, albeit with smearing. We speculate that better recovery of the anelastic properties could be achieved by including higher frequencies. For instance, regional subduction zone studies sometimes consider frequencies as high as 15–20 Hz (Rychert *et al.* 2008). Yet, achieving such high frequencies is still very computationally expensive in full waveform modelling. This hypothesis is supported by FWI studies that used approximate Hessian kernels and recovered Q anomaly shapes and sharp edges by using very high frequencies, for example

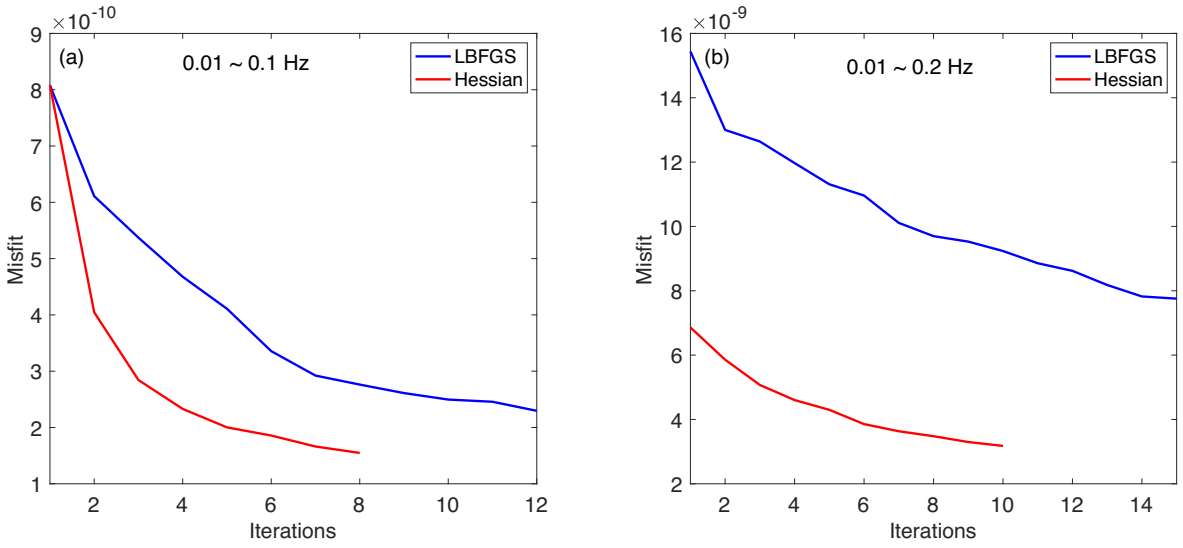

Figure 20. A comparison of the misfit convergence rates between the L-BFGS and the Hessian kernel inversions. The latter improves the convergence rate substantially (a). Panel (b) shows the misfit for the frequency stage of 0.01–0.2 Hz, where the misfit of the Hessian kernel inversion is consistently lower than that of the L-BFGS inversion, including the first iteration.

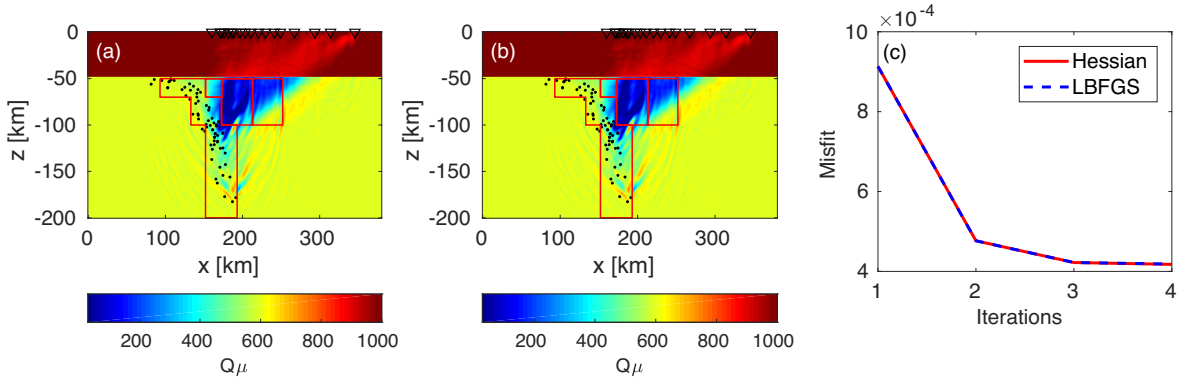

Figure 21. A comparison of the L-BFGS inversion with the full Hessian kernel inversion for the Q_μ model. Panel (a) shows the Q_μ recovered using L-BFGS inversion. Panel (b) shows the Q_μ model recovered by using the full Hessian kernel. Panel (c) shows the convergence of the Hessian (red) and L-BFGS (blue dashed). The misfit of the two inversions is equivalent for this model after using 10 Q_μ perturbations for the full Hessian calculation. The recovered structures are also very similar. The black dots indicate the earthquake locations and the black inverted triangles indicate the stations.

up to 30 or 50 Hz (e.g. Fabien-Ouellet *et al.* 2017; Pan & Wang 2020; Yang *et al.* 2020). However, the size of models that were interrogated were also very small, sometimes smaller than 1 km by 1 km, given the computational costs.

5.3 3-D example: the 3-D scatterer model

5.3.1 3-D elastic inversion

Fig 22 shows the 3-D elastic L-BFGS FWI and the full Hessian kernel inversion with kernels calculated computed on the fly with the QuadSEM3D package for the frequency stage up to 0.1 Hz at the 12th iteration. We find that the full Hessian kernel inversion and the L-BFGS FWI perform similarly for the first frequency stage, up to 0.1 Hz. The *Vp* magnitudes are updated to values that range from –15 to 10 m s^{-1} relative to the background models, i.e., much less than the true anomaly perturbation from the background model of –340 m s^{-1}. The *Vs* magnitudes are better recovered, with values ranging from –400 to 10 m s^{-1}, encapsulating the true model perturbation from the background model of –390 m s^{-1}. Additional iterations or higher frequency stages may need to be performed to image the edges of the scatterers for the *Vp* and *Vs* models. The densities magnitudes are updated by –100 to 50 kg m^{-3}, which encapsulates the true density perturbation of –95.7 to 0 kg m^{-3}. However, the shapes of the anomalies are not very well-recovered, particularly for *Vp* and density. For the investigated frequency stage, up to 0.1 Hz, the full Hessian kernel inversion is about 2.25 times more computationally expensive than the L-BFGS FWI.

Fig 23 shows that the *Vp* and *Vs* model recovery for the frequency stage of 0.2 Hz is better than that of the first frequency stage of 0.1 Hz (Fig 22). However, the recovered density from the frequency stage of 0.2 Hz is worse than that of the frequency stage of 0.1 Hz. Compared

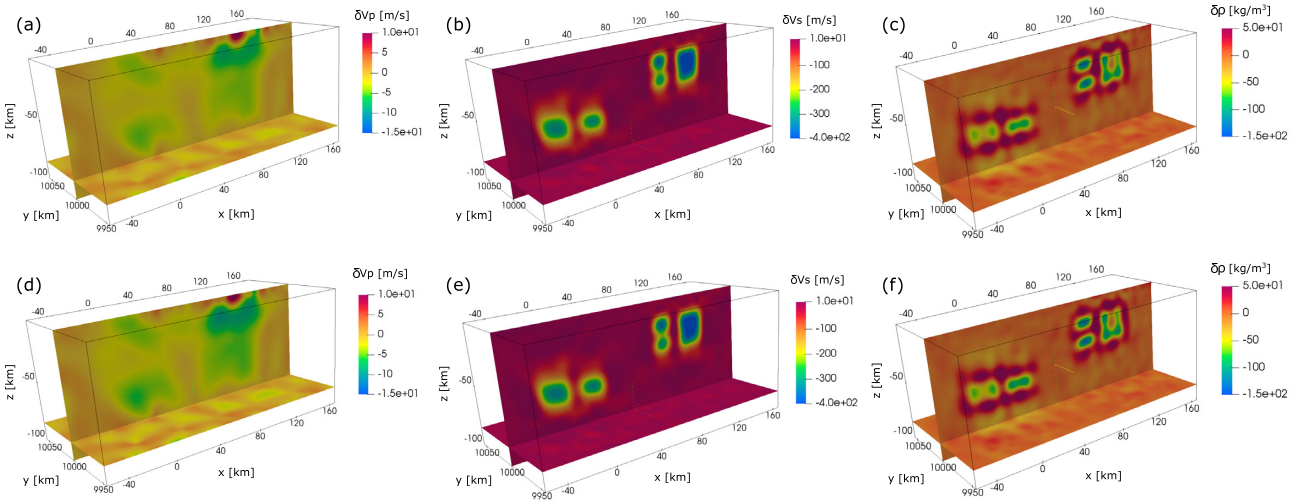

Figure 22. A comparison of the 3-D L-BFGS inversion with the 3-D full Hessian kernel inversion for the frequency range from 0.01 to 0.1 Hz for the 3-D elastic model. Panels (a)–(c) show the L-BFGS inversion for compressional wave speed, Vp [m s^{-1}], shear wave speed, Vs [m s^{-1}] and density, ρ [kg m^{-3}], respectively. The second row shows the corresponding inversion results using the full Hessian kernels. We show relative model perturbation to the background model for visualization purposes.

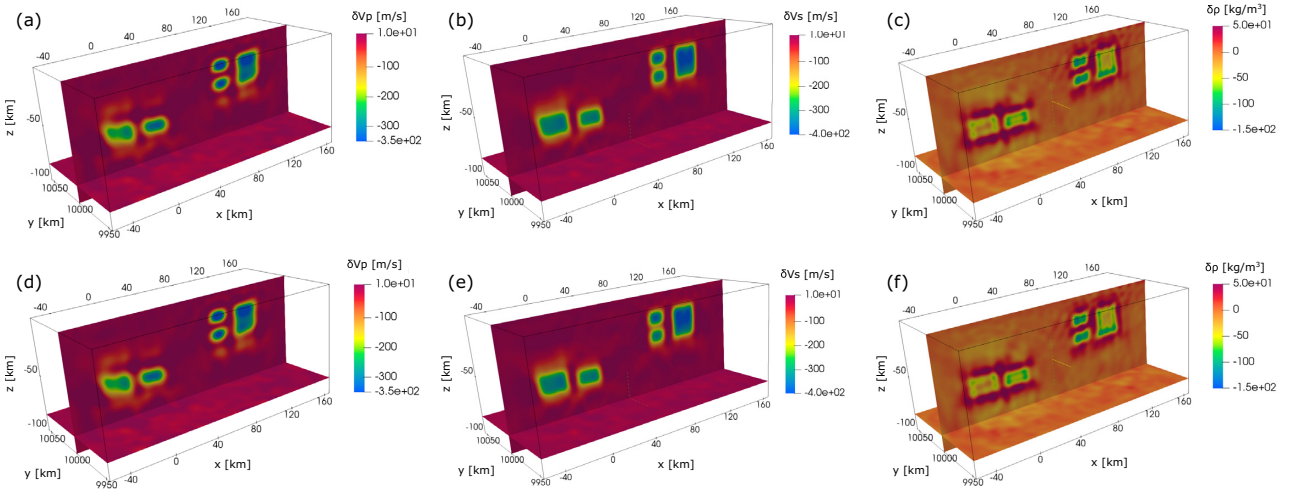

Figure 23. The same as Fig 22 but for the frequency range from 0.01 to 0.2 Hz.

to the Vs model recovery, the Vp model recovery is still less well-imaged due to the lower sensitivity of the small P-waves in comparison to that of the larger S-wave phases. We then perform one additional P-wave inversion to recover the Vp model while using the inversion at the stage of 0.2 Hz as an initial input. The left-hand column of Fig 23 shows the recovered Vp model after the P phase inversion. The decreased recovery of density structure at higher frequency likely reflects the interparameter trade-off with Vp .

Fig 24 shows the inversion difference to the true model, and we find that the Hessian kernel inversion recovers the Vp model better than the L-BFGS inversion, although they show similar misfit values (Fig 25). For the frequency stage of 0.2 Hz, the full Hessian kernel inversion shows the same level of misfit reduction at the 9th iteration as the L-BFGS inversion shows at the 13th iteration (Fig 25). The full Hessian kernel inversion is about 1.55 times expensive than the L-BFGS inversion at this point. This is an improvement over the 2.25-fold computational cost at the end of the first frequency stage. Given this and the improved convergence rates that we observe at higher frequencies, we speculate that the full Hessian kernel may show greater convergence properties at higher frequencies than we were able to test here. The full Hessian kernel inversion may be computationally advantageous after a good low-frequency model has been determined using L-BFGS.

For the 3-D FWI using the full Hessian kernels, we use the classical Newton-CG algorithm (Nocedal & Wright 2006, p. 112 and p. 119). The Newton-CG inversion also depends on variations in the Newton-CG algorithm version that is used. In this algorithm the updated CG direction, $\mathbf{p}$, is normalized to a fixed scalar value for the Hessian kernel calculation. This algorithm works well for 2-D inversions, as evidenced by the testing performed here. However, we find that this approach fails in the 3-D tests as the product of the $\mathbf{p}$ with the Hessian vector product $\mathbf{Hp}$, that is the $\mathbf{p}^{\mathrm{T}}\mathbf{Hp}$, can be negative around the initial model or around one iteration of the inverted model during the inversion. If negative curvature is detected at the beginning, the classical Newton-CG algorithm will return the negative gradient as the Newton direction. To partially deal with this issue, we use a line search to find the optimal $\mathbf{p}$ value for the model update. We store a previous search direction after

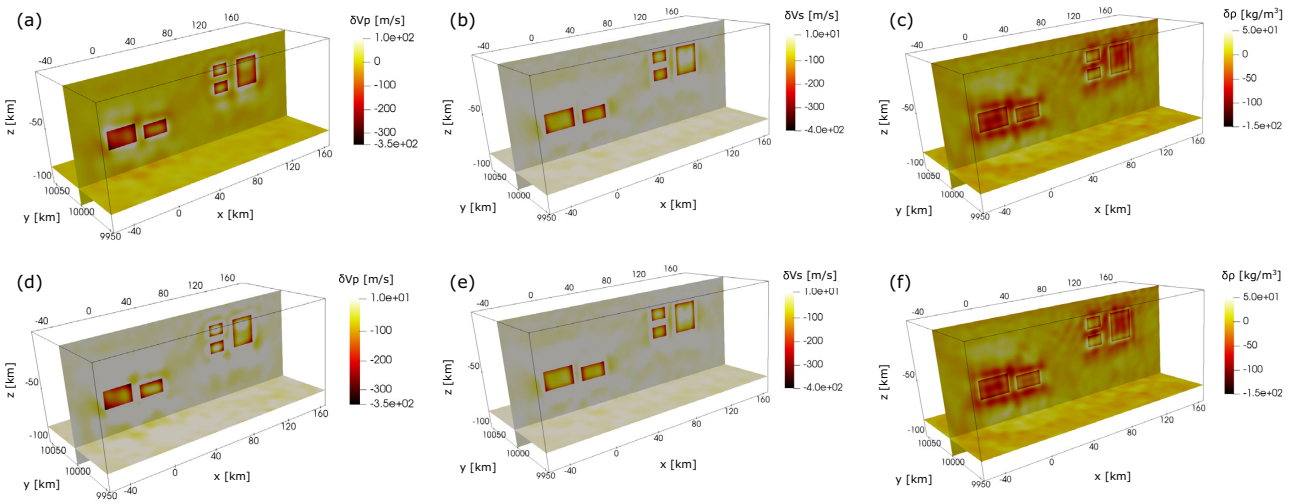

Figure 24. A comparison of the 3-D inversions to the true models for the investigated frequency range from 0.01 to 0.2 Hz. Panels (a)–(c) show the difference of the L-BFGS inversion to the true model for *Vp*, *Vs* and ρ. Panels (d)–(f) show the difference from the true model for the full Hessian kernel inversion. We show relative model perturbation to the background model for visualization purposes.

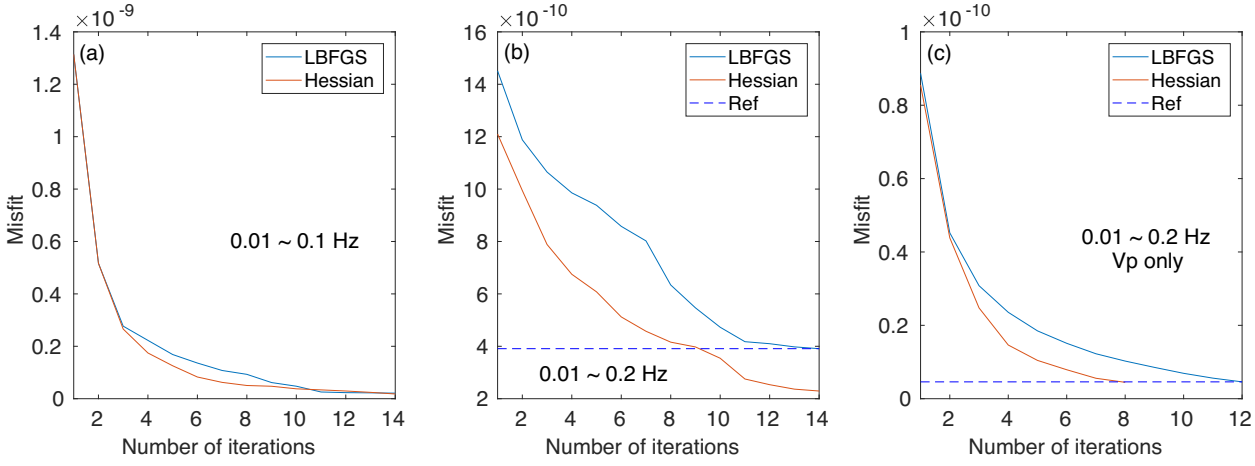

Figure 25. 3-D waveform misfits versus iterations for the frequency range from 0.01 to 0.1 Hz (left-hand panel) and from 0.01 to 0.2 Hz (middle panel) and *Vp* only from 0.01 to 0.2 Hz (right-hand panel). The blue dashed reference line shows the same misfit for the two methods. The full Hessian kernel inversion improves the convergence rate when the maximum investigated frequency is increased from 0.1 to 0.2 Hz.

the first iteration. The previous and current search directions can be used to calculate a new CG direction as well, in case the **p** optimisation does not work. Alternative Newton-CG algorithms or similar algorithms that use the Hessian vector product may be developed in the future for using the negative curvature of the Hessian. For the inversion, we constrain the updated model within physical limits. In other words, we constrain each updated model with a –2000 unit smaller than the minimum of the initial and true model, and a +2000 unit greater than the maximum of the initial and true model.

5.3.2 3-D simultaneous elastic and Q inversion

In the work, we perform a simultaneous inversion of the 3-D elastic and Q models using the improved Q Fréchet kernels rederived in this work. The improvement can save about 1/3 of the computational cost in comparison to the classical anelastic adjoint tomography that uses the shifted adjoint source. Fig 26 shows the simultaneous elastic and Q inversion, with the initial model provided by the 3-D elastic L-BFGS inversion. We observed that the Q amplitude and the anomaly shape are less well-recovered in comparison to the elastic inversion for the frequency investigated, up to 0.5 Hz. Fig 26(c) shows the misfit function of the simultaneous inversion, in which the misfit converges after three iterations. The left upper Q anomaly is recovered due to sufficient S-wave illumination. The right bottom anomaly is not well-recovered since only one event is located at the bottom of the Q anomaly, which is also very close to the event location. The source and receiver geometry is very important in Q inversion, in particular for the S-waves illuminations.

A simultaneous inversion of the elastic and Q models is challenging since Q has a smaller effect on the waveforms than the elastic parameters. Attenuation trades-off with velocity owing to dispersion. The elastic perturbations superimposed upon an anelastic reference

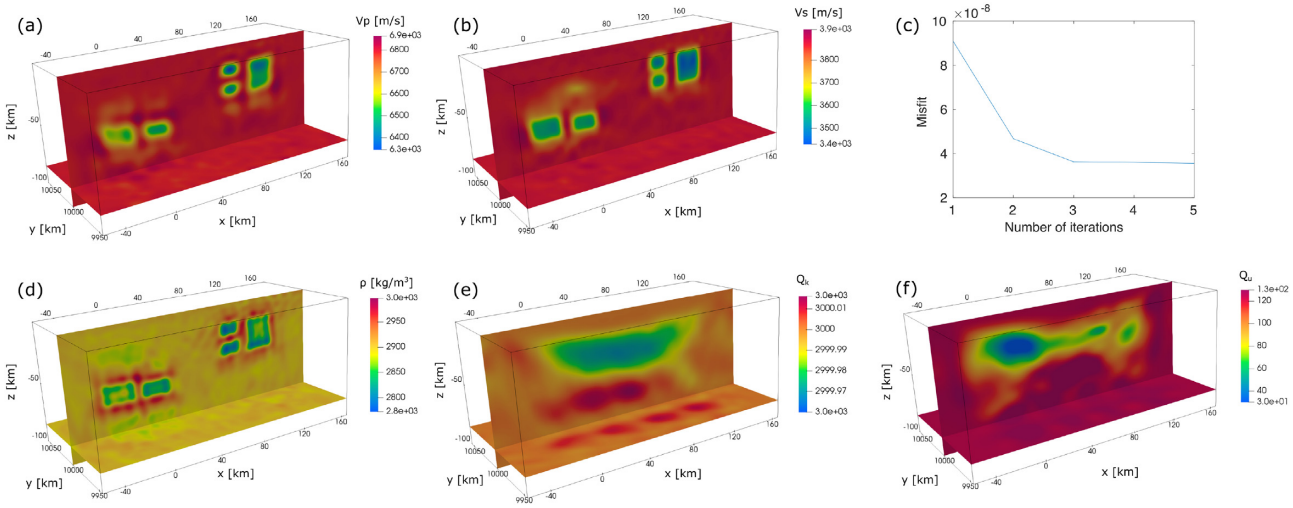

Figure 26. A simultaneous inversion of the five model parameters. Panels (a), (b) and (d) show the inversion for the compressional wave speed, Vp [m/s], shear wave speed, Vs [m s^{-1}] and the density, ρ [kg m^{-3}], respectively. Panel (c) shows the misfit varied with iterations. Panels (e) and (f) show the inversion for quality factors Q_κ and Q_μ.

model may be computed like those of the elastic case provided the required wavefields are calculated in the anelastic reference model (Liu & Tromp 2008). Unlike the elastic inversion where the kernels are computed using the phase and amplitude difference or the full waveform difference, in the simultaneous elastic and Q inversion, we compute the elastic kernels using the phase difference. There are four sets of elastic moduli in the viscoelastic simulation: that is the moduli at the reference and central frequency, and the relaxed and unrelaxed moduli. In the inversion, the initial and true models should be considered under the same moduli state, that is, assuming which moduli of the initial and true models were specified. In the codes, the moduli are initially shifted from the reference frequency to the central frequency, then shifted to the unrelaxed state for stress calculation. In the simultaneous inversion, one may invert the elastic properties first, then perform a simultaneous inversion of the elastic and Q models. Alternatively, one may also perform a simultaneous inversion of the elastic and Q models from the beginning of the inversion. In the latter case, a larger weight (ω_ϕ) may be used to enhance the elastic contribution in comparison to a smaller weight (ω_A) used for the Q. These weights are applied to the misfit, adjoint sources, and possibly the kernels to balance the two contributions. We estimate the initial weights (χ_ϕ and χ_A) by the relative phase and amplitude difference of the misfit when normalized at the same level in magnitude. The weights are then adaptively estimated via the phase and amplitude misfits (χ_ϕ and χ_A) with respect to the elastic and Q models during the inversion. The line search also needs to be adjusted for the elastic and Q inversion. The use of the same step length for the elastic and Q update may fail because of the different magnitudes of the velocity versus the attenuation parameters.

The Q inversion for real data would be very challenging. One reason is that the uncertainty in the elastic model typically remains high even in well-resolved models with good source receiver geometries. The second reason is that the source and receiver geometry, in which the Q inversion may require more ray illumination in comparison to the elastic inversion. This requires more investigations in the future. In this work, we fix all source parameters, which could be fine for active source seismology, since these parameters are known. For passive source seismology, the uncertainty of source parameters would further increase the difficulty of the Q inversion. We perform the simultaneous elastic and Q inversion using the Q kernel rederived in this work for efficiency purposes. As the 3-D elastic and Q inversions are very expensive at the regional scale, we show the inversions only up to 0.5 Hz. The 3-D full Hessian kernel inversion for Q would again double the computational cost, which is computationally prohibitive for most medium-scale high-performance computing platforms at present. The associated inversions require applying for larger computing grants for supporting the computation. Given its very high computational cost, it may not be advisable to use it at this time.

6 CONCLUSIONS

We revisit the Q Fréchet kernels via the perturbation of the modulus in both the frequency and the time domain and show that the frequency-independent Q Fréchet kernels of Fichtner & van Driel (2014) are approximately equivalent to the Q Fréchet kernels of Tromp *et al.* (2005) when derived for $\delta Q/Q$ or δQ^{-1}. The Q Fréchet kernels are computed while computing the regular kernels for the elastic properties, using the forward strain and a shift of the adjoint strain. This is explicit for the displacement-stress formulation that uses the strain without recomputing its time derivative and associated memory variable, as an extension of the work of Fichtner & van Driel (2014). This development reduces the need for an additional simulation in the computation of frequency-independent Q Fréchet kernels and extends it to the computation of frequency-dependent Q Fréchet kernels for the spectral element and adjoint methods. We found that the use of the shifted adjoint source is less accurate than the shifted adjoint field. We then derive the full Hessian kernels for frequency-(in)dependent Q based on the work of Fichtner & Trampert (2011) and Fichtner & van Driel (2014), and compute the frequency-independent Q full Hessian kernels by a combination of the on-the-fly QuadSEM approach (Xie *et al.* 2021) with the checkpointing parsimonious storage method (Komatitsch *et al.* 2016), in which the

two forward fields are simultaneously recomputed from the checkpoints while simultaneously computing two adjoint fields during the adjoint simulation. We call this approach QuadSEM-*Q*.

For the idealized case with many sources and receivers and small anomalies with sharp edges and also for the more realistic Nicaragua subduction zone example, the on-the-fly QuadSEM inversion for 2-D elastic properties demonstrates improved structure and amplitude recovery in comparison to the L-BFGS inversions. For the L-BFGS inversion, we limit our implementation to the publically available Seisflows, and we slightly adjust the internal parameters to obtain optimal performance of the L-BFGS algorithm. We observed that the use of full Hessian kernels improves the convergence rate, and this improvement effectively compensates for the higher computational cost per iteration in comparison to the L-BFGS. This results in smaller computational costs for the same level of misfit for the tested models. The full Hessian kernel also recovers structure at lower frequency than that required for the L-BFGS, which could further reduce the computational costs for the two examples. The 2-D Q_μ anomalies are recovered using both L-BFGS and the full Hessian kernel, albeit with some smearing .

The 3-D elastic full Hessian kernel inversion shows similar misfits as the L-BFGS inversion after the first frequency stage, up to 0.1 Hz. The use of full Hessian kernel increases the convergence rate at the second frequency stage, up to 0.2 Hz, although the computational cost is still 1.5 times that of L-BFGS to achieve the same level of misfit. We speculate that this trend will continue and that at higher frequency stages the full Hessian kernel inversion will be more computationally advantageous than the L-BFGS for models with small scatterers or small-scale heterogeneity as the Hessian kernels are better accounting for second-order scattered fields. The investigations at higher frequencies would be very computational expensive in 3-D. This will be investigated in the future when further computing sources are available. One of the reasons why the performance of the 3-D full Hessian kernel inversion is not as good as the 2-D case may be because the negative curvature of the Hessian in this 3-D case, which requires a development of new algorithms to use the negative curvature information. A simultaneous inversion of the elastic and *Q* models via the spectral-element and adjoint method in the 3-D case is very challenging even for synthetic models due to the small scale of the *Q* relative to the elastic properties. Compared to the classical anelastic adjoint tomography that uses the additionally defined adjoint source, the proposed method using the rederived *Q* kernels reduces the computational cost at a factor of 1/3 as no additional adjoint simulations are performed for the *Q* kernel calculations. The recovery of *Q* in 3-D is similar to the recovery we observed in 2-D, with many artefacts visible for our tests up to 0.5 Hz. Inversions at higher frequencies will exponentially increase the computational cost, which will be investigated in the future.

DATA AND RESOURCES

No field data were used in this work. We use the QuadSEM (https://github.com/yujiangxie/QuadSEM) for 2-D full Hessian kernel calculations for elastic and anelastic properties, which were developed based on the public SPECFEM2D package (https://github.com/geodynamics/specfem2d). The 2-D elastic full Hessian kernels are computed on the fly via the QuadSEM and the 2-D anelastic full Hessian kernels are computed partially on the fly via checkpointing implemented in QuadSEM-Q, as a complementary to the elastic QuadSEM. We use QuadSEM3D (https://github.com/yujiangxie/QuadSEM3D) for 3-D elastic full Hessian kernel calculations, which were developed based on the public SPECFEM3D Cartesian package (https://github.com/geodynamics/specfem3d). The 3-D elastic full Hessian kernels are computed on the fly. The reformulated *Q* Fréchet kernels computed by the forward strain and a shift of the adjoint strain (or the adjoint memory variable) in the 2-D and 3-D cases are initially included in the QuadSEM-Q and QuadSEM3D packages. We use the Seisflows package (https://github.com/rmodrak/seisflows) for the inversion workflow, with adding a few lines for the simultaneous elastic and *Q* inversion. For the full Hessian kernel inversion, we use the Newton-CG algorithms (Nocedal & Wright 2006).

ACKNOWLEDGMENTS

The authors acknowledge funding support from the Natural Environment Research Council (NE/M003507/1), the European Research Council (GA 638665), and the National Science Foundation (NSF-EAR-2147918). We acknowledge the use of the IRIDIS High-Performance Computing Facility, and associated support services at the University of Southampton in the completion of this work. We thank editor Carl Tape and two anonymous reviewers provided constructive comments which improved the paper. We thank Geoff Abers for providing parameter files from Nicaragua.

DATA AVAILABILITY

This manuscript uses no data as described in the data availability statement.

REFERENCES

Abers , G.A., Fischer , K.M., Hirth , G., Wiens, D.A., Plank, T., Holtzman, B.K., McCarthy, C. & Gazel, E., 2014. Reconciling mantle attenuation-temperature relationships from seismology, petrology, and laboratory measurements, *Geochem. Geophys. Geosyst.,* **15,** 3521–3542.

Agius, M.R., Rychert, C.A., Harmon, N., Tharimena, S. & Kendall, J.M., 2021. A thin mantle transition zone beneath the equatorial mid-Atlantic ridge, *Nature,* **589,** 562–566.

Askan, A., Akcelik, V., Bielak, J. & Ghattas, O., 2007. Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures, *Bull. seism. Soc. Am.,* **97,** 1990–2008.

Bamberger, A., Chavent, G., Hemons, C. & Lailly, P., 1982. Inversion of normal incidence seismograms, *Geophysics,* **47,** 757–770.

Bamberger, A., Chavent, G. & Lailly, P., 1977. Une application de la théorie du contrôle à un problème inverse sismique, *Ann. Geophys.,* **33,** 183–200.

Blanc, É., 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, *Geophys. J. Int.,* **205,** 427–439.

Blanch, J., Robertsson, J. & Symes, W., 1995. Modelling of a constant Q: methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique, *Geophysics,* **60,** 176–184.

Boehm, C., Hanzich, M., de la Puente, J. & Fichtner, A., 2016. Wavefield compression for adjoint methods in full-waveform inversion, *Geophysics,* **81**(6), R385–R397.

Boehm, C. & Ulbrich, M., 2015. A semismooth Newton-CG method for constrained parameter identification in seismic tomography, *SIAM J. Sci. Comput.,* **37,** S334–S364.

Bonnans, J.F., Gilbert, J., Lemaréchal, C. & Sagastizábal, C., 2006. *Numerical Optimization, Theoretical and Practical Aspects,* Springer Science & Business Media. .

Bozdağ, E., Peter, D., Lefebvre, M., Komatitsch, D., Tromp, J., Hill, J., Podhorszki, N. & Pugmire, D., 2016. Global adjoint tomography: first-generation model, *Geophys. J. Int.,* **207,** 1739–1766.

Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.,* **185,** 845–870.

Carcione, J.M., 1988. Wave propagation simulation in a linear viscoelastic medium, *Geophys. J.,* **95,** 597–611.

Carcione, J.M., 2001. *Wave Fields in Real Media: Wave Propagation in Anisotropic, Anelastic and Porous Media,* Elsevier.**31,** 1–390.

Chen, P., Zhao, L. & Jordan, T.H., 2007a. Full 3D tomography for the crustal structure of the Los Angeles region, *Bull. seism. Soc. Am.,* **97,** 1094–1120.

Chichester, B., Rychert, C., Harmon, N., van der Lee, S., Frederiksen, A. & Zhang, H.. 2018. Seismic Imaging of the North American Midcontinent Rift Using S-to-P Receiver Functions, *J. Geophys. Res.,* **123,** 7791– 7805. https://doi.org/10.1029/2018JB015771

Cornell, C., 1968. Engineering seismic risk analysis, *Bull. seism. Soc. Am.,* **58,** 1583–1606.

Dahlen, F., Hung, S.H. & Nolet, G., 2000. Fréchet kernels for finite-frequency traveltimes—I. Theory, *Geophys. J. Int.,* **141,** 157–174.

Dalton, C.A., Ekstrom, G. & Dziewonski, A.M., 2008. The global attenuation structure of the upper mantle, *J. geophys. Res.,* **113,** B09303. doi:10.1029/2007JB005429.

Eberhart-Phillips, D., Reyners, M., Upton, P. & Gubbins, D., 2018. Insights into the structure and tectonic history of the southern south island, new zealand, from the 3-d distribution of P- and S-wave attenuation, *Geophys. J. Int.,* **214,** 1481–1505.

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

Epanomeritakis, I., Akçelik, V., Ghattas, O. & Bielak, J., 2008. A Newton-CG method for large-scale three-dimensional elastic full-waveform seismic inversion, *Inverse Problems,* **24**(3), 034015. doi:10.1088/0266-5611/24/3/034015.

Fabien-Ouellet, G., Gloaguen, E. & Giroux, B., 2017. Time domain viscoelastic full waveform inversion, *Geophys. J. Int.,* **209,** 1718–1734.

Faul, U. & Jackson, I., 2015. Transient creep and strain energy dissipation: an experimental perspective, *Ann. Rev. Earth planet. Sci.,* **43,** 541–569.

Fehmers, G. & Höcker, C., 2003. Fast structural interpretation with structure-oriented filtering, *Geophysics,* **68,** 1286–1293.

Fichtner, A., 2010. *Full Seismic Waveform Modelling and Inversion,* pringer Science & Business Media.

Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.P., 2009. Full waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.,* **179,** 1703–1725.

Fichtner, A. & Trampert, J., 2011. Hessian kernels of seismic data functionals based upon adjoint techniques, *Geophys. J. Int.,* **185,** 775–798.

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

Fichtner, A. & Villaseñor, A., 2015. Crust and upper mantle of the western Mediterranean – constraints from full-waveform inversion, *Earth planet. Sci. Lett.,* **428,** 52–62.

Fischer, K.M., Ford, H.A., Abt, D.L. & Rychert, C.A., 2010. The lithosphere-asthenosphere boundary, *Ann. Rev. Earth planet. Sci.,* **38,** 551–575.

Fischer, K.M., Rychert, C.A., Dalton, C.A., Miller, M.S., Beghein, C. & Schutt, D.L., 2020. A comparison of oceanic and continental mantle lithosphere, *Phys. Earth planet. Int.,* **309,** 106600. https://doi.org/10.1016/j.pepi.2020.106600.

French, S.W. & Romanowicz, B., 2015. Broad plumes rooted at the base of the Earth's mantle beneath major hotspots, *Nature,* **525,** 95–99.

Gao, H., 2018. Three-dimensional variations of the slab geometry correlate with earthquake distributions at the Cascadia subduction system, *Nat. Commun.,* **9,** 1204. https://doi.org/10.1038/s41467-018-03655-5

Gao, Y., Tilmann, F., van Herwaarden, D.P., Thrastarson, S., Fichtner, A., Heit, B., Yuan, X. & Schurr, B., 2021. Full waveform inversion beneath the central andes: Insight into the dehydration of the Nazca slab and delamination of the back-arc lithosphere, *J. geophys. Res.,* **126**(7), e2021JB021984, doi:10.1029/2021JB021984.

Gauthier, O., Virieux, J. & Tarantola, A., 1986. Two-dimensional nonlinear inversion of seismic waveforms: numerical results, *Geophysics,* **51,** 1387–1403.

Hale, D., 2013. Implementing an anisotropic and spatially varying Matérn model covariance with smoothing filters, CWP Report 815.

Harmon, N., *et al.* 2021. Widespread hydration of the back arc and the link to variable hydration of the incoming plate in the Lesser Antilles from Rayleigh Wave Imaging., *Geochem. Geophys. Geosys.,* **22,**doi:10.1029/2021GC009707.

Harmon, N., Rychert , C.A., Agius, M.R., Tharimena, S., Le Bas , T. P., Kendall, J.M. & Constable, S.. 2018. Marine geophysical investigation of the Chain Fracture Zone in the equatorial Atlantic from the PI-LAB Experiment, *J. geophys. Res.,* **123,**doi:10.1029/2018jb015982.

Harmon, N., Rychert, C.A., Kendall, J.M., Agius, M., Bogiatzis, P. & Tharimena, S., 2020. Evolution of the oceanic lithosphere in the equatorial atlantic from Rayleigh wave tomography, evidence for small-scale convection from the PI-LAB experiment, *Geochem. Geophys. Geosyst.,* **21**(9), e2020GC009174, doi:10.1029/2020GC009174.

Harmon, N., Salas De La Cruz, M., Rychert, C., Abers, G. & Fischer, K., 2013. Crustal and mantle shear velocity structure of Costa Rica and Nicaragua from ambient noise and teleseismic Rayleigh wave tomography, *Geophys. J. Int.,* **195,** 1300–1313.

Harmon, N., Wang, S., Rychert, C.A., Constable, S. & Kendall, J.M.. 2021. Shear velocity inversion guided by resistivity structure from the PI-LAB Experiment for integrated estimates of partial melt in the mantle, *J. Geophys. Res.,* **126,**e2021JB022202.

Havlin, C., Holtzman, B. & Hopper, E., 2021. Inference of thermodynamic state in the asthenosphere from anelastic properties, with applications to North American upper mantle, *Phys. Earth planet. Inter.,* **314,** 106639. https://doi.org/10.1016/j.pepi.2020.106639

Igel, H., Djikpesse, H. & Tarantola, A., 1996. Waveform inversion of marine reflection seismograms for P impedance and Poisson's ratio, *Geophys. J. Int.,* **124,** 363–371.

Jackson, I., 2015. *Treatise in Geophysics,* 2nd edn, Schubert, G., Elsevier.

Jackson, I. & Faul, U.H., 2010. Grainsize-sensitive viscoelastic relaxation in olivine: towards a robust laboratory-based model for seismological application, *Phys. Earth planet. Inter.,* **183,** 151–163.

Kappel, F. & Kuntsevich, A., 2000. An implementation of Shor'S R-algorithm, *Comput. Optim. Appl.,* **15,** 193–205.

Karaoğlu, H. & Romanowicz, B., 2018. Global seismic attenuation imaging using full-waveform inversion: a comparative assessment of different choices of misfit functionals, *Geophys. J. Int.,* **212,** 807–826.

Ko, Y.T., Kuo, B.Y. & Hung, S.H., 2012. Robust determination of earthquake source parameters and mantle attenuation, *J. geophys. Res.,* **117**(B4), doi:10.1029/2011JB008759.

Kolsky, H., 1965. The propagation of stress pulses in viscoelastic solids, *Phil. Mag.,* **8,** 693–710.

Komatitsch, D., 1999. *Specfem3D cartesian [software]* doi: GITHASH8, https://geodynamics.org.

Komatitsch, D. & Tromp, J., 1999. Introduction to the spectral-element method for 3-D seismic wave propagation, *Geophys. J. Int.,* **139,** 806–822.

Komatitsch, D., Xie, Z., Bozdağ, E., Andrade, E. S.d., Peter, D., Liu, Q. & Tromp, J., 2016. Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion, *Geophys. J. Int.,* **206,** 1467–1478.

Lailly, P., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Conference on Inverse Scattering—Theory and Application,* pp. 206–220. Bednar, J.B., Robinson, E. & Weglein, A., (eds.)SIAM.

Lavayssiere, A., Rychert , C.A, Harmon, N., Keir, D., Hammond, J.O.S., Kendall, J.M., Doubre, C. & Leroy, S.. 2018. maging lithospheric discontinuities beneath the northern East African Rift using S-to-P receiver functions, *Geochem. Geophys. Geosyst.,* **19,**doi:10.1029/2018gc007463.

Liu, H., Anderson, D. & Kanamori, H., 1976. Velocity dispersion due to anelasticity: implications for seismology and mantle composition, *Geophys. J. R. astr. Soc.,* **47,** 41–58.

Liu, Q. & Gu, J.Y., 2012. Seismic imaging: from classical to adjoint tomography, *Tectonophysics,* **566,** 31–66.

Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.,* **96,** 2383–2397.

Liu, Q. & Tromp, J., 2008. Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods, *Geophys. J. Int.,* **174,** 265–286.

Luo, Y., Modrak, R. & Tromp, J., 2014. Strategies in adjoint tomography, in *Handbook of Geomathematics,* 2nd edn, pp. 1943–2001, eds Freeden, W., Nashed, M. & Sonar, T., Springer.

McCarthy, C., Takei, Y. & Hiraga, T., 2011. Experimental study of attenuation and dispersion over a broad frequency range: 2. The universal scaling of polycrystalline materials, *J. geophys. Res.,* **116,** B09207. doi:10.1029/2011JB008384.

Métivier, L., Bretaudeau, F., Brossier, R., Operto, S. & Virieux, J., 2014. Full waveform inversion and the truncated newton method: quantitative imaging of complex subsurface structures, *Geophys. Prospect.,* **62,** 1353–1375.

Métivier, L. & Brossier, R., 2016. The seiscope optimization toolbox: a large-scale nonlinear optimization library based on reverse communication, *Geophysics,* **81,** F1–F15.

Métivier, L., Brossier, R., Virieux, J. & Operto, S., 2013. Full waveform inversion and the truncated newton method, *SIAM J. Sci. Comput.,* **35,** B401–B437.

Moczo, P. & Kristek, J., 2005. On the rheological models used for time-domain methods of seismic wave propagation, *Geophys. Res. Lett.,* **32,** L01306. doi:10.1029/2004GL021598.

Modrak, R., Borisov, D., Lefebvre, M. & Tromp, J., 2018. Seisflows—flexible waveform inversion software, *Comput. Geosci.,* **195,** 88–95.

Müller, G., 1983. Rheological properties and velocity dispersion of a medium with power-law dependence of Q on frequency, *J. Geophys.,* **54,** 20–29.

Myers, S.C., Beck, S., Zandt, G. & Wallace, T., 1998. Lithospheric-scale structure across the Bolivian andes from tomographic images of velocity and attenuation for P and S waves, *J. geophys. Res.,* **103,** 21 233–21 252.

Nocedal, J. & Wright, S.J., 2006. *Numerical Optimisation,* Springer.

Pakravan, A., Kang, J. & Newtson, C., 2016. A gauss–newton full-waveform inversion for material profile reconstruction in viscoelastic semi-infinite solid media, *Inverse Prob. Sci. Eng.,* **24,** 393–421.

Pan, W. & Wang, Y., 2020. On the influence of different misfit functions for attenuation estimation in viscoelastic full-waveform inversion: synthetic study, *Geophys. J. Int.,* **221,** 1292–1319.

Pan, W., Wang, Y. & Innanen, K., 2021. Full-waveform adjoint Q tomography in viscoelastic medium with central-frequency measurements, *Authorea,* February 06, 2021, doi:10.1002/essoar.10506130.1.

Peter, D., *et al.* 2011. Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes, *Geophys. J. Int.,* **186,**721–739.

Possee , D., Rychert , C.A., Harmon, N. & Keir, D.. 2021. Seismic Discontinuities across the North American Caribbean Plate Boundary from S-to P- receiver functions, *Geochem. Geophys. Geosys.,* **22,** doi:10.1029/2021GC009723

Pozgay, S.H., Wiens, D.A., Conder, J.A., Shiobara, H. & Sugioka, H., 2009. Seismic attenuation tomography of the Mariana subduction system: implications for thermal structure, volatile distribution, and slow spreading dynamics, *Geochem. Geophys. Geosyst.,* **10**(4), Q04X05.

Pratt, R.G., Shin, C. & Hicks, G.J., 1998. Gauss-Newton and full newton methods in frequency-space seismic waveform inversion, *Geophys. J. Int.,* **133,** 341–362.

Ruan, Y.Y., Forsyth, D.W. & Bell, S.W., 2018. Shear attenuation beneath the Juan de Fuca Plate: implications for mantle flow and dehydration, *Earth planet. Sci. Lett.,* **496,** 189–197.

Rychert, C A, *et al.* 2021. A dynamic lithosphere-asthenosphere boundary at the equatorial Mid-Atlantic Ridge, *Earth & Planet. Sci. Let.,* doi:10.1016/j.epsl.2021.116949

Rychert, C. A., *et al.* 2012. Volcanism in the Afar Rift sustained by decompression melting with minimal plume influence, *Nature Geosciene,* **6**(), 406–409.

Rychert, C. A., Harmon, N. & Ebinger, C.. 2014. Receiver function imaging of lithospheric structure and the onset of melting beneath the Galapagos Archipelago, *Earth and Planetary Science Letters,* **388,**156–165.

Rychert, C. A., Harmon, N. & Tharimena, S.. 2018a. Scattered wave imaging of the oceanic plate in Cascadia, *Sci. Adv.,* **4**(), doi:10.1126/sciadv.aao1908

Rychert, C. A., Rondenay, S. & Fischer, K. M.. 2007. P-to-S and S-to-P imaging of a sharp lithosphere-asthenosphere boundary beneath eastern North America, *J. Geophys. Res.,* **112** B08314. doi:10.1029/2006JB004619

Rychert, C. A. & Shearer, P.. 2009. A Global View of the Lithosphere-Asthenosphere Boundary, *Science,* **324**(), 495–498.

Rychert, C.A & Shearer, P.. 2011. Imaging the lithosphere-asthenosphere boundary beneath the Pacific using SS waveform modeling, *J. geophys. Res.,* **116,** B07307. doi:10.1029/2010jb008070

Rychert, C.A., Fischer, K.M., Abers, G.A., Plank, T., Syracuse, E., Protti, J.M., Gonzalez, V. & Strauch, W., 2008. Strong along-arc variations in attenuation in the mantle wedge beneath Costa Rica and Nicaragua, *Geochem. Geophys. Geosyst.,* **9**(10), Q10S10. doi:10.1029/2008GC002040.

Rychert, C.A., Fischer, K.M. & Rondenay, S., 2005. A sharp lithosphere-asthenosphere boundary imaged beneath eastern North America, *Nature,* **436,** 542–545.

Rychert , C.A., Harmon, N. & Armitage, J.. 2018b. Seismic imaging of thickened lithosphere resulting from plume pulsing beneath Iceland, *Geochem. Geophys. Geosyst.,* **19**(), (eds H. Yuan and B. Romanowicz). . https://doi.org/10.1002/9781119249740.ch4

Rychert, C.A., Harmon, N., Constable, S. & Wang, S., 2020. The nature of the lithosphere-asthenosphere boundary, *J. geophys. Res.,* **125,** e2018JB016463, doi:10.1029/2018JB016463.

Rychert, C.A., Harmon, N. & Tharimena, S.. 2018c. Seismic Imaging of the Base of the Ocean Plates, *Lithospheric Discontinuities, AGU monograph,* **Ch.4** doi:10.1002/9781119249740.ch4

Rychert , C.A. & Harmon, N.. 2017. Constraints on the anisotropic contributions to velocity discontinuities at ∼60 km depth beneath the Pacific, *Geochem. Geophys. Geosyst.,* **18**(), doi:10.1002/2017gc006850

Rychert, C.A., Shearer, P.M. & Fischer, K.M.. 2010. Scattered wave imaging of the lithosphere-asthenosphere boundary, *Lithos,* **120**(), 173–185.

Rychert, C.A., Tharimena, S., Harmon, N., Wang, S., Constable, S., Kendall, J.M., Bogiatzis, P., Agius, M.R. & Schlaphorst, D., 2021. A dynamic lithosphere–asthenosphere boundary near the equatorial mid-Atlantic Ridge, *Earth planet. Sci. Lett.,* **566,** doi:10.1016/j.epsl.2021.116949.

Saikia, U., Rychert, C.A., Harmon, N. & Kendall, J.M.. 2021. Upper mantle anisotropic shear velocity structure at the equatorial Mid-Atlantic Ridge constrained by Rayleigh wave group velocity analysis from the

PI-LAB experiment, *Geochem. Geophys. Geosys.,* **22,** e2020GC009495. doi:10.1029/2020GC009495

Saikia, U., Rychert, C.A., Harmon, N. & Michael Kendall, J., 2021. Seismic attenuation at the equatorial mid-Atlantic Ridge constrained by local Rayleigh wave analysis from the pi-lab experiment, *Geochem. Geophys. Geosyst.,* **22**(12), e2021GC010085, doi:10.1029/2021GC010085.

Savage, B., Dimitri, K. & Jeroen, T., 2010. Effects of 3D attenuation on seismic wave amplitude and phase measurements, *Bull. seism. Soc. Am.,* **100,** 1241–1251.

Schurr, B., Asch, G., Rietbrock, A., Trumbull, R. & Haberland, C., 2003. Complex patterns of fluid and melt transport in the central Andean subduction zone revealed by attenuation tomography, *Earth planet. Sci. Lett.,* **215,** 105–119.

Stachnik, J.C., Abers, G.A. & Christensen, D.H., 2004. Seismic attenuation and mantle wedge temperatures in the Alaska subduction zone, *J. geophys. Res.,* **109**(B10), doi:10.1029/2004JB003018.

Syracuse, E.M., Abers, G.A., Fischer, K., MacKenzie, L., Rychert, C., Protti, M., González, V. & Strauch, W., 2008. Seismic tomography and earthquake locations in the nicaraguan and costa rican upper mantle, *Geochem. Geophys. Geosyst.,* **9,** doi:10.1029/2008GC001963.

Takanami, T., Sacks, I.S. & Hasegawa, A., 2000. Attenuation structure beneath the volcanic front in northeastern Japan from broad-band seismograms, *Phys. Earth planet. Inter.,* **121,** 339–357.

Takeuchi, N., Kawakatsu, H., Shiobara, H., Isse, T., Sugioka, H., Ito, A. & Utada, H., 2020. Inversion of longer-period OBS waveforms for P structures in the oceanic lithosphere and asthenosphere, *J. geophys. Res.,* **125**(7), e2019JB018810, doi:10.1029/2019JB018810.

Tao, K., Grand, S.P. & Niu, F., 2018. Seismic structure of the upper mantle beneath eastern Asia from full waveform seismic tomography, *Geochem. Geophys. Geosyst.,* **19,** 2732–2763.

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., 1988. Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation, *Pure appl. Geophys.,* **128,** 365–399.

Tharimena, S., Rychert, C.A., Harmon, N. & White, P.. 2017. Imaging Pacific lithosphere seismic discontinuitiesInsights from SS precursor modeling, *J Geophys Res,* **122**(3), doi:10.1002/2016jb013526

Tharimena S., Rychert , C.A. & Harmon, N.. 2016. Seismic imaging of a mid-lithospheric discontinuity beneath Ontong Java Plateau, *Earth and Planetary Science Letters,* **450**(), 62–70.

Tharimena, S., Rychert , C.A. & Harmon, N.. 2017. A unified continental thickness from seismology and diamonds suggests a melt-defined plate, *Science,* **357**(), 580–583.

Tromp, J. , Komatitsch, D. & Liu, Q.. 2008. Spectral-element and adjoint-methods in seismology, *Communications in Computational Physics,* **3**(), 1–32.

Tromp, J., 2020. Seismic wavefield imaging of Earth's interior across scales, *Nat. Rev. Earth Environ.,* **1,** 40–53.

Tromp, J., Tape, C. & Liu, Q., 2005. Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels, *Geophys. J. Int.,* **160,** 195–216.

Tsumura, N., Matsumoto, S., Horiuchi, S. & Hasegawa, A., 2000. Three-dimensional attenuation structure beneath the northeastern Japan arc estimated from spectra of small earthquakes, *Tectonophysics,* **319,** 241–260.

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

Wang, S., Constable, S., Rychert, C.A. & Harmon, N.. 2020. A lithosphere-asthenosphere boundary and partial melt estimated using marine magnetotelluric data at the central Middle Atlantic Ridge, *Geochem Geophys Geosys,* **21**(), doi:10.1029/2020GC009177.

Wei, S.S. & Wiens, D.A., 2018. P-wave attenuation structure of the Lau back-arc basin and implications for mantle wedge processes, *Earth planet. Sci. Lett.,* **502,** 187–199.

Wei, S.S., Wiens, D.A., Zha, Y., Plank, T., Webb, S.C., Blackman, D.K., Dunn, R.A. & Conder, J.A., 2015. Seismic evidence of effects of water on melt transport in the Lau back-arc mantle, *Nature,* **518,** 395–398.

Weickert, J., 1999. Coherence-enhancing diffusion filtering, *Int. J. Comput. Vis.,* **31,** 111–127.

Xie, Y., Rychert, C., Harmon, N., Liu, Q. & Gajewski, D., 2021. On-the-fly full Hessian kernel calculations based upon seismic wave simulations, *Seismol. Res. Lett.,* **92,** 3832–3844.

Xing, G. & Zhu, T., 2019. Modeling frequency-independent Q viscoacoustic wave propagation in heterogeneous media, *J. geophys. Res.,* **124,** 11 568–11584.

Xing, G. & Zhu, T., 2020. Hessian-based multiparameter fractional viscoacoustic full-waveform inversion, in *SEG Technical Program Expanded Abstracts 2020,* pp. 895–899.

Yamauchi, H. & Takei, Y., 2016. Polycrystal anelasticity at near-solidus temperatures, *J. geophys. Res.,* **121,** 7790–7820.

Yang, J., Zhu, H., Li, X., Ren, L. & Zhang, S., 2020. Estimating *P*-wave velocity and attenuation structures using full waveform inversion based on a time domain complex-valued viscoacoustic wave equation: the method, *J. geophys. Res.,* **125**(6), e2019JB019129, doi:10.1029/2019JB019129.

Yang, P., Brossier, R., Méetivier, L. & Virieux, J., 2016a. Checkpointing-assisted reverse forward simulation: an optimal recomputation method for FWI and RTM, in *SEG Technical Program Expanded Abstracts 2016,* pp. 1089–1093.

Yang, P., Brossier, R., Métivier, L. & Virieux, J., 2016b. Wavefield reconstruction in attenuating media: a checkpointing-assisted reverse-forward simulation method, *Geophysics,* **81**(6), R349–R362.

Yang, P., Brossier, R., Métivier, L., Virieux, J. & Zhou, W., 2018. A time-domain preconditioned truncated newton approach to visco-acoustic multiparameter full waveform inversion, *SIAM J. Sci. Comput.,* **40,** B1101–B1130.

Yang, Y.J., Forsyth, D.W. & Weeraratne, D.S., 2007. Seismic attenuation near the east pacific rise and the origin of the low-velocity zone, *Earth Planet. Sci. Lett.,* **258,** 260–268.

Yong, P., Brossier, R. & Métivier, L., 2020. Parsimonious time-domain truncated-newton method in FWI thanks to fourier-domain and "full scattered field" approximation, in *SEG Technical Program Expanded Abstracts 2020,* pp. 830–834.

Zhang, C., Xie, Z., Komatitsch, D., Cristini, P. & Matzen, R., 2016. Revisiting the 1/L problem in rheological models for time-domain seismic wave propagation, in *Proceedings of the 86th annual meeting of the Society of Exploration Geophysics (SEG'2016),* pp. 1–3, Dallas, TX, United States.

Zhu, H., Bozdağ, E., Peter, D. & Tromp, J., 2012. Structure of the European upper mantle revealed by adjoint tomography, *Nat. Geosci.,* **5,** 493–498.

Zhu, H., Bozdağ, E. & Tromp, J., 2015. Seismic structure of the European upper mantle based on adjoint tomography, *Geophys. J. Int.,* **201,** 18–52.

Zhu, H. & Tromp, J., 2013. Mapping tectonic deformation in the crust and upper mantle beneath Europe and the North Atlantic Ocean, *Science,* **314,** 871–875.

APPENDIX A: A GENERAL FORM OF THE FULL HESSIAN KERNEL

The Hessian operator H applied to two arbitrary model perturbations may be expressed as (see Fichtner & Trampert 2011)

$$H(\delta\mathbf{m}_1, \delta\mathbf{m}_2) = \int_V K_m^1 \delta\mathbf{m}_2 \, d^3\mathbf{x} = \int_V \left(K_m^{1\to 2} + K_m^{2\leftrightarrow 1} + K_m^{1\to 1}\right) \delta\mathbf{m}_2 d^3\mathbf{x}, \tag{A1}$$

where V denotes the model volume. $K_m^1 = K_m^{1\to 2} + K_m^{2\leftrightarrow 1} + K_m^{1\to 1}$ denotes the full Hessian kernels, which may be expressed by the forward displacement field, $\mathbf{s}$, the adjoint displacement field, $\mathbf{s}^\dagger$ and their perturbed fields, $\delta_1\mathbf{s}$ and $\delta_1\mathbf{s}^\dagger$,

$$K_m^1 = \int_0^T \left[\mathbf{s}^\dagger \nabla_m \mathbf{L}(\delta_1\mathbf{s}) + \delta_1\mathbf{s}^\dagger \nabla_m \mathbf{L}(\mathbf{s}) + \mathbf{s}^\dagger \nabla_m \nabla_m \mathbf{L}(\mathbf{s})(\delta\mathbf{m}_1)\right] dt. \tag{A2}$$

The $\delta_1\mathbf{s}^\dagger$ may be decomposed into $\delta_1\mathbf{s}_m^\dagger$ and $\delta_1\mathbf{s}_s^\dagger$ with $\delta_1\mathbf{s}^\dagger = \delta_1\mathbf{s}_m^\dagger + \delta_1\mathbf{s}_s^\dagger$, which corresponds to the separation of the total perturbation that is owing to the model perturbation and adjoint source perturbation, respectively. Hence, eq. (A2) may be rewritten to

$$K_m^1 = \int_0^T \left[\mathbf{s}^\dagger \nabla_m \mathbf{L}(\delta_1\mathbf{s}) + \delta_1\mathbf{s}_m^\dagger \nabla_m \mathbf{L}(\mathbf{s}) + \delta_1\mathbf{s}_s^\dagger \nabla_m \mathbf{L}(\mathbf{s}) + \mathbf{s}^\dagger \nabla_m \nabla_m \mathbf{L}(\mathbf{s})(\delta\mathbf{m}_1)\right] dt. \tag{A3}$$

For the sake of simplicity, spatial and temporal dependencies of the vectors or matrices are omitted unless stated otherwise. The $\delta\mathbf{m}_1$ and $\delta\mathbf{m}_2$ are two model perturbations. The $\delta_1\mathbf{s}$ is the perturbed forward field due to the perturbation of the model, $\delta\mathbf{m}_1$. The two portions, $\delta_1\mathbf{s}_m^\dagger$ and $\delta_1\mathbf{s}_s^\dagger$, are the perturbed adjoint fields related to the perturbed model and perturbed adjoint source, respectively. They can be computed separately or simultaneously in two or one adjoint simulation(s) by substituting the adjoint source term. See the work of Xie *et al.* (2021) for implementation of the elastic case. A similar derivation of the Hessian kernels (also called Hessian vector product) was given by Métivier *et al.* (2013, 2014) using the derivation ideas of the adjoint method, which are typically computed by the velocity-stress wave equations using the finite-difference approach for wave simulations and the associated adjoint state method for gradient calculations. Once the forward and adjoint fields are computed from the two models $\mathbf{m}_1$ and $\mathbf{m}_1 + v\delta\mathbf{m}_1$ are available, construction of the full Hessian kernels for the elastic media is straightforward using eqs (A1)–(A3). However, practical challenges include the large disk space and I/O requirements to store and transmit these wavefields during a single Hessian kernel construction or iterative Hessian kernel constructions for full-waveform inversion. Therefore, the wavefield compression approach (Boehm & Ulbrich 2015; Boehm *et al.* 2016) or the on-the-fly approach (Xie *et al.* 2021) may be used to decrease the storage and I/O issues. The Hessian kernel expressions with a specific parametrization of model parameters can be found from the kernels with another parametrization through the chain rule (see e.g. Fichtner & Trampert 2011; Xie *et al.* 2021).

APPENDIX B: POST-PROCESSING OF FULL HESSIAN KERNELS

B1 Removing strong artefacts around the source and receiver locations

For simplicity, the full Hessian kernel is written as $\mathbf{H}$ in this section. We remove the strong amplitude around the sources and receivers for the full Hessian kernels based on a Gaussian typed function, which is defined by

$$\mathbf{H}(x_i) = \int_V \mathbf{H}_{raw}(x_i) \left[1 - \exp\left[-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{i,d} - x_{sr,d})^2}{R_d^2}\right]\right] dV, \tag{B1}$$

where $\mathbf{H}_{\text{raw}}(x_i)$ indicates the raw Hessian kernel computed by the SEM and second-order adjoint method directly at position x_i where i indicates the spatial index. The R_d controls the semi-axis of the ellipse (2-D) or ellipsoid (3-D) from the centre point, $x_{sr,\,d}$ (source or receiver location). The semi-axes are selected by several candidates tested from a few earthquakes. The exponential factor, $\left[1 - \exp\left[-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_{i,d}-x_{sr,d})^2}{R_d^2}\right]\right] \in [0\ 1)$, has the minimum value of zero at the centre point. A smaller R_d is selected for the receivers than that of the sources due to the strong source injection. The computation of Hessian kernels includes the computation of Fréchet kernels in which the operator eq. (B1) is applied to the Fréchet kernels as well. The Hessian kernel is multiplied by the exponential factors for all receivers of one event and that is repeated for all events, which takes a few seconds for the 2-D examples as discussed in Section 5. We apply this operator to each event kernel separately before the kernel summation for all events. Once the summed Hessian kernels are obtained, the smoothing operators are applied to stabilize the inversions. In this work, we remove the source–receiver artefacts for both the Fréchet and Hessian kernels using the above operator. The radius R for these kernels are set to $30(\mathrm{d}x + \mathrm{d}z)/2$ and $15(\mathrm{d}x + \mathrm{d}z)/2$, respectively, where $\mathrm{d}x$ and $\mathrm{d}z$ represent the grid spacing for the model mesher.

B2 Smoothing

Similar to the post-processing for Fréchet kernels, we smooth full Hessian kernels in order to solve the Newton equation of eq. (27). The Gaussian function used to smooth the Fréchet kernels (e.g. Zhu *et al.* 2015) is used to smooth the raw full Hessian kernels $\mathbf{H}_{\text{raw}}(\mathbf{x}^{'})$, which may be regarded as a regularization procedure,

$$\mathbf{H}(\mathbf{x}) = \frac{1}{W(\mathbf{x})}\int_V \mathbf{H}_{raw}(\mathbf{x}') \exp\left[-(x' D_a)^2/\left(2\sigma_a^2\right)\right] \exp\left[-(x - x')^2/\left(2\sigma_x^2\right)\right] d^3\mathbf{x}', \tag{B2}$$

$$W(\mathbf{x}) = \int_V \exp[-(x' D_a)^2/(2\sigma_a^2)] \exp[-(x - x')^2/(2\sigma_x^2)] d^3\mathbf{x}', \tag{B3}$$

where D_a denotes the azimuthal distance between points $\mathbf{x}$ and $\mathbf{x}^{'}$. The $x = ||\mathbf{x}||$ and $x^{'} = ||\mathbf{x}^{'}||$. The two quantities σ_a and σ_x controls the half-width of the Gaussian function in the azimuthal and radial direction, respectively. The operator degrades to 2-D while omitting the azimuthal terms. Different from eq. (B1) which applies to the source and receiver locations, respectively, the operator of eq. (B2) applies to each of the GLL points within the kernels, making its implementation more expensive than eq. (B1) if the same smoothing aperture is applied. The half widths of the Gaussian function are chosen based on the wavelengths of structure which can be resolved by the investigated frequency bands. We use the same smooth aperture of $\sigma_a = 14\,000$ and $\sigma_x = 3500$ for both the Fréchet and Hessian kernels at the low frequency stage.

At high frequency stages or if the model includes heterogeneity, the raw Hessian kernel may not be simply smoothed by the Gaussian filter that correlates varies only with distance, without a variation in direction for each GLL point. Therefore, we introduce the structure-oriented smoothing operator (Weickert 1999; Fehmers & Höcker 2003; Hale 2013) to preserve structure features for the Hessian kernel in our codes. Equivalently, next to the convolution with the exponential coefficient as in eq. (B2), smoothing the Hessian kernel at each point $\mathbf{x}$ may be defined as

$$\mathbf{H}(\mathbf{x}) = \mathbf{C}(\mathbf{x})\, \mathbf{H}_{raw}(\mathbf{x}), \tag{B4}$$

where $\mathbf{C}(\mathbf{x})$ denotes the kernel covariance operator. Similar to the work of Hale (2013), we apply the covariance operator $\mathbf{C}(\mathbf{x})$ by solving the following partial differential equations (Hale 2013),

$$|\mathbf{D}|^{-\frac{1}{4}}(\mathbf{x})(1 - \alpha\nabla \cdot \mathbf{D}(\mathbf{x}) \cdot \nabla)^l(1 - \beta\nabla \cdot \mathbf{D}(\mathbf{x}) \cdot \nabla)|\mathbf{D}|^{-\frac{1}{4}}\mathbf{H}(\mathbf{x}) = \gamma\mathbf{H}_{raw}(\mathbf{x}). \tag{B5}$$

eq. (B5) may be efficiently solved by solving the following sequence of equations instead (Hale 2013):

$$\begin{aligned}
&\mathbf{H}_0(\mathbf{x}) = |\gamma^2\mathbf{D}(\mathbf{x})|^{\frac{1}{4}}\ \mathbf{H}_{raw}(\mathbf{x}),\\
&(1 - \alpha\nabla \cdot \mathbf{D}(\mathbf{x}) \cdot \nabla)\ \mathbf{H}_l(\mathbf{x}) = \mathbf{H}_{l-1}(\mathbf{x}),\\
&(1 - \beta\nabla \cdot \mathbf{D}(\mathbf{x}) \cdot \nabla)\ \mathbf{H}_L(\mathbf{x}) = \mathbf{H}_l(\mathbf{x}),\\
&\mathbf{H}(\mathbf{x}) = |\gamma^2\mathbf{D}(\mathbf{x})|^{\frac{1}{4}}\mathbf{H}_L(\mathbf{x}),
\end{aligned} \tag{B6}$$

where the $\mathbf{H}_l$ can be computed sequentially based on its previous $\mathbf{H}_{l-1}$ with $l = 1, 2, 3,..., L-1$. The tensor field $\mathbf{D}(\mathbf{x})$ may be computed from the input Hessian kernel $\mathbf{H}_{raw}(\mathbf{x})$ by

$$\mathbf{D}(\mathbf{x}) = [\nabla\mathbf{H}_{raw}(\mathbf{x})]\ [\nabla\mathbf{H}_{raw}(\mathbf{x})]^T. \tag{B7}$$

We use $l = 2$, $\alpha = 1$ and $\beta = 0$ for the 2-D examples. Besides the two operators for the kernel post-processing, no other pre-conditioner or regularization terms are used in this work. In the 3-D case, we have not yet used this smoothing method since the Gaussian smoothing codes provided by the SPECFEM3D worked well for the investigated frequency up to 0.2 Hz in our tests. We may consider this smoothing method in the 3-D in the future.

APPENDIX C: ANELASTIC KERNEL CALCULATIONS USING SPECFEM2D/3-D BASED ON THE ADJOINT MEMORY VARIABLES

C1 SPECFEM2D using the adjoint memory variables for kernel calculations.

To compute the Fréchet kernels for viscoelastic media, the method presented by Fichtner & van Driel (2014) performs well and it does not require additional computational cost to compute the Q Fréchet kernels. . In this section, we show how to compute the Fréchet kernels for viscoelastic properties based on the adjoint memory variables using the public SPECFEM2D package and the associated codes are published with this work. To do this, we revisit the stress–strain relation in a viscoelastic media (e.g. Emmerich & Kornm 1987; Blanch *et al.* 1995; Carcione 2001, and therein). The stress–strain relation in a viscoelastic medium may be written as

$$\sigma_{ij}(t) = \int_{-\infty}^{\infty} C_{ijkl}(t - t')\,\dot{\epsilon}_{kl}(t')\,dt' = \int_{-\infty}^{\infty} \dot{C}_{ijkl}(t - t')\,\epsilon_{kl}(t')\,dt' = \int_{-\infty}^{\infty} M_{ijkl}(t - t')\,\epsilon_{kl}(t')\,dt', \tag{C1}$$

where $C_{ijkl}(t - t^{'})$ represents the elastic tensor and the dot $\cdot$ above the letter denotes the time derivative. The integral interval depends on the simulation system. The $C(t)$ for the generalized Zener model may be defined as (e.g. Liu *et al.* 1976; Carcione 2001; Moczo & Kristek 2005; Fichtner & van Driel 2014)

$$C(t) = M^r\left[1 + \frac{1}{N}\sum_{l=1}^{N}\left(\frac{\tau_\epsilon^{(l)}}{\tau_\sigma^{(l)}} - 1\right)e^{-t/\tau_\sigma^{(l)}}\right]H(t) = M^r\left[1 + \frac{1}{Q}\sum_{l=1}^{N}D^{(l)}e^{-t/\tau_\sigma^{(l)}}\right]H(t), \tag{C2}$$

where M^r indicates the relaxed modulus, $\tau_\epsilon^{(l)}$ and $\tau_\sigma^{(l)}$ are the relaxation times and $H(t)$ is the Heaviside function. Differentiating eq. (C2) with respect to time t, yields

$$M(t) = M^r\left[1 + \frac{1}{Q}\sum_{l=1}^{N} D^{(l)} e^{-t/\tau_\sigma^{(l)}}\right]\delta(t) - M^r\left[\frac{1}{Q}\sum_{l=1}^{N}\frac{D^{(l)}}{\tau_\sigma^{(l)}} e^{-t/\tau_\sigma^{(l)}}\right] H(t), \tag{C3}$$

where $\delta(t)$ is the delta function. For an isotropic media, inserting eq. (C3) into eq. (C1) leads to

$$\sigma_{ij}(t) = M_\kappa^r\left[1 + \frac{1}{Q_\kappa}\sum_{l=1}^{N} D_\kappa^{(l)}\right]\epsilon_{kk}(t)\delta_{ij} + 2M_\mu^r\left[1 + \frac{1}{Q_\mu}\sum_{l=1}^{N} D_\mu^{(l)}\right]\tilde{\epsilon}_{ij}(t) + M_\kappa^r\sum_{l=1}^{N} R_\kappa^{(l)}\delta_{ij} + 2M_\mu^r\sum_{l=1}^{N}\tilde{R}_{ij,\mu}^{(l)}, \tag{C4}$$

where M_κ^r and M_μ^r denote the relaxed elastic modulus for κ and μ, respectively. The SPECFEM2D uses the unrelaxed elastic modulus, for the adjoint simulation, eq. (C4) may be written as

$$\sigma_{ij}^\dagger = M_\kappa^U\,\epsilon_{kk}^\dagger\delta_{ij} + 2M_\mu^U\,\tilde{\epsilon}_{ij}^\dagger + M_\kappa^U\left[1 + \frac{1}{Q_\kappa}\sum_{l=1}^{N} D_\kappa^{(l)}\right]^{-1}\sum_{l=1}^{N} R_\kappa^{(l)\dagger}\delta_{ij} + 2M_\mu^U\left[1 + \frac{1}{Q_\mu}\sum_{l=1}^{N} D_\mu^{(l)}\right]^{-1}\sum_{l=1}^{N}\tilde{R}_{ij,\mu}^{(l)\dagger}, \tag{C5}$$

where M_κ^U and M_μ^U are the unrelaxed elastic modulus, respectively. The last two terms in eq. (C5) include the adjoint memory variables, which may be defined as

$$E_\kappa^{(l)\dagger} = \left[1 + \frac{1}{Q_\kappa}\sum_{l=1}^{N} D_\kappa^{(l)}\right]^{-1} R_\kappa^{(l)\dagger} = -\left[1 + \frac{1}{Q_\kappa}\sum_{l=1}^{N} D_\kappa^{(l)}\right]^{-1}\frac{1}{Q_\kappa}\frac{D_\kappa^{(l)}}{\tau_{\sigma,\kappa}^{(l)}}\int_{-\infty}^{\infty}\epsilon_{kk}^\dagger e^{-(t'-t)/\tau_{\sigma,\kappa}^{(l)}} H(t'-t)\,dt', \tag{C6}$$

$$\tilde{E}_{ij,\mu}^{(l)\dagger} = \left[1 + \frac{1}{Q_\mu}\sum_{l=1}^{N} D_\mu^{(l)}\right]^{-1}\tilde{R}_{ij,\mu}^{(1)\dagger} = -\left[1 + \frac{1}{Q_\mu}\sum_{l=1}^{N} D_\mu^{(l)}\right]^{-1}\frac{1}{Q_\mu}\frac{D_\mu^{(l)}}{\tau_{\sigma,\mu}^{(l)}}\int_{-\infty}^{\infty}\tilde{\epsilon}_{ij}^\dagger e^{-(t'-t)/\tau_{\sigma,\mu}^{(l)}}\,H(t'-t)\,dt'. \tag{C7}$$

Before the forward and adjoint simulation, for each grid point of the model, the $\frac{D^{(l)}}{Q}$ and $\tau_\sigma^{(l)}$ are searched for in order to approximate the input Q. Eqs (C6) and (C7) link to Eqs (16) and (13). The Q kernels can be computed when the adjoint memory variable and the $D^{(l)}$ and $\tau_\sigma^{(l)}$ are available (see Section 2). Note that in the 2-D case, we use $\kappa = \lambda + \mu$. Alternatively, the memory variables computed by the SPECFEM2D can be transformed to the memory variables $\tilde{\mathbf{R}}$ used in this work based on eqs (C6) and (C7). These expressions are only valid for the displacement–stress formulation that uses the adjoint strain for the adjoint memory variable calculations. For the adjoint memory variable that is computed by the time-derivative of the adjoint strain, one can refer to the work of Fichtner & van Driel (2014).

C2 SPECFEM3D Cartesian using the adjoint memory variables for the kernel calculations.

The relaxation times in the SPECFEM3D Cartesian package are computed differently in comparison to that of the SPECFEM2D package due to the efficient Q simulation, which the 3-D version uses the same, N evenly spaced stress relaxation times, τ_σ, for the Q_κ and Q_μ simulation, and the optimization is performed only to find the τ_ϵ for Q_κ and Q_μ. This strategy can be updated to optimize the τ_σ and τ_ϵ (or $D^{(l)}$ and $\tau_\sigma^{(l)}$) simultaneously for Q_κ or Q_μ in the subroutine of the τ_ϵ optimization, as similarly done in the SPECFEM2D. We then slightly update the 'get_attenuation_model' to get the two scaling factors for κ and μ using $D^{(l)}$ and $\tau_\sigma^{(l)}$, in which the κ and μ are scaled from the reference to the central frequency then to the unrelaxted state for the stress calculation later. The two relaxation times are also stored together with the two scaling factors at the *attenuation.bin*. The relaxation times are then used for the Q kernel calculation in the adjoint simulation. When the reference frequency is set to the central frequency, it is equivalent to scaling the modulus from the central frequency to the unrelaxted state. The 3×3 symmetric matrix memory variable $\tilde{\mathbf{R}}$ is computed in the subroutine of the memory variable calculation. Note that the SPECFEM3D Cartesian computes the memory variable based on the relaxation spectrum (Komatitsch & Tromp 1999; Savage *et al.* 2010). The memory variable of the SPECFEM3D Cartesian is numerically equivalent to the memory variable computed in this work assuming the stress terms are numerically accurately computed in both the two implementations: one in the SPECFEM3D Cartesian and one via eq. (C4) or eq. (C5). We compute five components (R11, R22, R12, R13 and R23) of the memory variable $\tilde{\mathbf{R}}$ for the stress and Q kernel calculation. The trace of $\tilde{\mathbf{R}}$ equals to zero. Once the relaxation times, scaling factors and the memory variables are computed, the memory variables are then used in the stress calculation based on eq. (C4) or eq. (C5), and also used for the Q kernel calculation in the 'compute_kernels' subroutine. Alternatively, instead of recomputing the memory variable, $\tilde{\mathbf{R}}$, in this work, one can still use the memory variables computed from the SPECFEM3D Cartesian, which are then transformed to the $\tilde{\mathbf{R}}$ based on eq. (C4) to eq. (C7) for the Q kernel calculation. In the 3-D case, we use $\kappa = \lambda + 2\mu/3$. The memory variables are expressed as trace for κ and as the deviator for μ. We have the moduli from the reference and the central frequency as well as the relaxed and unrelaxed moduli when attenuation is used in the simulation. We compute the stress terms using the unrelaxed modulus and the Q kernels can be computed via the relaxed or unrelaxed modulus. These moduli can be transformed into each other when the relaxation times are computed. Codes are given in the QuadSEM3D.

APPENDIX D: COMPUTATIONAL COSTS

D1 2-D full Hessian kernel construction on the fly or via checkpointing

Construction of the anelastic Hessian kernel using the QuadSEM-Q requires a 2-fold CPU time and memory cost in comparison to the construction of the anelastic Fréchet kernels alone, assuming the same number of computer cores are used. Double disk space is required when compared to the construction of the anelastic Fréchet kernels since two models are used. Due to the computer cluster used for the simulations and also to stabilize the run for the anelastic wave equations for recomputing the two anelastic forward fields in this work, we use 100 checkpoints to save the attenuated forward fields during the forward simulation, resulting in 100 chunks for the forward field reconstruction during the adjoint simulation. Each chunk includes 100 time steps ($100 \times 100 = 10\,000$). A different number of checkpoints can be considered as well, for instance, by increasing the number of cores when the anelastic wave simulation is stable. The cost for constructing the anelastic Hessian kernels is mostly taken up by the anelastic wavefield computation, which takes about 7 min the forward simulation and about 14 min for the adjoint simulation using 10 CPU cores (Intel® Xeon® Gold 6138 CPU @ 2.00GHz). For the wavefield storage, only about 1.5 GB of disk space is required by the QuadSEM-Q, which reduces the disk space requirement to two orders of magnitude smaller than the classical wavefield storage approach which would take about 230 GB of disk space for the full Hessian kernel calculations. The parsimonious storage implementation (Komatitsch *et al.* 2016) that is used for the anelastic Fréchet kernel calculations would take about 152 GB of disk space since two sets of fields still need to be stored for the full Hessian kernel calculations. Note that the calculation of the disk space requirements here accounts for only 100 checkpoints for computing one full Hessian kernel. If the classical full wavefield storage approach is used, more than 10 times the disk space (i.e. 2300 GB) and 20 times the I/O overhead (i.e. 4600 GB) would be required if the waves are simulated for 1000 seconds (∼16.6 min). In each iteration, the full Hessian kernels need to be computed three times if three CG iterations are used, which requires 3 times the I/O.. The disk space and I/O requirements are still very high while using the parsimonious storage method alone. A combination of the on-the-fly approach (Xie *et al.* 2021) with the parsimonious storage method (Komatitsch *et al.* 2016) can be used to reduce the disk space and I/O requirements.

One snapshot of a forward field takes about 7.8 MB of memory per core when 10 cores are used. The more cores that are used, the smaller the memory that is used on each core. The 7.8 MB of memory per core is much smaller than the 4000 MB of memory per core in most present-day HPC computing platforms. Large HPC clusters or supercomputers are readily available nowadays, where users typically use only a small portion of the memory per core, typically between 5 and 30 per cent, as generally a relatively large number of processor cores are used to keep the calculation relatively fast (see Komatitsch *et al.* 2016, for more discussions). The cost for recomputing the forward fields may be compensated by the fewer number of iterations due to the fact that the full Hessian kernels are used to improve the convergence rate. As an extension of the traditional parsimonious storage method (Komatitsch *et al.* 2016), we do not adopt the solution that doubles the parsimonious storage method in a parallel fashion because it still has the massive I/O issue for transmitting the large numbers of wavefields over computing nodes or cores in order to access these required fields simultaneously. We use all wavefields simultaneously in the computer memory because they are simultaneously computed for each time step.

D2 3-D Fréchet and full Hessian kernel construction

In the 3-D case, wavefield storage and I/O costs become challenging, making the classical wavefield storage method computationally prohibitive. In the 3-D example of this work for one forward simulation, about 32.22 MB disk space is required to store the *Database.bin*, and about 386.19 MB is required to store the model parameters (e.g. *vp.bin*) and the coordinates (e.g. *x.bin*). About 880 MB disk space is required to store the *external_mesh.bin* and about 680 MB disk space to store the *attenuation.bin*. The *attenuation.bin* typically stores the scaling factors for M_κ and M_μ when attenuation is used in the simulation. Hence, the total disk space requirement excluding the storage of the wavefields and waveforms is about 1.6 GB.

In the elastic case, the wavefield storage is used by the last snapshot of the forward fields and the absorbing boundary fields. The former uses about 210 MB disk space and the latter uses about 21.37 GB disk space. The waveform storage for all stations is only about 5.75 MB. The storage requirement for the elastic Fréchet kernels in terms of *Vp*, *Vs* and ρ is about 128.75 MB. The same storage is required for the full Hessian kernel since both are stored at the same GLL level. Thereby, the total disk space for the elastic Fréchet kernel calculation uses about 23.30 GB disk space, of which most of the disk space is used to store the absorbing boundary fields. A simultaneous computation of the 3-D elastic Fréchet and full Hessian kernels on the fly would approximately double the disk space requirement, which would need about 46 GB disk space under the on-the-fly feature. The disk space requirement here is calculated for one event. If the on-the-fly approach is not used, the disk storage requirements will be greater.

In the viscoelastic case, the wavefields are stored at the checkpoints, and a storage of the last snapshot of the forward fields and the absorbing boundary fields is not required. The disk space requirement for all forward fields at the checkpoints is about 60.54 GB for 50 checkpoints, where each checkpoint uses 1.21 GB disk space. If all 5000 time steps (simulated for 100 s) are stored, it would be taken about 5.91 TB disk space. The 50 evenly spaced checkpoints (100 time steps per chunk) may be enough for this case. The possible minimum number of checkpoints, N_{cp}, can be estimated via a comparison of the waveforms computed from several different N_{cp}. Except for the storage requirement for the wavefields, most of the other disk space requirements are the same as that of the elastic case. The viscoelastic simulation

needs to store the *attenuation.bin*, and the kernels are stored for Vp, Vs, ρ, Q_κ and Q_μ. The disk space requirement for the viscoelastic case is calculated for one event.

The computational time for computing the elastic Fréchet kernels is about 33 min, which includes about 8.47 min for the forward simulation and about 24.10 min for the adjoint simulation when 10 CPUs are used (Intel® Xeon® Gold 6138 CPU @ 2.00GHz). A simultaneous computation of the elastic Fréchet and full Hessian kernels doubles the computational time when compared to the computation of the Fréchet kernel alone. The computational time here is calculated for one event. The viscoelastic simulation again approximately doubles the computational time when compared to the pure elastic simulation. The forward simulation with attenuation takes about 15 min and the adjoint simulation takes about 40 min. The computation times may be influenced by the method of implementation.

D3 Cost of the 2-D and 3-D inversions

The cost of the 2-D and 3-D inversions can be estimated approximately based on the cost of kernel calculation for one event, the number of events, the number of frequency stages, and the number of iterations for each frequency stage. Additional costs are needed for debugging, parameter testing, and the calls in the inversion workflow, etc. The maximum disk space and memory requirements can be estimated via one iteration. We remove the temporary files or useless files when moving to next iteration and next frequency stage.