Efficient time-domain 3D elastic and viscoelastic full-waveform inversion using a spectral-element method on flexible Cartesian-based mesh

Phuong-Thu Trinh[1], Romain Brossier[2], Ludovic Métivier[3], Laure Tavard[4], and Jean Virieux[2]

ABSTRACT

Viscoelastic full-waveform inversion is recognized as a challenging task for current acquisition deployment at the crustal scale. We have developed an efficient formulation based on a time-domain spectral-element method on a flexible Cartesian-based mesh. We consider anisotropic elastic coefficients and isotropic attenuation. Complete gradient expressions including the attenuation contribution spread into those of elastic components are given in a consistent way. The influence of attenuation on the P-wave velocity reconstruction is illustrated through a toy configuration. The numerical implementation of the forward problem includes efficient matrix-vector products for solving second-order elastodynamic equations for 3D geometries: An original high-order integration for topography effects is performed at nearly no extra cost. Combined adjoint and forward field recomputation from the final state and previously saved boundary values allows the estimation of misfit gradients for density, elastic parameters, and attenuation factors with no I/O efforts. Two-level parallelism is implemented over the sources and domain decomposition, which is necessary for a realistic 3D configuration. The misfit gradient preconditioning is performed by a so-called Bessel filter using an efficient differential implementation based on finite-element ingredients on the forward mesh instead of the often-used, costly convolution approach. A 3D synthetic illustration is provided on a subset $(2 \times 7 \times 3 \text{ km})$ of the SEG Advanced Modeling (SEAM) Phase II Foothills model with 4 lines of 20 sources. The structurally based Bessel filter and a simple data hierarchy strategy considering early body waves before all waves including surface waves allow a precise reconstruction of the P- and S-wavespeeds while keeping a smooth density description.

INTRODUCTION

In the majority of earth imaging applications, the physics of wave propagation can be described through partial differential equations (PDEs). The coefficients of such PDEs, related to the physical and mechanical parameters of the subsurface, can be adjusted through comparisons between the recorded and the numerically simulated data. Among the various techniques, the full-waveform inversion (FWI) technique (Lailly, 1983; Tarantola, 1984) offers the possibility to extract high-resolution (down to half the local wavelength) and quantitative multiparameter models of the subsurface (for reviews of the method, see Virieux and Operto, 2009; Virieux et al., 2017).

Because the pioneering 2D FWI applications with surface seismic data (Ravaut et al., 2004; Operto et al., 2006; Brenders and Pratt, 2007), most of FWI applications at the crustal scale have been performed under the acoustic approximation, generally for marine environments (Plessix and Perkins, 2010; Sirgue et al., 2010; Vigh et al., 2013; Warner et al., 2013; Operto et al., 2015). In such a flat-surface condition, numerical methods designed with finite-difference (FD) discretization appear to be a reasonable and efficient choice due to their numerical efficiency, their relatively simple implementation, and the fact that optimized modeling kernels developed for reverse time migration can be shared. When considering the elastic-wave equation, fluid/solid contrasts can also be implicitly and efficiently considered for representing simple bathymetry with FD. However,

[1]Total E&P, Pau, France and Univ. Grenoble Alpes, ISTerre, Grenoble, France. E-mail: phuong-thu.trinh@univ-grenoble-alpes.fr (corresponding author).
[2]Univ. Grenoble Alpes, ISTerre, Grenoble, France. E-mail: romain.brossier@univ-grenoble-alpes.fr; jean.virieux@univ-grenoble-alpes.fr.
[3]Univ. Grenoble Alpes, ISTerre, CNRS, Grenoble, France. E-mail: ludovic.metivier@univ-grenoble-alpes.fr.
[4]Univ. Grenoble Alpes, ISTerre, GRICAD, Grenoble, France. E-mail: laure.tavard@univ-grenoble-alpes.fr.

standard FD formulations are often limited to a regular-spacing grid, although there are interesting extensions for deformed grids (Moczo, 1989; de la Puente et al., 2014; Petersson and Sjögreen, 2015; Shragge, 2016). In addition, significant extra effort should be performed when considering surface topography or important geologic interfaces (Robertsson, 1996; Bohlen and Saenger, 2006). Extensions such as the summation-by-parts approaches (Petersson and Sjögreen, 2012), immersed-boundaries implementation (LeVeque, 2007; Lombard et al., 2008; Gao et al., 2015; Huiskes et al., 2017), or hybrid techniques (Moczo et al., 1997) have been proposed. However, they have not yet displayed superior efficiency compared with finite-element (FE) methods, which are based on a variational formulation, naturally honoring the boundary conditions. Moreover, the FE formulation can easily consider deformed meshes, allowing us to represent complex geometries. Among many FE families, the so-called spectral-element methods (SEMs) have become popular in geophysics for regional and global seismology problems (Faccioli et al., 1997; Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999). The particular choice of tensorial-based hexahedral elements, Lagrange polynomial basis functions, and Gauss-Lobatto-Legendre (GLL) integration points embedded in conventional SEM formulations leads to a diagonal mass matrix by construction, resulting naturally in an efficient and accurate explicit time-marching scheme, without any mass-lumping strategy (Marfurt, 1984). Realistic geometries, such as geologic basin structures or even the approximated ellipsoidal shape of the earth, can be described using hexahedral elements in unstructured meshes, allowing accurate numerical calculation of seismic wavefields (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999). One of the main issues for these wave simulations turns out to be the hexahedral mesh buildup, which can require significant human effort (Peter et al., 2011).

When considering on-shore seismic acquisitions, free-surface-related effects can be quite important in the wavefield. For example, significant elastic and viscoelastic effects can be observed in the presence of weathering and unconsolidated near-surface, complex topographies and strong geologic interfaces (shale-carbonate for example). These complex effects cannot be fully removed or compensated by data preprocessing, such as attenuation compensation or ground-roll removal, implying that a correct description of the physics is strongly advisable for accurate model parameter estimation. Moreover, considering the complete physical phenomena of the wave propagation would allow taking the benefit of each piece of recorded data, for more accurate results and higher resolution. Regarding anelastic media, attenuation affects the amplitudes and phases of the seismic wavefield; it is thus necessary to consider its effect precisely during the modeling and imaging processes. In active seismic applications, the approximation of constant Q-value over a frequency band is widely accepted. Viscous effects have been widely considered in acoustic-based frequency-domain modeling and FWI, at least as passive parameters during inversion, thanks to the straightforward and free implementation through complex-valued velocity. Impressive 3D viscoacoustic illustrations have been shown in Operto et al. (2015) and Amestoy et al. (2016) for velocity model building by monoparameter frequency-domain FWI based on a sparse direct solver, or more recently with a joint reconstruction of the Q model (Operto and Miniussi, 2017). When moving toward elastic and/or viscoelastic approximation, especially for 3D land data, frequency-domain approaches with direct solvers are out of reach for actual computer hardware technology, as mentioned by

Gosselin-Cliche and Giroux (2014). Alternative iterative solvers would require efficient and robust preconditioning for the modeling part (Li et al., 2014; Kostin et al., 2016). In addition, dispersive and complex wavefield would require considering a significant number of discrete frequencies, an additional argument reducing the attractability of frequency-domain approaches (Sirgue and Pratt, 2004; Brossier et al., 2009). Thus, the time-domain formulation is the currently selected approach when performing FWI for elastic media at various scales (Tape et al., 2010; Peter et al., 2011; Fichtner et al., 2013; Vigh et al., 2014; Borisov and Singh, 2015; Zhu et al., 2015). The time-domain formulation also allows the application of time windowing, wavefield separation, and various signal processing techniques, which favors strategies based on data-windowing hierarchy in FWI. It should be noted that these strategies are difficult to achieve in the frequency domain with few discrete frequencies, aside from the product by exponent decay of the Laplace transform proposed by Shin and Cha (2008). For example, regional and global scales FWI benefit from this time-domain formulation, in which the phases can be separated and selected before inversion (Maggi et al., 2009). At the exploration scales, seismic phases are however mixed up, and prior separation can be more challenging. In comparison to the frequency-domain approach, the time-domain formulation leads to a memory intensive and numerically expensive forward problem when attenuation is considered (Bohlen, 2002; van Driel and Nissen-Meyer, 2014). FWI gradient estimation also involves the zero-lag crosscorrelation in time between the incident and adjoint wavefields, weighted by the spatial radiation pattern. Although elastic modeling allows recomputing the incident field backward in time from the final snapshot and boundary field values thanks to the reversibility of the wave equation (Dussaud et al., 2008), the attenuation breaks this property. In such a case, checkpointing-based approaches (Griewank and Walther, 2000; Symes, 2007; Anderson et al., 2012; Komatitsch et al., 2016; Yang et al., 2016b) or compressed storage strategies (Boehm et al., 2016) are required, increasing again the computational cost in CPU time and memory of the FWI in viscoelastic media.

In this manuscript, we aim at developing an efficient, flexible, and accurate viscoelastic FWI methodology, dedicated to exploration-scale targets. This method is based on the SEM for the modeling part. In the first section, we review the second-order viscoelastic and elastic-wave equation for the displacement field. The associated FWI problem is formulated, and the FWI gradient expressions are derived for anisotropic elastic parameters and isotropic attenuation coefficients, based on a least-squares misfit function. As an illustration of these exact expressions, the impact of the attenuation when recovering elastic parameters will be pointed out in a toy configuration in the third section: No implicit approximation is considered on attenuation mechanism aside from the standard-linear-solid (SLS) description. We shall discuss the similarities and differences with previous expressions provided in the literature. In the second section, we describe our efficient SEM-based implementation in the FWI scheme, including flexible mesh design, optimized modeling, and inversion kernels. Our strategy focuses on the algorithmic balance between numerical efficiency, memory requirement, and simulation accuracy, especially for 3D viscoelastic problems. The geologic prior knowledge can be incorporated through the Bessel function-based gradient preconditioning, directly implemented by an SEM formulation on the same mesh as the wavefield modeling (Trinh et al., 2017b). In the third section, numerical performance

for the modeling part is evaluated. Similar computational behaviors are observed compared with the open-source FD modeling package SW4 V1.1 (Petersson and Sjögreen, 2013) and SEM-based tool SPECFEM V2.0 (Peter et al., 2011). Furthermore, a careful analysis on the viscoelastic gradient of compressional velocity in an ideal configuration emphasizes the impact of the attenuation on the reconstruction of velocity parameters. The fourth section is devoted to 3D elastic FWI examples at the exploration scale, which is a subset of the SEAM Phase II Foothill benchmark (Oristaglio, 2012, 2016; Regone et al., 2017). The examples illustrate the importance of all the individual components that we have developed, including a data-windowing hierarchy for simultaneous P- and S-wave velocity estimation. The required computer resources for 3D elastic and viscoelastic FWI on this subset are also provided.

VISCOELASTIC MODELING AND FWI

The viscoelastic wave propagation can be written in its complete form as

$$\begin{cases} \rho(\mathbf{x})\partial_{tt}u_i(\mathbf{x},t)\partial_j\sigma_{ij}(\mathbf{x},t) + f_i(\mathbf{x},t), \\ \sigma_{ij}(\mathbf{x},t) = M_{ijkl}(\mathbf{x},t) *_t \varepsilon_{kl}(\mathbf{x},t) + \mathcal{T}_{ij}(\mathbf{x},t), \end{cases} \quad (1)$$

where the density is denoted by ρ, the displacement by u, and the second-order stress and strain tensors by σ and ε, respectively. The external force is denoted by the vector $\mathbf{f}$, and the tensor $\mathcal{T}$ is the possible stress failure. The attenuation effect is described by the relaxation rate M_{ijkl}, and the symbol $*_t$ stands for convolution in the time domain. These equations use the Einstein convention (summation over repeated indices).

Elastic forward problem

The pure elastic-wave equation can be seen as a particular case of the viscoelastic equation with the specific relaxation rate

$$M_{ijkl}(\mathbf{x},t) = c_{ijkl}(\mathbf{x})\delta(t), \quad (2)$$

where c_{ijkl} is the elastic (or *unrelaxed*) stiffness coefficient and $\delta(t)$ is a Dirac delta function, resulting in the elastic-wave equation:

$$\begin{cases} \rho(\mathbf{x})\partial_{tt}u_i(\mathbf{x},t) = \partial_j\sigma_{ij}(\mathbf{x},t) + f_i(\mathbf{x},t), \\ \sigma_{ij}(\mathbf{x},t) = c_{ijkl}(\mathbf{x})\varepsilon_{kl}(\mathbf{x},t) + \mathcal{T}_{ij}(\mathbf{x},t). \end{cases} \quad (3)$$

Following the Voigt indexing and matrix notation, the second-order elastic-wave equation of displacement field $\mathbf{u}$ can be written as

$$\rho\partial_{tt}\mathbf{u} = DCD^T\mathbf{u} + \mathbf{S}, \quad (4)$$

where the spatial derivative operator in the Cartesian space is denoted by D

$$D = \begin{pmatrix} \partial_1 & 0 & 0 & 0 & \partial_3 & \partial_2 \\ 0 & \partial_2 & 0 & \partial_3 & 0 & \partial_1 \\ 0 & 0 & \partial_3 & \partial_2 & \partial_1 & 0 \end{pmatrix}, \quad (5)$$

the elastic stiffness tensor is denoted by C, the transposed operator by ".^{T}," and the source term by $\mathbf{S}$. The elastic equation is self-adjoint, implying that we can develop an adjoint system similar to the

incident problem. This property makes possible to use the same numerical scheme for incident and adjoint fields propagation.

Viscoelastic forward problem

The attenuation effect in seismic data is characterized by the energy loss per cycle of the phase, measured by the quality factor. The seismic attenuation is then described by the quality-factor tensor $Q_{ijkl}(\mathbf{x})$ (Emmerich and Korn, 1987; Carcione et al., 1988). The coefficients of this tensor are the model parameters that we consider for describing the attenuation. In the frequency domain, the inverse of these parameters is defined as the ratio of the imaginary and real parts of the complex relaxation rate $\hat{M}_{ijkl}(\mathbf{x},\omega)$, in which the " $\hat{}$ " notation denotes the Fourier transform. For realistic crustal scale application, the approximation of a constant Q over the considered frequency band $[\omega_{\min}, \omega_{\max}]$, together with the assumption that $Q \gg 1$ are usually accepted. Assuming that L SLS systems at specific reference frequencies $\omega_\nu \in [\omega_{\min}, \omega_{\max}]$ are used to fit a constant Q parameter over this frequency range, one may introduce anelastic coefficients Y_ν^{ijkl}, $\nu = 1, \ldots, L$, in the definition of the complex relaxation rate $\hat{M}_{ijkl}(\mathbf{x},\omega)$ as (Emmerich and Korn, 1987; Blanch et al., 1995; Moczo and Kristek, 2005; van Driel and Nissen-Meyer, 2014)

$$\hat{M}_{ijkl}(\mathbf{x},\omega) = c_{ijkl}(\mathbf{x})\left(1 - \sum_{\nu=1}^{L} Y_\nu^{ijkl}(\mathbf{x})\frac{\omega_\nu}{\omega_\nu + i\omega}\right). \quad (6)$$

Instead of computing L anelastic coefficients $Y_\nu^{ijkl}(\mathbf{x})$ for each spatial location $\mathbf{x}$, Yang et al. (2016a) propose the approximation $Y_\nu^{ijkl}(\mathbf{x}) \approx y_\nu Q_{ijkl}^{-1}(\mathbf{x})$ for the entire medium, in which the scalars y_ν being L dimensionless anelastic coefficients. The complex relaxation rate can thus be written as

$$\hat{M}_{ijkl}(\mathbf{x},\omega) = c_{ijkl}(\mathbf{x}) - \underbrace{c_{ijkl}(\mathbf{x})Q_{ijkl}^{-1}(\mathbf{x})\sum_{\nu=1}^{L} y_\nu \frac{\omega_\nu}{\omega_\nu + i\omega}}_{\text{attenuation mechanisms}}.$$

$$(7)$$

In equation 7, the product between the two fourth-order tensors c_{ijkl} and Q_{ijkl}^{-1} is a term-by-term product. The coefficients y_ν are estimated from the following least-squares minimization problem:

$$\min_{y_\nu}\left\{\int_{\omega_{\min}}^{\omega_{\max}}\left[Q_{\text{ref}}^{-1}\left(y_\nu\frac{\omega_\nu\omega}{\omega^2 + \omega_\nu^2} - 1\right)\right]^2\right\}, \quad (8)$$

which minimizes the distance between a given constant-value Q_{ref}^{-1} with its approximation over the frequency band $[\omega_{\min}, \omega_{\max}]$. The value of this constant is usually chosen such that $Q_{\text{ref}} \in [\min_{ijkl,\mathbf{x}}(Q_{ijkl}(\mathbf{x})), \max_{ijkl,\mathbf{x}}(Q_{ijkl}(\mathbf{x}))]$. Equations 7 and 8 incorporate explicitly the attenuation parameters $Q_{ijkl}^{-1}(\mathbf{x})$ into the wave equation, following a cheap-memory strategy: instead of storing the $L \times 21$ anelastic coefficients $Y_\nu^{ijkl}(\mathbf{x})$ associated with 21 coefficients $M_{ijkl}(\mathbf{x})$ at each spatial location (Komatitsch and Tromp, 1999), we only need to store L scalar y_ν for the entire medium and the heterogeneous attenuation parameters over the medium. We also store L reference frequencies ω_ν associated with L SLSs.

To separate the elastic rheology (represented by the unrelaxed stiffness coefficients $c_{ijkl}(\mathbf{x})$ in equation 7) with the attenuation mechanisms, which are frequency-dependent, we introduce attenuative stiffness coefficients c^a_{ijkl}. In the case of fully anisotropic attenuation, these parameters can be linked to the unrelaxed quantities through

$$c^a_{ijkl}(\mathbf{x}) = c_{ijkl}(\mathbf{x})Q^{-1}_{ijkl}(\mathbf{x}). \qquad (9)$$

The convolution relationship between stress and strain in equation 1 then becomes

$$\sigma_{ij}(\mathbf{x},t) = c_{ijkl}(\mathbf{x})\varepsilon_{kl}(\mathbf{x},t) - c^a_{ijkl}(\mathbf{x})\sum_{\nu=1}^{L}\psi_{\nu;kl}(\mathbf{x},t) + \mathcal{T}_{ij}(\mathbf{x},t), \qquad (10)$$

where the memory variable $\psi_{\nu;kl}(\mathbf{x},t)$ satisfies the first-order ordinary differential equation (ODE):

$$\partial_t\psi_{\nu;kl}(\mathbf{x},t) + \omega_\nu\psi_{\nu;kl}(\mathbf{x},t) = \omega_\nu y_\nu\varepsilon_{kl}(\mathbf{x},t). \qquad (11)$$

The mathematical demonstration is given in Appendix A, which is coherent with Moczo and Kristek (2005) and Yang et al. (2016a). Similar to the elastic case (equation 4), following the Voigt indexing, the second-order viscoelastic wave equation can thus be written as

$$\begin{cases} \rho\partial_{tt}\mathbf{u} = DCD^T\mathbf{u} - DC^a\sum_{\nu=1}^{L}\boldsymbol{\psi}_\nu + \mathbf{S}, \\ \partial_t\boldsymbol{\psi}_\nu + \omega_\nu\boldsymbol{\psi}_\nu = y_\nu\omega_\nu D^T\mathbf{u}, \nu = 1, \ldots, L, \end{cases} \qquad (12)$$

where $\boldsymbol{\psi}_\nu$ is the memory-variable vector associated with each SLS. The matrix $C^a = (C^a_{IJ})_{6\times6}$ contains the attenuative stiffness coefficients, with 21 independent components in the case of fully anisotropic attenuation. Due to the memory variables, the viscoelastic equation 12 is not self-adjoint. This property is related to the energy dissipation; however, we shall show that the considering adjoint system can be transformed into a similar structure as the incident problem. It should be noted that the memory variables can act either on the stress fields as it is the case here (equation 10), or on the displacement fields (Petersson and Sjögreen, 2012). The former strategy requires $L \times 6$ variables and two spatial derivatives per forward problem, whereas $L \times 3$ incident memory variables with at least three spatial derivatives are needed in the latter approach: two derivatives for the displacement fields and at least one for the memory variables. Because the estimation of the spatial derivatives is the most expensive operation in our SEM-based modeling kernels, we inject the memory variables into the stress fields as shown in equation 10 to optimize the computational cost.

Isotropic attenuation

Practical applications are generally limited to isotropic attenuation because constraining the anisotropic attenuation parameters is difficult for surface seismic acquisition. It is important to highlight that the isotropic attenuation regime implies that the attenuative stiffness tensor $(C^a_{IJ})_{6\times6}$ is isotropic but not the inverse quality factor matrix $(Q^{-1}_{IJ})_{6\times6}$ (Moczo et al., 1997). Hence, the Voigt-noted matrix can be written in the form

$$C^a = \begin{bmatrix} \lambda^a + 2\mu^a & \lambda^a & \lambda^a & 0 & 0 & 0 \\ \lambda^a & \lambda^a + 2\mu^a & \lambda^a & 0 & 0 & 0 \\ \lambda^a & \lambda^a & \lambda^a + 2\mu^a & 0 & 0 & 0 \\ 0 & 0 & 0 & \mu^a & 0 & 0 \\ 0 & 0 & 0 & 0 & \mu^a & 0 \\ 0 & 0 & 0 & 0 & 0 & \mu^a \end{bmatrix}, \qquad (13)$$

where λ^a and μ^a are the attenuative Lamé coefficients. Thus, we consider the effective mechanisms with

$$\lambda^a + 2\mu^a = \frac{1}{3}Q^{-1}_p(C_{11} + C_{22} + C_{33}),$$
$$\mu^a = \frac{1}{3}Q^{-1}_s(C_{44} + C_{55} + C_{66}), \qquad (14)$$

where the heterogeneous model parameters Q_p and Q_s are the attenuation parameters associated with the P- and S-wavespeeds (V_P, V_S) that we use. In the case of the isotropic-attenuation mechanism and isotropic-elastic rheology, the proposed solution is consistent with the development suggested by Moczo et al. (1997). Equation 14 also implies that isotropic attenuation has an isotropic impact on compressional and shear components even when considering anisotropic elasticity, which is physically meaningful. By doing so, the attenuation parameters Q_p and Q_s are explicitly incorporated in the wave equation, even for anisotropic elasticity, and therefore they can be naturally considered in the FWI framework. Let us mention that the anisotropy feature is important for the characterization of the earth, which might come from intrinsic and/or structure-induced anisotropy such as layering structures. Therefore, considering isotropic attenuation while taking into account elastic anisotropy is a good approximation for describing seismic wave propagation in surface acquisitions.

Inversion problem

The standard FWI problem attempts to reduce the data misfit between the calculated $\mathbf{d}_{cal}$ and the observed data $\mathbf{d}_{obs}$ at receiver locations by iteratively minimizing the following least-squares norm:

$$\chi(\mathbf{m}) = \frac{1}{2}\|\mathbf{d}_{cal}(\mathbf{m}) - \mathbf{d}_{obs}\|^2. \qquad (15)$$

The data $\mathbf{d}_{cal}$ are computed on the model parameters $\mathbf{m}$. The formulation implies a summation over all sources and receivers and an integral over the recording time. Standard FWI is a local optimization method, in which the descent direction relies on the gradient of the misfit function, computed by the adjoint method (Plessix, 2006). We follow the Lagrangian strategy to develop the displacement and memory variable adjoint fields $(\bar{\mathbf{u}}, \bar{\psi}_\nu)$ based on the second-order forward problem presented in equation 12. The mathematical details are provided in Appendix B. It should be highlighted that the second-order adjoint system has nearly identical equations as the incident problem

$$\begin{cases} \rho\partial_{tt}\bar{\mathbf{u}} = DCD^T\bar{\mathbf{u}} - DC^a\sum_{\nu=1}^{L}\bar{\boldsymbol{\psi}}_\nu - R^\dagger\Delta d_{\mathbf{u}}, \\ \partial_t\bar{\psi}_\nu - \omega_\nu\bar{\psi}_\nu = -y_\nu\omega_\nu D^T\bar{\mathbf{u}}, \end{cases} \qquad (16)$$

except for the sign "–" in the memory variables. The adjoint wavefield propagates backward in time from a final condition. The "–"

sign in front of the memory variables thus implies that the back propagation of the adjoint wavefield is stable in time, as is the forward propagation of the incident wavefield. In addition, the algorithm workflow for the forward and backward propagation can be similar.

The optimal choice of model parameters $\mathbf{m}$, especially in an attenuating medium, is not obvious and is expected to be dependent on the medium complexity, acquisition design, and radiation pattern. As a reference and natural parameterization, we compute the gradient for the density ρ, the independent components of the unrelaxed stiffness components (C_{IJ}), and the inverse of the quality factors Q_p and Q_s

$$\mathbf{m} = (\rho, C_{IJ;(1 \leq I \leq J \leq 6)}, Q_p^{-1}, Q_s^{-1}). \qquad (17)$$

The mathematical details for the gradient computation of each model parameter are presented in Appendix B. For the elastic and viscoelastic problems, the gradient on density ρ is the zero-lag crosscorrelation of the adjoint displacement field and the incident acceleration field

$$\frac{\partial \chi(\mathbf{m})}{\partial \rho} = (\bar{\mathbf{u}}, \partial_{tt}\mathbf{u})_{\Omega,t}. \qquad (18)$$

The gradient on attenuation parameters Q_p^{-1} and Q_s^{-1} can be estimated as

$$\frac{\partial \chi(\mathbf{m})}{\partial Q_{p,s}^{-1}} = -\left(\bar{\boldsymbol{\varepsilon}}, \sum_{\nu=1}^{L} \frac{\partial C^a}{\partial Q_{p,s}^{-1}} \boldsymbol{\psi}_\nu \right)_{\Omega,t}, \qquad (19)$$

which is the zero-lag crosscorrelation of the adjoint strain field $\bar{\varepsilon}$ and the incident memory variable field $\boldsymbol{\psi}_\nu$. The explicit expression of the scaled quantity $\partial C^a / \partial Q_{p,s}^{-1}$ is provided in Appendix B. The obtained expression 19 is coherent with the Q gradient proposed by Fichtner and van Driel (2014).

Following the separation of the elastic rheology and attenuation mechanism in equation 7, the C_{IJ} elementary gradient has two terms:

$$\frac{\partial \chi(\mathbf{m})}{\partial C_{IJ}} = \left(\bar{\boldsymbol{\varepsilon}}, \frac{\partial C}{\partial C_{IJ}} \boldsymbol{\varepsilon} \right)_{\Omega,t} - \left(\bar{\boldsymbol{\varepsilon}}, \sum_{\nu=1}^{L} \frac{\partial C^a}{\partial C_{IJ}} \boldsymbol{\psi}_\nu \right)_{\Omega,t}. \qquad (20)$$

The first term has an identical expression as the elastic gradients proposed by Tromp et al. (2005); Virieux et al. (2017), which is the zero-lag crosscorrelation of the adjoint $\bar{\varepsilon}$ and the incident strain fields ε. The adjoint and incident strain fields are calculated in the anelastic medium. The second term is related to the memory variables $\boldsymbol{\psi}_\nu$. The explicit expressions of the scaled quantities $\partial C / \partial C_{IJ}$ and $\partial C^a / \partial C_{IJ}$ are given in Appendix B, which confirm that the attenuation coefficient Q_p has an isotropic impact on compressional components, thus having an isotropic impact on V_P estimation (a similar interpretation for the attenuation coefficient Q_s). The development of $\partial C^a / \partial C_{IJ}$ also highlights that, in a weakly attenuative medium, the second term in the viscoelastic C_{IJ} gradient (Equation 20) can be neglected and tends to zero in a purely elastic medium. Because the memory variables $\boldsymbol{\psi}_\nu$ obey the first-order ODE 11 with the incident strain field ε as the source, the magnitude of the second term is generally small compared with the first term. However, we shall expect an accumulative influence on the

amplitude of the model parameter reconstruction. A numerical illustration of this effect on a simple inclusion example is presented in the "Numerical investigations" section.

This interpretation is also consistent with other theoretical investigations in the literature. Tarantola (1988) shows that the gradient with respect to the general relaxation rate is the crosscorrelation of the adjoint and the incident strain fields. By considering the chain rule with respect to the stiffness coefficients, in which the relaxation function M is given in equation 7, a two-term expression for the C_{IJ} gradient, as for our formulation 20, shall be obtained. Liu and Tromp (2008) also express the gradient of the misfit function through the crosscorrelation of the adjoint and incident strain fields, scaled by the perturbation of the relaxation rate δM. The time dependency of this perturbation is related to the second term in our gradient expression. For numerical implementation, Liu and Tromp (2008) and Komatitsch et al. (2016) assume that δM is time independent, resulting in a simplification of the gradient kernels, explaining why they obtained similar expression as for the elastic case.

Furthermore, the gradient for any parameter α (seismic velocity, anisotropic parameter, impedance, etc.) can be computed by the chain rule using the density ρ, C_{IJ} and $Q_{p,s}^{-1}$ elementary gradients

$$\frac{\partial \chi(\mathbf{m})}{\partial \alpha} = \sum_{I=1}^{6} \sum_{J=I}^{6} \frac{\partial \chi}{\partial C_{IJ}} \frac{\partial C_{IJ}}{\partial \alpha} + \frac{\partial \chi}{\partial \rho} \frac{\partial \rho}{\partial \alpha} + \frac{\partial \chi}{\partial Q_p^{-1}} \frac{\partial Q_p^{-1}}{\partial \alpha}$$
$$+ \frac{\partial \chi}{\partial Q_s^{-1}} \frac{\partial Q_s^{-1}}{\partial \alpha}. \qquad (21)$$

EFFICIENT SEM-BASED IMPLEMENTATION

The section introduces an efficient SEM-based 3D viscoelastic FWI methodology designed for crustal-scale exploration. Our implementation, embedded in the SEM46 code (SEM for Seismic Imaging at eXploration scale) relies on several key elements, compared with existing modeling and FWI packages: (1) optimized SEM computing kernels for incident and adjoint fields, coupled with low-memory attenuation implementation (Deville et al., 2002; Yang et al., 2016a), (2) flexible Cartesian-based deformed mesh with high-order geometry representation to capture complex topographies and variable element sizes to reduce the numerical cost (Trinh et al., 2017a), (3) two message passing interface (MPI)-based parallelism levels for tackling large-scale and multiple-shot experiments, (4) various nonlinear optimization methods thanks to the SEISCOPE optimization toolbox (Métivier and Brossier, 2016), (5) FWI gradient building with the checkpointing-assisted reverse-forward simulation (CARFS) method, combining reverse propagation with compressed boundary storage and checkpointing (Yang et al., 2016b, 2016c), and (6) structurally based nonstationary and anisotropic smoothing filter implemented with SEM on the modeling mesh (Trinh et al., 2017b).

We use a classic hexahedron-based SEM frame (Faccioli et al., 1997; Komatitsch and Tromp, 1999) for wave modeling, in which the physical domain Ω is decomposed into a set of nonoverlapping hexahedral elements. The same mesh is used for the inversion steps. Each hexahedral element can be mapped to the reference cube $[-1,1] \otimes [-1,1] \otimes [-1,1]$. The cube is further discretized into a set of $(N+1)^3$ GLL points $(\xi_{k_1}, \eta_{k_2}, \zeta_{k_3})$; $k_1, k_2, k_3 = 0, \ldots, N$, where N refers to the interpolation order and k_1, k_2, and k_3 are the indexing of GLL points in z-, x-, and y-dimensions, respectively.

These collocation points are used to define $(N+1)^3$ basis functions; each function is a triple product of Lagrange polynomials of degree N.

Optimized and low-memory viscoelastic modeling kernels

Considering the choice of Lagrange polynomials as basis functions and the GLL quadrature for numerical integration, the weak form of the second-order viscoelastic wave equation 12 of the displacement field **u** can be written as

$$\begin{cases} \mathbf{M}\partial_{tt}\mathbf{u} = -\mathbf{K}\mathbf{u} + \mathcal{D}^w C^a \sum_{\nu=1}^{L} \boldsymbol{\psi}_\nu + \mathbf{F}, \\ \partial_t \boldsymbol{\psi}_\nu + \omega_\nu \boldsymbol{\psi}_\nu = \omega_\nu y_\nu \mathcal{D}\mathbf{u}. \end{cases} \quad (22)$$

The global mass and stiffness matrices are denoted by **M** and **K**, respectively, and the source term is denoted by **F**. The global mass matrix **M** is diagonal by construction for SEM approaches. The operator $\mathcal{D}$ estimates the spatial derivatives of a vector in the Cartesian space. The operator $\mathcal{D}^w$ is equivalent to a spatial derivative operator weighted by GLL weights. The free-surface condition is naturally taken into account by the weak formulation.

Considering a high-order interpolation N allows using a coarse mesh design. However, it also implies a strict Courant-Friedrichs-Lewy (CFL) stability condition on the time stepping, which significantly increases the number of time steps for the forward simulation. Our SEM implementation is based on limited interpolation orders for test functions with $N = 4$ or 5. It has been shown that these orders provide a good compromise between the numerical accuracy and the CFL constraint (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999). In addition, because our SEM modeling kernel aims at performing seismic modeling in heterogeneous media with spatial resolution down to half of the wavelength (optimal resolution of FWI; Virieux and Operto, 2009), higher order SEM coupled with coarse meshes would not easily represent the medium without using the homogenization theory (Capdeville et al., 2010). Further research on the efficiency and accuracy of using higher orders should be conducted.

Algorithm 1. Viscoelastic Newmark scheme.

1. **for** $it = 1, \ldots, nt$ **do**
2.   Prediction phase
3. $\mathbf{u}^{it} = \mathbf{u}^{it-1} + \Delta t \mathbf{v}^{it-1} + \frac{\Delta t^2}{2} \mathbf{a}^{it-1}$
4. $\mathbf{v}^{it-1/2} = \mathbf{v}^{it-1} + \frac{\Delta t}{2} \mathbf{a}^{it-1}$
5.   Resolution phase
6. $\boldsymbol{\varepsilon}^{it} = \mathcal{D}\mathbf{u}^{it}$
7. $\boldsymbol{\psi}_\nu^{it+1/2} = \gamma_\nu \boldsymbol{\psi}_\nu^{it-1/2} + \beta_\nu \boldsymbol{\varepsilon}^{it}$ (equation 24)
8. $\boldsymbol{\sigma}^{it} = C\boldsymbol{\varepsilon}^{it} - C^a \sum_{\nu=1}^{L} \frac{\boldsymbol{\psi}_\nu^{it+1/2} + \boldsymbol{\psi}_\nu^{it-1/2}}{2}$
9. $\mathbf{a}^{it} = -\mathbf{M}^{-1}\mathcal{D}^w\boldsymbol{\sigma}^{it} + \mathbf{F}^{it}$
10.   Correction phase
11. $\mathbf{v}^{it} = \mathbf{v}^{it-1/2} + \frac{\Delta t}{2} \mathbf{a}^{it}$
12. **end**

Optimized modeling kernels for the product of the displacement vector by the stiness matrix

The most computationally intensive part of the modeling kernel is the product of the displacement vector by the stiness matrix. The stiffness matrix **K** can be factorized as

$$\mathbf{K} = \mathcal{D}^w C \mathcal{D}. \quad (23)$$

The implementation of the product of the displacement vector by the matrix **K** is thus based on this factorization through a decomposition into three steps: the estimation of the spatial derivatives of the displacement leading to the strain component, the product with the stiffness coefficient leading to the stress components, and the last estimation of the spatial derivatives. Each spatial-derivative calculation is matrix free, and it is implemented following Deville et al. (2002) approach, taking benefit from the tensorial properties of hexahedral elements, optimizing the loop vectorization and cache usage by smart management of fast and slow dimensions in tables combined with manual loop unrolling for inner loops.

Low-memory viscoelastic modeling kernel

A second-order explicit Newmark scheme is implemented for the time integration to compute the displacement field at each time step (Komatitsch and Vilotte, 1998; Komatitsch and Tromp, 1999). The first-order ODE governing the memory variables (equation 22) needs to be incorporated into the time scheme without degrading the second-order accuracy. To do so, the memory variables are updated at half time steps compared with the displacement field

$$\boldsymbol{\psi}_\nu^{it+1/2} = \underbrace{\frac{2 - \omega_\nu \Delta t}{2 + \omega_\nu \Delta t}}_{\gamma_\nu} \boldsymbol{\psi}_\nu^{it-1/2} + \underbrace{\omega_\nu y_\nu \frac{2\Delta t}{2 + \omega_\nu \Delta t}}_{\beta_\nu} \mathcal{D}\mathbf{u}^{it}, \quad (24)$$

as shown in the viscoelastic Newmark scheme in Algorithm 1 for $\nu = 1, \ldots, L$. Following this expression, the memory variables at the next step are explicitly computed from the previous step and the already known displacement field. Therefore, we only need to store a single entity of the $\boldsymbol{\psi}_\nu$ fields, i.e., $L \times 6$ tables of memory variables per time step. In addition, as highlighted in equations 7 and 8, we only store L scalar anelastic coefficients y_ν to reduce the memory requirement related to SLS attenuation mechanism (Yang et al., 2016a). These two strategies allow us to reduce the memory footprint of viscoelastic simulation while preserving the accuracy. The simulation quality of our approach is illustrated in Appendix C, through the LOH benchmarks generated by Pacific Earthquake Engineering Research Center (Moczo et al., 2006). We obtain an excellent agreement with elastic and viscoelastic semianalytical solutions on layered medium, calculated from the boundary integral method (Coutant, 1989), even for the cross-comparison of different attenuation mechanisms.

Flexible Cartesian-based mesh design

In a completely unstructured FE mesh, a setup stage is required at the beginning of any simulation in which the indices or any neighborhood look-up tables are precomputed. We simplify the mesh design by considering a Cartesian-based deformed mesh to combine the accurate representation of the topography allowed by the FE

meshes and the convenience of an implementation on an FD grid (Cupillard et al., 2011; Gokhberg and Fichtner, 2016). The spatial position of each element can be obtained directly from its x, y, and z indices, without using any neighborhood look-up-tables or searching steps over the global mesh. The topography variation is represented by vertically deformed elements as shown in Figure 1a. The numbers of elements in the x-, y-, and z-directions are constant, although the element size can vary. For the basis functions at interpolation order $N = 4$ or 5, SEM allows to accurately model elastic waves propagation with approximately six GLL nodes per shortest wavelength (Komatitsch, 1997). This condition is referred to as the *volume condition*.

Cartesian-based mesh with variable element-size

In most geophysical targets, the velocity is spatially varying, leading to spatially varying wavelengths. The element size should thus be adapted to the variation of the local shortest wavelength. Figure 1b shows an example of such a flexible mesh design for a 3D target, in which the element size varies from 49 to 142 m in three directions. The mesh is built upon the condition that six GLL points per shortest wavelength are required. Under the same constraint, Figure 1a illustrates the constant element-size mesh, in which the element size is 40 m in each direction. In this cross section, by using a variable element-size mesh, the number of elements is reduced by a factor of 3.03 times in the z-direction, 1.57 times in the x-direction, and 1.47 times in the y-direction: almost a factor seven in total, while keeping the Cartesian-based topology of the mesh. The reduction of the number of elements systematically reduces the numerical cost by the same factor, or even more if it induces the relaxation of the CFL stability condition of the explicit time-marching scheme.

Complex topography representation

When considering significant topography variations, hexahedral elements can be vertically deformed as illustrated in Figure 1. The shape of each element is defined by a set of $(n + 1)$ control points in

each direction, leading to $(n + 1)^3$ control points and associated shape functions in 3D geometry. The spatial position of GLL points inside the element is computed from these control points and shape functions through

$$\mathbf{x}(\xi, \eta, \zeta) = \sum_{k_1=1}^{n+1} \sum_{k_2=1}^{n+1} \sum_{k_3=1}^{n+1} \ell_{\hat{k}}(\xi, \eta, \zeta) \mathbf{x}_{\hat{k}}, \quad (25)$$

where $\hat{k}$ stands for the triple indexes k_1, k_2, and k_3. The associated shape functions are a triple product of Lagrange polynomials of degree n: $\ell_{\hat{k}}(\xi, \eta, \zeta)$. The number of control points and shape functions $(n + 1)$ is not related to the interpolation order N of the test functions needed for solving the wave-equation PDE.

Figure 2a and 2b show the relative position of GLL points inside a deformed elements under P_1 (using eight corners of each element as control points and linear shape functions) and P_4 representation $((4 + 1)^3$ GLL points are used as control points and order 4 shape functions). The P_1 representation of the surface cannot honor sharp spatial variation of the free surface, as shown in Figure 2c, which might affect the accuracy of the simulation due to the interaction between elastic waves and the complex surface. Decreasing the element size is one way for following the rapid variation of the topography, namely, the *surface condition*: At least two spatial sampling points per shortest topographic wavelength are required to honor the interaction at the surface. This criterion might be stricter than the volume condition, and it would significantly increase the computational cost. The surface condition limitation can be overcome by using higher order P_n representation, such as $n = 4$ (Figure 2d) where all GLL points inside the element are considered as control points.

Figure 2d further confirms that the P_4 representation provides a better description of the complex topography because $(4 + 1)^2$ GLL points are used in each element to capture the topography map, instead of $(1 + 1)^2$ points for the P_1 case. Thus, this leads to a more accurate spatial positioning of GLL points inside the element. To evaluate the influence of the topography representation on
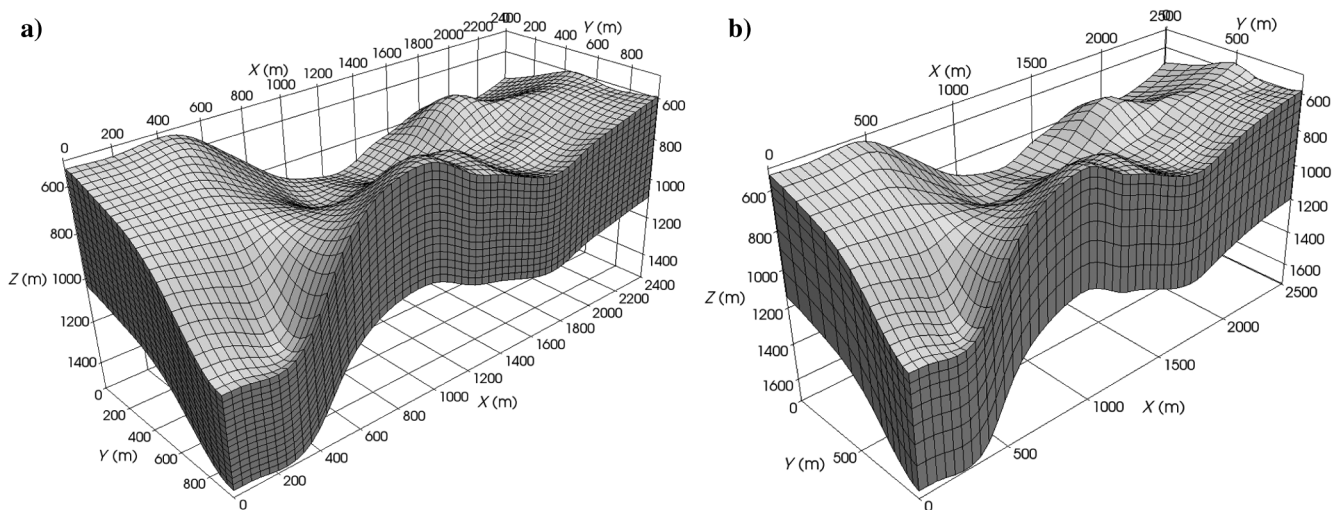

Figure 1. The 3D mesh generation on a realistic example (4 km in the z-direction, 8 km in the x-direction, and 2 km in the y-direction). (a) A subset of the constant element-size mesh with $dz = dx = dy = 40$ m. (b) A subset of the variable element-size mesh, in which the number of elements is reduced 3.03 times in the z-direction, 1.57 times in the x-direction, and 1.47 times in the y-direction: $dz \in [49 - 142]$ m, $dx \in [49 - 139]$ m, and $dy \in [49 - 100]$ m.

the accuracy of the wavefield simulation, we use a homogeneous model with a complex surface extracted from the SEAM Phase II Foothills benchmark as shown in Figure 3a. The medium is meshed at 100 m element-size, under P_1 and P_4 topography representations.
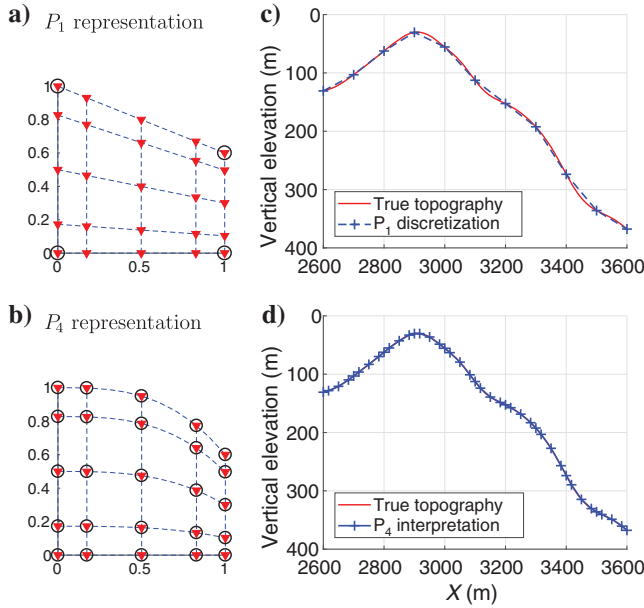

Figure 2. Left: Position of GLL points (the red triangles) of a cross section inside a 3D element for interpolation order $N = 4$ using (a) the eight corners of each element and linear shape function and (b) the $(4+1)^3$ GLL control points associated with P_4 shape functions. The control points are highlighted by the black circles. Right: The associated with topography description (for a cross section extracted from SEAM II Foothills model) using (c) P_1 and (d) P_4 representation. The positions of control points are marked by the sign "+," showing that the P_4 representation provides a better representation of the complex topography. The element size is 100 m for both cases.
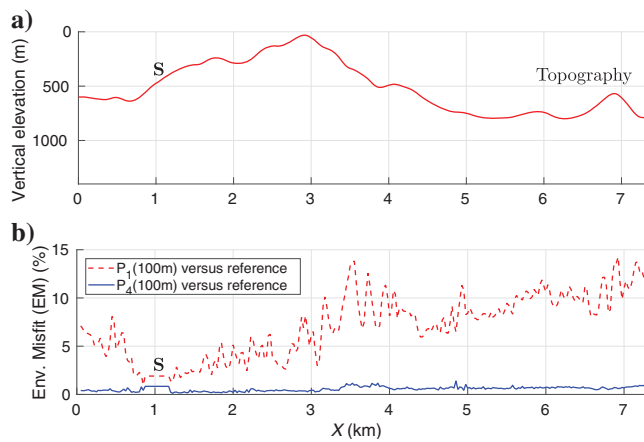
Figure 3. (a) A complex topography extracted from the SEAM II Foothills model. (b) Envelope misfit comparison: red — between the P_1 (100 m) mesh with the reference solution and blue — between the P_4 (100 m) mesh with the reference solution, which shows significant improvement in the numerical accuracy. The reference solution is obtained from the simulation on P_1 (25 m) mesh.

A single-valued envelope misfit estimated at each receiver location (Kristeková et al., 2006) is used to compare the observed seismograms with the reference solution, computed from a 25 m mesh under P_1 topography representation. The comparison is shown in Figure 3b: For the same number of elements, thus the identical numerical cost, the P_4 surface representation provides a more accurate estimation of the wavefield. Moreover, the simulation error of the P_1 (100 m) mesh is accumulated with offset, which might damage the information at far offsets, thus the velocity estimation of the near surface during the inversion. Trade-offs between topography representation and velocity parameters in acoustic land-FWI have been investigated numerically by Huiskes et al. (2017). When considering surface waves, the influence of topography representation on elastic FWI is expected to be even more important.

It should be noticed that only the volumetric Jacobian matrix associated with the mapping from the reference space to the deformed space is required for the wave propagation. Using high-order geometric representation (P_n) therefore only affects the mesh building step and the computation of the Jacobian, which are computed only once in the FWI workflow. Moreover, similar optimized strategy as the matrix-free spatial-derivative estimation in the modeling kernels section can be applied for the computation of the volumetric Jacobian. The computational cost of the wavefield modeling is thus unaltered, while the simulation accuracy related to the complex wave phenomena at the free surface is significantly improved. The combination of the variable element-size mesh design with high-order topography representation makes possible using as coarse elements as possible without degrading the medium geometries' representation, then providing a good balance between computational cost and numerical accuracy. We are aware that such high-order shape functions might affect the SEM convergence properties. However, our numerical experiments in Figure 3 confirm a rather low simulation error compared with a reference solution. In addition, it is widely known that the convergence characteristics of SEM are already lost when solving the wave equation in a heterogeneous medium inside each element, which is the cases of majority practical applications.

Parallel implementation

Our implementation relies on a two-level MPI-based parallelization. The inner level is designed on a Cartesian-based domain decomposition, which ensures an efficient load-balancing and an easy implementation. In particular, this avoids the use of a third-party mesh partitioner, even if the number of possible subdomains should be constrained by the mesh splitting in each direction. This parallel level is quite standard and efficient in SEM as gathering information through MPI communications is equivalent to the assembly procedure of FE schemes. This communication involves a single-depth "layer" of degree of freedom at each domain interface whatever is the SEM interpolation order, different from FD schemes. The outer MPI level is performed over seismic shots managed in parallel. This level is embarrassingly parallel because the communications steps only involve the summation of the misfit function and gradient contributions of each seismic shot and the scatter of physical parameters models and descent directions. Practical applications show that this second MPI-based level generally satisfies the perfect theoretical efficiency. It has to be noted that, compared with other SEM implementations relying on a single shot framework, our framework generally involves a large number of shots in parallel. The domain

decomposition per shot thus contains few tens of CPU cores, leading to a strong memory and computational load per core. In practical exploration FWI-liked applications, the communication between subdomains is negligible compared with the computation on each domain. The asynchronous communication would thus only provide a marginal saving.

The parallel efficiency of our implementation has been assessed on different hardware architectures (Intel CPU-based) with different interconnect technologies, showing very good properties. An example of strong scaling for the domain decomposition level is conducted on a benchmark with $192 \times 192 \times 192$ elements (4.5×10^8 degrees of freedom [dof]). The efficiency is assessed from an architecture embedding Intel Omni-Path interconnect and Intel Xeon E5-2697 v4 nodes (36 cores/node). Figure 4a shows the efficiency for intranode parallelization. As usual for PDE-based numerical schemes, the memory bandwidth appears to be a strong bottleneck: For the low number of MPI processes, we benefit from the whole memory bandwidth leading to artificially good performance (scalability larger than 90% up to eight process). When the number of MPI processes is increased, memory-bandwidth-bound effects appear, leading to an efficiency decrease (down to 60% for 32 process). This effect is directly associated with the memory-bandwidth bound, and when the hardware is fully used as in production, only the 32 processes number should be considered as a reference. In Figure 4b, the strong scaling is pushed up to 512 MPI process, showing stable and good scaling property, greater than 94%.

Nonlinear optimization with the SEISCOPE optimization toolbox

The inversion kernel relies on the SEISCOPE optimization toolbox (Métivier and Brossier, 2016). The library embeds various nonlinear optimization methods, coupled with the same line-search algorithm satisfying the Wolfe conditions (Nocedal and Wright, 2006). The library works through a reverse-communication interface, requesting the user to provide misfit function and gradient, and potentially preconditioner application. The optimization kernel is therefore fully included in the FWI main code, without any upper level wrapper.

Balancing memory requirement and computational cost for viscoelastic gradient building

FWI gradient estimation, required as the input of the optimization process, involves the zero-lag crosscorrelation in time between the incident and adjoint wavefields, weighted by the spatial radiation pattern (equations 18–20). The adjoint-state approach is thus considered for the gradient estimation, which requires access to the incident and adjoint wavefields at the same time (Plessix, 2006). We might store the incident wavefield in the core-memory or out-of-core memory at the expense of the I/O cost (Boehm et al., 2016). To avoid a heavy memory requirement, the incident field can be recomputed from the last snapshot and the boundary wavefield simultaneously with the propagation of the adjoint field. For elastic medium, the incident field can be perfectly reconstructed by the reverse propagation in time from

the last snapshot and the stored wavefield at the boundaries, synchronously with the propagation of the adjoint field backward in time. Gradients are directly accumulated during this process, resulting in a cheap operation with limited I/O operations (Dussaud et al., 2008).

Typical SEM simulation requires several time steps on the order of 10^4, which might lead to significant memory requirement for storing the boundary wavefields. Our implementation relies on compressed boundary storage to mitigate the memory cost (Yang et al., 2016c). The time stepping given by the CFL condition is generally much smaller than the Nyquist requirement, and the compression technique can be quite efficient: During forward propagation, the boundary wavefield is saved at every r time step ($r \gg 1$), where r is the decimation ratio. When reconstructing the incident field, at each time step, the boundary wavefield is interpolated from the decimated time series through Lagrange polynomial interpolation. Numerical experiments show that only the displacement field $\mathbf{u}$ (and not its spatial derivatives, nor time derivatives) needs to be saved at boundaries, after the prediction phase in the viscoelastic Newmark time-marching scheme (Algorithm 1), for an accurate reconstruction of the incident displacement $\mathbf{u}$ and velocity $\mathbf{v}$ fields.

In an anelastic medium, propagation of the adjoint field backward in time (equation 16) is as numerically stable as the forward propagation of the incident fields (equation 12) because they share the same equations. However, the irreversibility of the viscoelastic wave equation makes the reverse propagation of the incident field unstable, which requires specific strategies for efficient gradient accumulation (Tarantola, 1988; Griewank and Walther, 2000; Liu and Tromp, 2008; Komatitsch et al., 2016). To overcome this issue, we implement a viscoelastic version of the CARFS algorithm (Yang et al., 2016b), which makes a smart decision between the reverse modeling and the forward modeling using checkpoints based on the strain-energy measure. Similar to standard checkpointing techniques, the checkpoint distribution is computed from the binomial law, the optimal number of checkpoint being related to the total number of time steps nt by $\log_2(nt)$. For each checkpoint, we store the displacement, velocity, and memory variable fields $\mathbf{u}^{it}, \mathbf{v}^{it}, \boldsymbol{\psi}_\nu^{it-1/2}$. During the forward propagation of the incident field (Algorithm 2), at each time step, a reference global strain energy is also recorded

$$E_{\text{ref}}^{it} = \frac{1}{2}(\boldsymbol{\sigma}^{it}\boldsymbol{\varepsilon}^{it})_\Omega. \quad (26)$$

This global measure is used for monitoring the stability of the reverse simulation of the incident wavefield based on a predefined energy
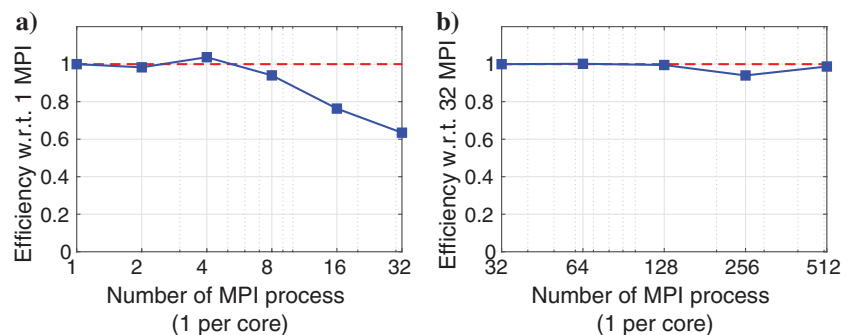

Figure 4. Strong scaling of the domain decomposition MPI level for a $192 \times 192 \times 192$ element application, on an Intel Omni-Path interconnect and Intel Xeon E5-2697 v4 node (36 cores/node) architecture: (a) intranode efficiency and (b) internode efficiency, showing stable and good scaling property.

tolerance. In our application, we use tol = 0.1, as shown in Algorithm 3. Above this tolerance, the nearest earlier in time checkpoint is considered. A forward simulation from this checkpoint until the current time level is performed. During this forward simulation, the checkpoints are also redistributed. Once done, the reverse simulation of the incident wavefield together with the propagation of the adjoint wavefield backward in time can be continued.

The combination of this compression technique through decimation and interpolation with the CARFS approach provides a good balance between the memory requirement, simulation accuracy, and the computational cost, which is directly linked to the recomputation ratio of the incident field. Additional compression strategies could also be used to further reduce the memory requirement and I/O requests (Boehm et al., 2016).

Structure-oriented Bessel preconditioning

In practical applications, the gradient vector $\mathbf{g}(\mathbf{x})$ can exhibit unconstrained or unrealistic high-wavenumber components, which are

---

Algorithm 2. Forward propagation.

---

1. **for** $it = 1, \ldots, nt$ **do**
2.   Forward propagation of the incident field: $F^{it} : \mathcal{W}^{it-1} \rightarrow \mathcal{W}^{it}$
3.   Record the reference strain-energy E_{ref}^{it} for $\mathcal{W}^{it}$
4.   **if** (At the checkpoint position) **then** Store the snapshot $\mathbf{u}^{it}, \mathbf{v}^{it}, \boldsymbol{\psi}_\nu^{it-1/2}$
5.   **if** (At the decimated location: $mod(it,r)=0$) **then** Store the boundary of the displacement field $\mathbf{u}^{it}$
6. **end**

Algorithm 3. Gradient estimation.

---

1. **for** $it = nt - 1, \cdots, 1$ **do**
2.   Adjoint time-step: $at = nt - it$
3.   Forward propagation of the adjoint field: $F^{at} : \bar{\mathcal{W}}^{at-1} \rightarrow \bar{\mathcal{W}}^{at}$
4.   **if** (At the checkpoint position) **then**
5.   Read the snapshot $\mathbf{u}^{it}, \mathbf{v}^{it}, \boldsymbol{\psi}_\nu^{it-1/2}$
6.   **else**
7.   Interpolate the boundary of the displacement field $\mathbf{u}^{it}$
8.   Backward propagation of the incident field: $(F^{it})^{-1} : \mathcal{W}^{it+1} \rightarrow \mathcal{W}^{it}$
9.   Measure the energy of the reconstructed wavefield E_{rec}^{it}
10.   **if** (Instability: $|E_{\text{ref}}^{it} - E_{\text{rec}}^{it}| > \text{tol} \cdot E_{\text{ref}}^{it}$) **then**
11.   Forward propagation from the closest checkpoint prior to the current time level it
12.   Replace the used checkpoints by new snapshot positions
13.   **end**
14.   **end**
15.   Gradient collection (equations 18–20);
16. **end**

---

incompatible with the intrinsic resolution of FWI. Often, these undesirable features of the misfit gradient are removed through convolution filtering with Gaussian or Laplace kernels. Moreover, by incorporating some geologic prior knowledge, this gradient smoothing can also be used for constraining the inversion (Guitton et al., 2012). With the same objective of removing artifacts and incorporating available prior knowledge of geologic information, we have introduced a nonstationary, anisotropic Bessel smoothing filter $B_{\text{3D}}(\mathbf{x})$ based on the modified spherical Bessel function of the second kind of order zero (Abramowitz and Stegun, 1972; Trinh et al., 2017b). The anisotropic filter shape is defined by variable coherent lengths: L_v is associated with the direction perpendicular to the local bedding plan; L_u and L_w are related to the planar structure of potential geologic features. The 3D orientation is controlled by azimuth θ and dip φ angles. The azimuth is the horizontal angle measured from the y-axis, and the dip is the angle created by the oriented plan with the horizontal direction. Their values range such that $\theta \in [-\pi, \pi]$ and $\varphi \in [-\pi/2, \pi/2]$.

Instead of convolving the original vector $\mathbf{g}(\mathbf{x})$ with the forward filter $B_{\text{3D}}(\mathbf{x})$ to get the smoothed vector $\mathbf{s}(\mathbf{x})$, we solve the following equation relying on the sparse inverse operator

$$B_{\text{3D}}^{-1}(\mathbf{x}) * \mathbf{s}(\mathbf{x}) = \mathbf{g}(\mathbf{x}). \qquad (27)$$

Due to the definition of the inverse operator (Trinh et al., 2017b), equation 27 is equivalent to a Helmholtz-like PDE, in which the original gradient $\mathbf{g}(\mathbf{x})$ appears in the right side. Under the assumption of the slow spatial variation of the filter parameters, their spatial derivatives can be neglected. Equation 27 governing the smoothing process can thus be approximated as

$$\mathbf{s}(\mathbf{x}) - \nabla_{z,x,y}^T \mathbf{P}(\mathbf{x})\mathbf{P}^T(\mathbf{x})\nabla_{z,x,y}\mathbf{s}(\mathbf{x}) = \mathbf{g}(\mathbf{x}), \qquad (28)$$

where the operator $\nabla_{z,x,y}$ is the spatial derivatives $(\partial/\partial z, \partial/\partial x, \partial/\partial y)^T$ and the upper symbol "T" stands for the transposed operator. The information related to the geologic variation of the medium (i.e., filter parameters) is embedded in the matrix

$$\mathbf{P}(\mathbf{x}) = \begin{bmatrix} L_v \cos \varphi & L_u \sin \varphi & 0 \\ -L_v \cos \theta \sin \varphi & L_u \cos \theta \cos \varphi & L_w \sin \theta \\ L_v \sin \theta \sin \varphi & -L_u \sin \theta \cos \varphi & L_w \cos \theta \end{bmatrix}, \qquad (29)$$

which can be interpreted as the projection between the Cartesian space and the locally rotated dimensionless coordinates system.

Following the weak formulation of SEM, equation 28 naturally yields a symmetric, positive definite, and well-conditioned linear system

$$(\mathbf{M}_b + \mathbf{K}_b)\mathbf{s} = \mathbf{M}_b\mathbf{g}. \qquad (30)$$

The mass matrix $\mathbf{M}_b$ associated with the application of Bessel filter is diagonal by construction, and the stiffness matrix $\mathbf{K}_b$ is symmetric by definition. We solve this linear system 30 through a parallel conjugate gradient iterative solver, using the same mesh and optimized kernels developed for the wave equation. The most expensive operator is the product of the sparse stiffness-matrix $\mathbf{K}_b$ with a given vector. Again, the factorization of this matrix

$$\mathbf{K}_b = \mathcal{D}^w(\mathbf{P}\mathbf{P}^T)\mathcal{D}, \qquad (31)$$

is used to achieve an efficient implementation (Deville et al., 2002). The overall scheme has a linear algorithmic complexity with respect to the coherent length ($\mathcal{O}(L)$), to be compared with the complexity $\mathcal{O}(L^3)$ for a standard windowed-convolution-based filter. In practical applications, the smoothing process can be performed at a negligible fraction of the computational cost of the forward/adjoint problems (less than 1%). A cascaded application can be used to produce other filter decays; for example, a double application of Bessel operators provides an accurate approximation of the Laplace filter (Trinh et al., 2017b). Illustrations on the gradient smoothing are shown in "3D case study" section.

NUMERICAL INVESTIGATIONS

The simulation quality of our elastic and viscoelastic forward modeling is illustrated through the comparison with semianalytical solutions in Appendix C. In this section, we show that our approach exhibits similar overall computational efficiency as some open-source FD and SEM schemes. We shall also illustrate the effect of the second term in the viscoelastic C_{IJ} gradient mentioned in equation 20.

Comparison with other open-source codes for the modeling part

For 3D elastic and viscoelastic seismic modeling, it is important to check relative performance of this newly designed code SEM46 with two well-established open-source programs: SW4 V1.1 and SPECFEM V2.0 (CPU version). SW4 uses a fourth-order FD approach (Sjögreen and Petersson, 2012), whereas SPECFEM V2.0 relies on a SEM approach (Peter et al., 2011). We choose SPECFEM V2.0 instead of SPECFEM V3.0 because we want to perform our comparison on CPU: SPECFEM V3.0 has been mainly optimized for GPU platforms (Komatitsch et al., 2010). The three programs are compiled with Intel 2015 compiler and optimized option-O3 on a Intel Sandy Bridge EP E5-2670 2.6 GHz with 64 Gb of RAM, Infiniband FDR nonblocking computing network.

We use a homogeneous model of $12 \times 12 \times 12$ km for the comparison. For viscoelastic comparison, in each code, quality factors are set as $Q_p = Q_s = 40$, described by three SLSs. In most practical applications, SW4 recommends to use at least six points per shortest wavelength for an acceptable accuracy, leading to the grid sampling at 25 m (Petersson and Sjögreen, 2013). Therefore, the FD grid contains $481 \times 481 \times 481$ points. Because approximately six GLL points per shortest wavelength allows us to accurately model elastic waves with P_4 spectral elements, the SEM mesh is filled with $120 \times 120 \times 120$ elements at 100 m size and the interpolation order $N = 4$ is used for SEM46 and SPECFEM. This setting leads to the same number of spatial points (1.1×10^8) for the FD grid and SEM meshes.

The total recording time is 6 s leading to 908 time steps in SW4 and 4000 time steps in SEM46, due to the CFL condition. A similar configuration is considered for SPECFEM with identical number of dof and time steps as for SEM46. Figure 5 compares the computational cost of SEM46 with the one of SW4 and SPECFEM for parallelism over different domain decomposition settings. Even with much larger number of time steps, SEM implementations exhibit comparable computational cost as SW4 for elastic and viscoelastic modeling. In this comparison, SEM46 uses a triclinic elasticity

coupled with an isotropic attenuation implementation, whereas isotropic elasticity and isotropic attenuation are considered in SPECFEM. The anelastic effect is described by independent Q_p and Q_s quality factors in SEM46. Only isotropic elasticity and shear quality factor Q_s is taken into account by the version of SPECFEM V2.0 we use, which somehow simplifies the numerical implementation. Even though our SEM46 implementation exhibits comparable computational performance as the ones of SPECFEM V2.0 for the forward 3D elastic and viscoelastic simulations, under the same computer and compilation environment. The efficiency of these different codes might be improved by possible local code rewriting, but overall they require similar, if not identical, computer resources. More quantitave benchmarking will require a careful analysis of numerical errors and interaction between teams who have developed these codes: This is not our present target.

Influence of the attenuation on the velocity gradient estimation

In equation 20, we have intensively discussed the expression of the viscoelastic C_{IJ} gradient

$$\frac{\partial \chi(\mathbf{m})}{\partial C_{IJ}} = \left(\bar{\boldsymbol{\varepsilon}}, \frac{\partial C}{\partial C_{IJ}} \boldsymbol{\varepsilon} \right)_{\Omega,t} - \left(\bar{\boldsymbol{\varepsilon}}, \sum_{\nu=1}^{L} \frac{\partial C^a}{\partial C_{IJ}} \boldsymbol{\psi}_\nu \right)_{\Omega,t}. \qquad (32)$$

It should be noticed that the second term has never been explicitly reported before, although correctly handled by Fichtner and van Driel (2014). To our understanding, it is generally ignored in practical applications (Liu and Tromp, 2008; Komatitsch et al., 2016). Its magnitude is small in a weakly attenuating medium compared with the first-term contribution. However, its influence should increase with the attenuation, especially at far offsets. Therefore, in a highly attenuating medium, ignoring the second term might affect the model parameter estimation.

The contribution of the second term in equation 32 is illustrated on a simple toy example. We use a simple setting with homogeneous background $V_P = 2300$ m/s, $V_S = 1500$ m/s, $\rho = 1000$ kg/m³,
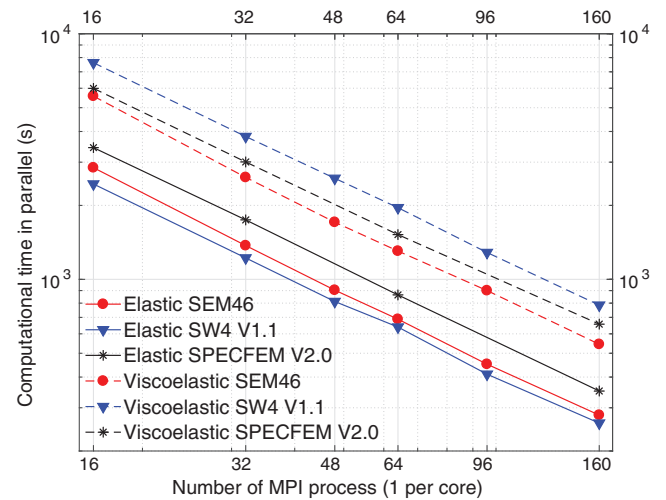

Figure 5. Numerical efficiency compared with SW4 V1.1 and SPECFEM V2.0 (CPU-based) packages over different domain decomposition.

and a V_P inclusion in the center of the model as shown in Figure 6a. The medium is strongly attenuating with $Q_p = Q_s = 40$.

In Figure 6a, the medium is illuminated by four sources, highlighted by the red triangles. For each source, a plan containing 6241 receivers is located at the opposite side of the inclusion; for example, when the source position is $x_s = 250$ m, the receiver plan is at $x_r = 6750$ m. Free-surface conditions are applied at all faces of the model to increase the illumination through the reflection at the boundary of the model. The source time function is a Ricker wavelet, centered at 10 Hz. With this setting, the distance between sources and receivers is approximately 30 dominant P-wavelengths. Because a very sparse acquisition is used in this toy example, we

a) True V_P model

b) Two-term V_P gradient

c) Gradient without the second term

d) Gradient difference Δ (**c** - **b**)
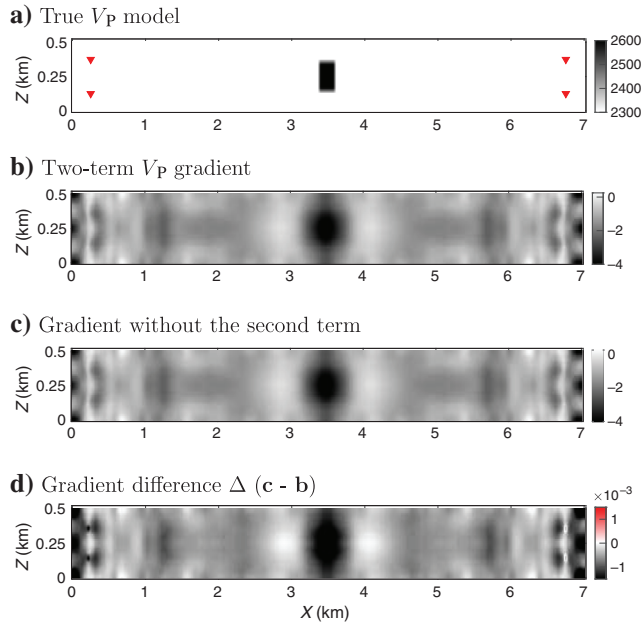
Figure 6. (a) Homogeneous background with a V_P inclusion in the center of the model. Source positions are highlighted by the red triangles. (b) First viscoelastic FWI V_P gradient computed from elementary C_{IJ} gradients in equation 32. (c) First FWI gradient without considering the second term. (d) Difference between gradients (c and b), which highlights the contribution related to the second-term gradient.
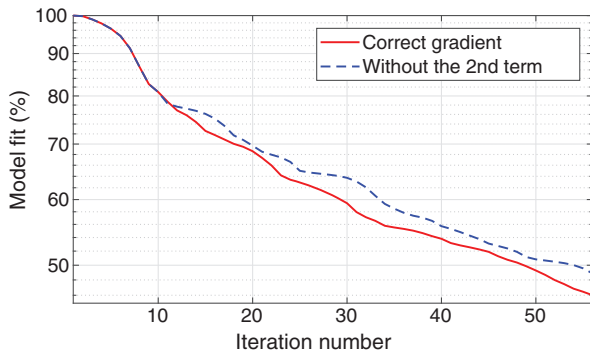
Figure 7. An example of the model fit at the inclusion position between the true V_P and the inverted model using the correct gradient estimation (red) and without considering the second term (blue), which highlights the accumulative effect on model parameter estimation.

use a long recording time (9 s) so that the medium can be sampled by a multiple reflected wavefield. We use a homogeneous initial model $V_P = 2300$ m/s. Figure 6b shows an estimation of the first V_P gradient computed from elementary C_{IJ} gradients in equation 32. The inversion problem is ill-constrained leading to significant artifacts in the gradients. However, we only apply a minimal Bessel smoothing to remove the numerical artifacts above the Nyquist wavenumber associated with the underlying mesh.

Figure 6c shows the V_P gradient estimation when ignoring the second term in the elementary C_{IJ} gradients. The difference between the gradients (c and b) in Figure 6d highlights the small amplitude modification at the inclusion position and the places intensively sampled by the wavefield. It should be noted that without the second term, the gradient (c) is underestimated at the inclusion position, implying a noncorrect estimation of the model update. On this setting, we run a FWI test with 60 iterations in which the model fit at the inclusion position is shown in Figure 7. The model fit decreases faster when considering the correct gradient. The comparison effectively shows that ignoring the second term in the gradient estimation might affect the reconstruction of the model parameter. The effect manifests at later stages, when the shape of the target has been found, and the inversion attempts to recover the correct amplitude of model parameter.

In practical applications, these effects might be negligible compared with other sources of uncertainties. However, the correct gradient can be collected without any extra cost because the incident adjoint fields ψ_ν are available during the simultaneous forward and backward propagations of the adjoint and the incident fields (Algorithm 3).

3D ELASTIC CASE STUDY

In this section, we apply our FWI frame on an isotropic elastic example coming from a 3D subtarget of the SEAM Phase II Foothill model (Oristaglio, 2012, 2016; Regone et al., 2017) to highlight the importance of each element which have been presented.

Model description and 3D acquisition design

The topography variation is significant for this target, with a maximal vertical variation of 800 m as shown in Figure 8. The bedding
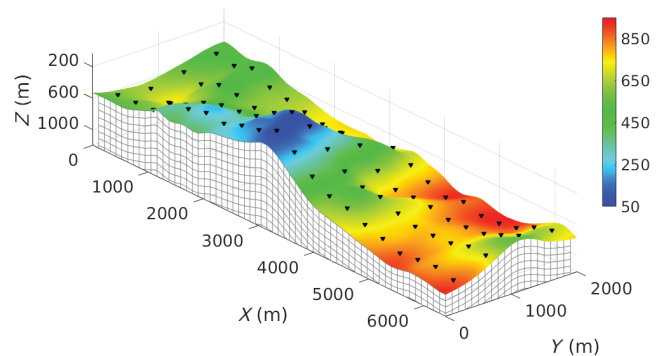

Figure 8. The 3D acquisition design superimposed on the topography and the underlying mesh, in which the colorbar illustrates the absolute depth Z_{abs} (m) from a predefined zero depth: The source positions are marked by the red triangles $\Delta S_x = 320$ m and $\Delta S_y = 500$ m. For each source, 3C receivers are deployed in the whole surface with $\Delta R = 12.5$ m.

plans are gently dipping in the x-direction with folding structures, which can be visualized from the depth section in Figure 9a, 10a, and 11a. The model also has an unconformity at 2–3 km depth, which might be difficult to recover through FWI. Velocity attributes extracted at 20 m below the free surface in Figure 12a show interesting features with complex geometry that FWI aims to reconstruct.

We use a 3D surface acquisition with four source lines, each line including 20 sources. The source positions are indicated by the red triangles in Figure 8, with inline and crossline source-spacing ΔS_x and ΔS_y taken at 300 and 500 m, respectively. For each source, a grid of 3C receivers is deployed on the whole surface, the distance between two adjacent receivers being 12.5 m. We use a vertical point force. The source-time function is a Ricker wavelet, centered at 3.5 Hz. The total recording time is equal to 6 s.

Inversion setup

SEM is used for the forward and inversion problem. The observed data are generated with a constant element-size mesh, whereas the inversion problem is computed over a variable element-size mesh, locally satisfying the volume condition. A sharp density model, as the velocity models in Figure 11a, is considered for the observed data generation. As illustrated in Figures 2c and 3, the P_4 representation should be used to describe rapid topography variations of the surface, especially for the variable element-size mesh. Because the sources and receivers are located at the free surface, the P_4 representation also helps to correctly position sources and receivers. As shown in the observed-data panel in Figure 13, the seismic wavefield is complex, including highly dispersive surface waves due to the rapidly varying topography. Significant backscattering of body and surface waves as well as mode conversion at steep-slope surface positions can also be observed, for example, at $X = 3$ km (the associated V_s cross section is presented in Figure 9a).

The initial V_P and V_S models that we use (Figures 11b and 12b) are smoothed versions of the true model. A similar smoothed model is used for density as the input of the inversion process. To investigate on different aspects of elastic FWI, such as gradient preconditioning and data-windowing hierarchy, we choose here a set of starting models compatible with the frequency-content of the data: The calculated data in the initial models are not cycle skipped compared with the observed data, as shown in Figure 13a. We invert simultaneously for V_P and V_S, and the density is kept unchanged, as a passive parameter. Each inversion sequence consists of 60 iterations of the limited-memory Broyden-Fletcher-Goldfarb-Shanno (l-BFGS) optimization method. We do not apply other preconditioning or regularization than the Bessel gradient smoothing as detailed below.
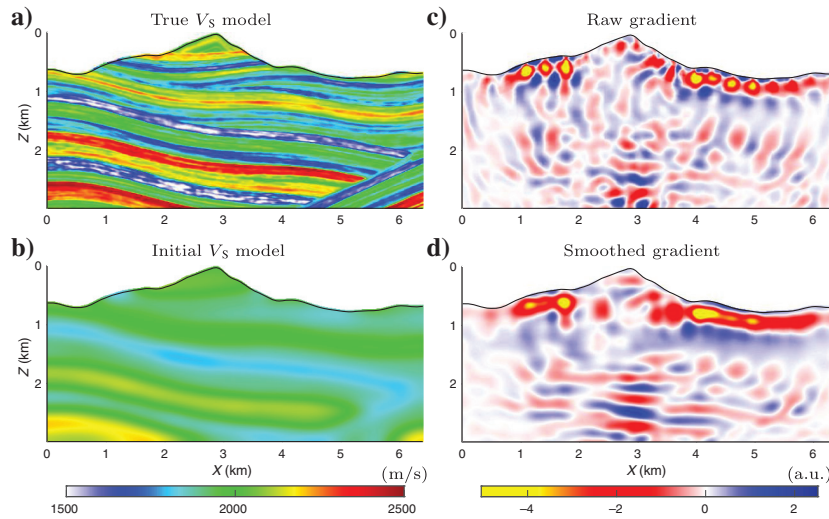

Figure 9. Gradient smoothing illustration underneath the source line at $Y = 750$ m: (a) true velocity model, (b) initial model, (c) original scaled gradient without any smoothing, showing significant acquisition footprint and artifacts due to a sparse acquisition. (d) Smoothed gradient by Bessel filter in which the parameters are described in Figure 10b and 10c, which clearly show the dip and azimuth filtering effect.
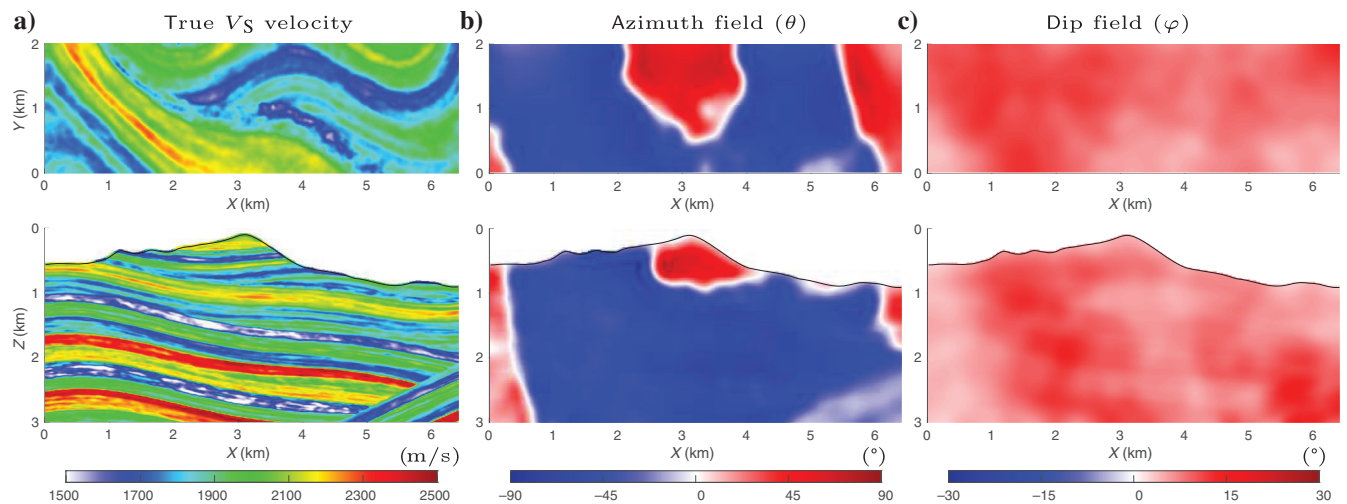

Figure 10. Top: depth section (xy-view) at $z = 1$ km, bottom: cross section (zx) at $Y = 100$ m. (a) True V_S velocity model with gentle dipping in zx-section and folding in xy-view. (b) Smooth azimuth field that nicely follows the folding the structures in xy-view. (c) Smooth dip field that is almost homogeneous.
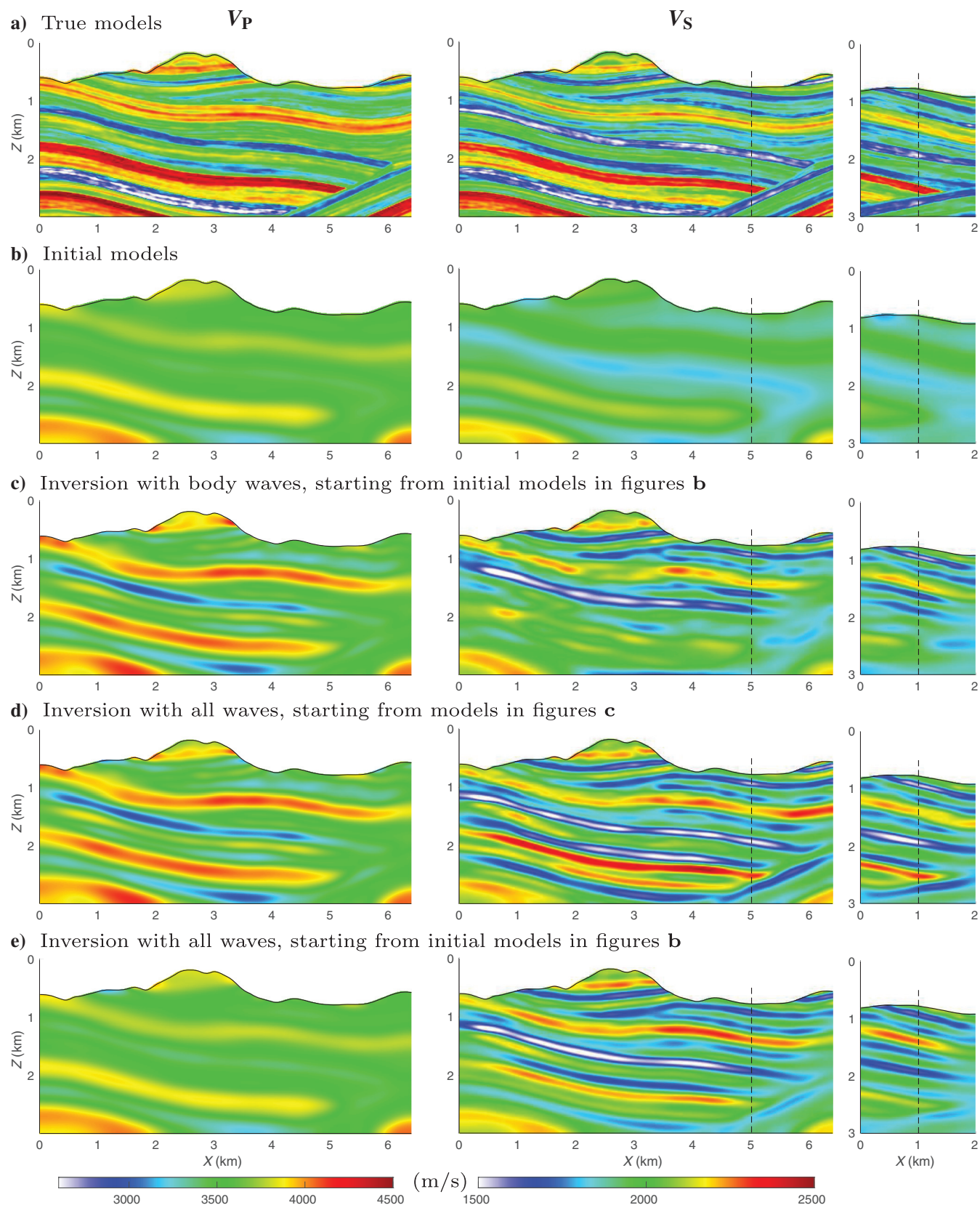
Figure 11. Left: V_P velocity cross section at $Y = 1$ km, middle: V_S cross section at $Y = 1$ km, and right: V_S cross section at $X = 5$ km. (a) True models, (b) initial models, and (c) inversion results by only using the early-body waves, starting from the initial models. (d) Inversion results by using all the wavefield, starting from models in (c), showing significant improvement for the V_S estimation. (d) Inversion results by using all the wavefield, starting from the initial models — no significant update in V_P estimation.

Structure-oriented gradient preconditioning

The raw gradient computed on the initial model is presented in Figure 9c, showing significant acquisition footprint at the near surface. Unrealistic oscillations also occur at greater depths, due to the limited illumination coming from the sparsity of the sources. However, the targets have considerable dipping and folding structures, which can be characterized by smooth azimuth and dip fields as shown in Figure 10b and 10c. Because the structures vary quicker in the vertical direction than the horizontal directions, we design a highly anisotropic filter shape in which the vertical coherent length L_v being 25 m, much smaller than for other directions. Due to the presence of an unconformity, the horizontal coherent lengths L_u and L_w increases from 25 m near the discontinuity position to 100 m elsewhere. By doing so, the filter has an isotropic shape near the unconformity to avoid any smearing effect across the unconformity. The anisotropic coherent lengths design together with the 3D rotation result in the oblate-spheroid shape of the filter volume, with the major plan toward the dipping direction of the bedding plan. In practical applications, the filter design can be based on an interpretation of the migrated image computed from the initial models or any geologic prior information about the medium such as well logs and geologic description.
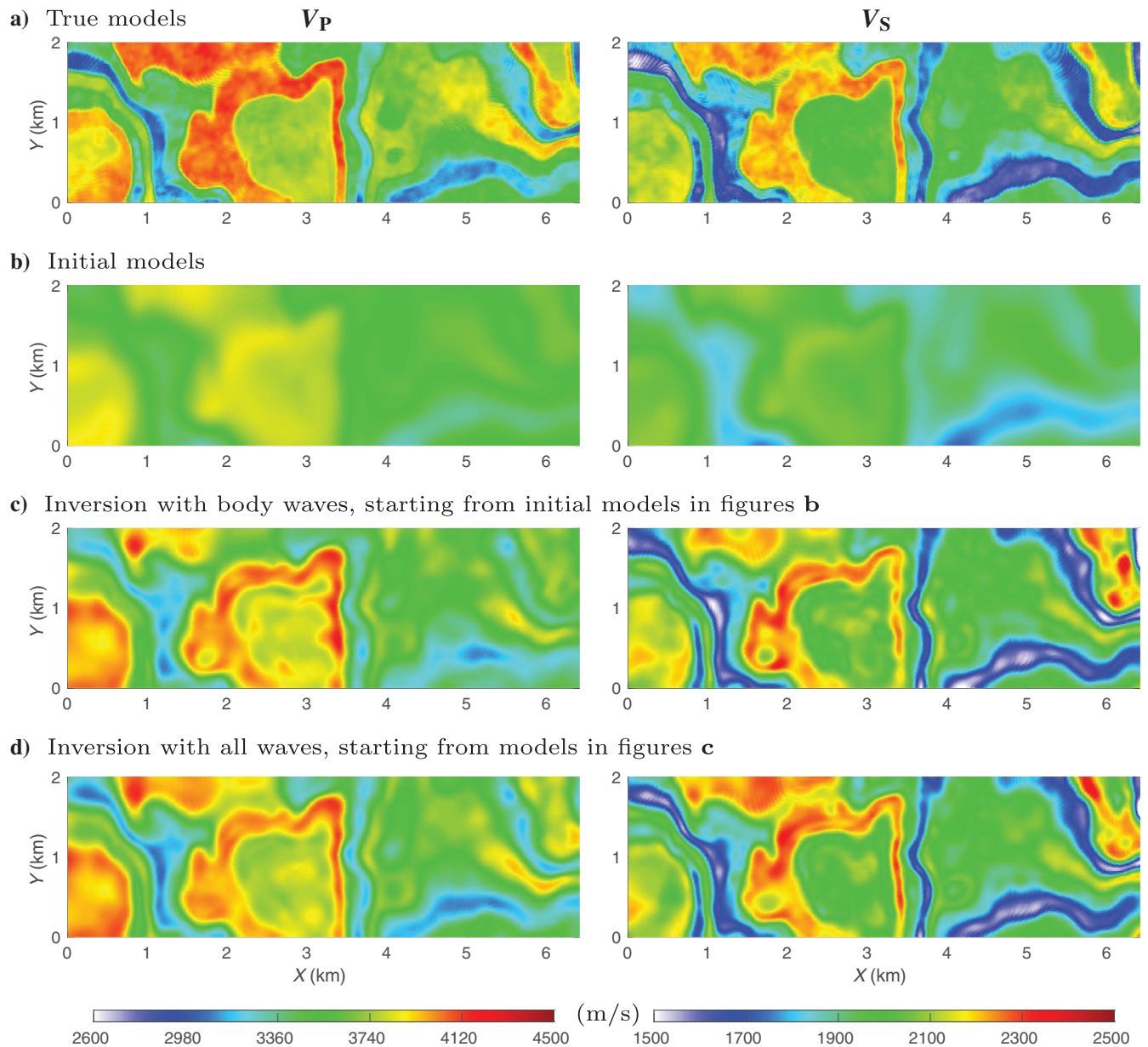
Figure 12. Velocity attributes extracted at the relative depth $\Delta Z = 20$ m from the free-surface. Left: V_P and right: V_S. (a) True models with interesting channels, (b) initial models, and (c) inversion results by only using the early-body waves, starting from the initial models. (d) Inversion results by using all the wavefield, starting from models in (c), showing that surface waves can bring additional information to improve inversion results at near-surface for V_P and V_S.

By considering these filter parameters in the structure-oriented Bessel preconditioning, we obtain the smoothed gradient in Figure 9d: The artifacts due to the acquisition footprint and the poor illumination are effectively reduced, without degrading the deeper structures. The continuity of the features at greater depths is enhanced because the oscillations are weaken. We also obtain a correct orientation of the geologic features. In this example, the smoothing process costs only 0.4% running time of one FWI iteration. The structure-oriented Bessel gradient precondition described above is systematically applied into any inversion sequence in the next section.

Data-windowing hierarchy

Due to the rough topography, the seismic wavefield is quite complex with dispersive surface waves and significant back-scattering waves as shown in Figure 13. A two-step data-windowing hierarchy is considered to mitigate the dominance of the V_S parameter over V_P, due to the presence of surface waves. The models are first inverted with early-body waves, arriving before the surface waves. The observed models will then be used as starting models for the inversion will the entire wavefield. During the first step, a bottom mute is used to separate the early-body waves with the surface waves and underlying reflected and back-scattering waves.
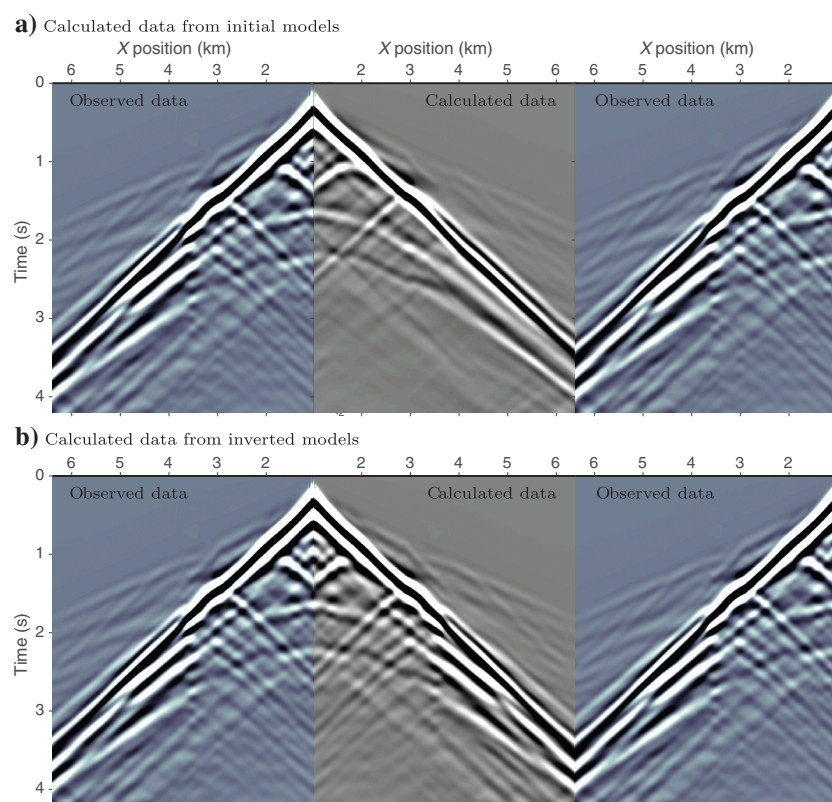
Figure 13. Data comparison at the shot line $Y = 750$ m between the observed data and calculated data: (a) $\mathbf{d}_{cal}$ estimated from initial models in Figures 11b and 12 and (b) $\mathbf{d}_{cal}$ computed from inverted models in Figures 11b and 12. Final inverted models significantly improve the data fit for the back-scattering, body, and surface waves. The inversion also recovers the body-wave amplitude, related to the radiation efficiency of the source from model perturbation.

Inversion results

The cross sections of the results obtained after the first step are shown in Figure 11c. The inversion successfully recovers main structures of the V_P model. The reconstruction of the V_S model is limited at 2 km depth due to the shallow penetration of the shear component. The models presented in Figure 11d and 12d are obtained after the second step, accounting for the whole data set. Adding surface waves and other parts of the wavefield does not degrade the V_P estimation and improve significantly the V_S model. The continuity of the near-surface features are strengthened, and the deeper structures are better resolved. The unconformity is well-reconstructed in the V_S model at the 2 km depth (Figure 11d, the middle and right panels). This is in part due to the nonstationary design of the Bessel filter. When looking at the velocity attributes extracted at 20 m below the free surface (Figure 12c and 12d), inverting for surface waves actually improves the thin-structure imaging and amplitude estimation for V_P and V_S.

Discussions

Importance of the data-windowing hierarchy

Because the calculated data from the initial models are not cycle skipped compared with the observed data, one can run the inversion will the entire wavefield, without any distinction between body and surface waves. The inverted results are shown in Figure 11e, in which the V_S is well-reconstructed. However, we observe insignificant updates of the V_P model. Because the data are dominated by surface waves, the least-squares misfit function is mainly sensitive to V_S. The data-oriented strategy we use in this 3D example is thus crucial: By focusing on early body waves before considering the whole data, our data-windowing hierarchy makes it possible to better constrain the P-wave velocity.

Figure 13 shows the comparison of observed data with the calculated data computed from the initial models and the final inverted models (Figures 11 and 12d). It should be remarked that compared with the observed data, the initial models produce simple surface waves estimation and weak amplitude of the first arrival due to wrong velocities at the source location. After the two-step data-windowing hierarchy, the inversion successfully recovers details in the velocity models, resulting in an improved agreement with the observed data for surface, body, and back-scattering waves.

For complex real data applications, especially in foothill environment, this data separation between early body waves and surface waves might be difficult to achieve. In such a case, a model hierarchy could be used on top of this data-windowing strategy, in which Bessel gradient preconditioning and/or model regularization could be used to monitor the wavenumber-content of the allowed model updates. At the first step, only

low-wavenumber V_S updates would be allowed to enhance the V_P reconstruction.

Structure-oriented Bessel preconditioning enhances the model estimation

The importance of the gradient preconditioning is highlighted in Figure 9. If prior information about the structures such as the local 3D rotation is not available, a simple stationary anisotropic Bessel preconditioning could be enough to remove artifacts in the gradient (Trinh et al., 2017b), for example, with coherent lengths $L_v = 25$ m, $L_u = L_w = 100$ m, and 0° dip and azimuth angles. It should be repeated that the anisotropic design relies on the fact that geologic features vary quicker in the vertical direction than in the horizontal direction.

To highlight the importance of our filter design with variable coherent lengths and 3D rotation, Figure 14 shows the difference between inverted results with the nonstationary structure-oriented preconditioning and the inversion with the stationary filter, under the same inversion hierarchy and setting. The comparison shows nonneligible modification of the amplitude estimation for V_P and V_S, especially following the dipping directions. In particular, significant difference can also be observed at the unconformity because the nonstationary filter is reduced to isotropic shape near this location. The model update is thus not smeared out across the unconformity, leading to a sharp interface as shown in the final V_S model in Figure 11d. It should be noted that we only impose

the relative position of the unconformity, and we simply reduce the filter effect at this location. By doing so, the filter does not artificially sharpen the interface but let FWI free to reconstruct it.

Computer memory and elapsed time

The 3D elastic inversion study has been performed on 1600 cores, in which the computation for each shot being performed on 20 cores (i.e., 80 sources and only 20 subdomains per shot). The incident fields are reconstructed from the decimated boundary wavefield with a decimation ratio equal to five. The total memory requirement per shot is thus approximately 44 Gb. Each gradient computation in parallel is performed in 20 min, leading to almost 21 h for one FWI sequence of 60 iterations, as shown in Table 1. It should be noted that storing the incident fields at each time step would require approximately 889.5 Gb per shot without any compression.

On the same 3D models with similar numerical setting, we add homogeneous viscous properties with $Q_p = Q_s = 50$, which is the value of quality factors in the near-surface region of the SEAM II Foothills benchmark. To derive benefit from whole memory while optimizing the elapsed time, 80 checkpoints are stored and a decimation ratio equal to five is used for wavefields saving at the boundaries. The energy tolerance for monitoring the reconstruction of incident fields is set as 5% (Algorithm 3), leading to a recomputation ratio equal to 3.2. The waves propagation in anelastic medium is more expensive than the elastic cases: One gradient costs 1.36 h
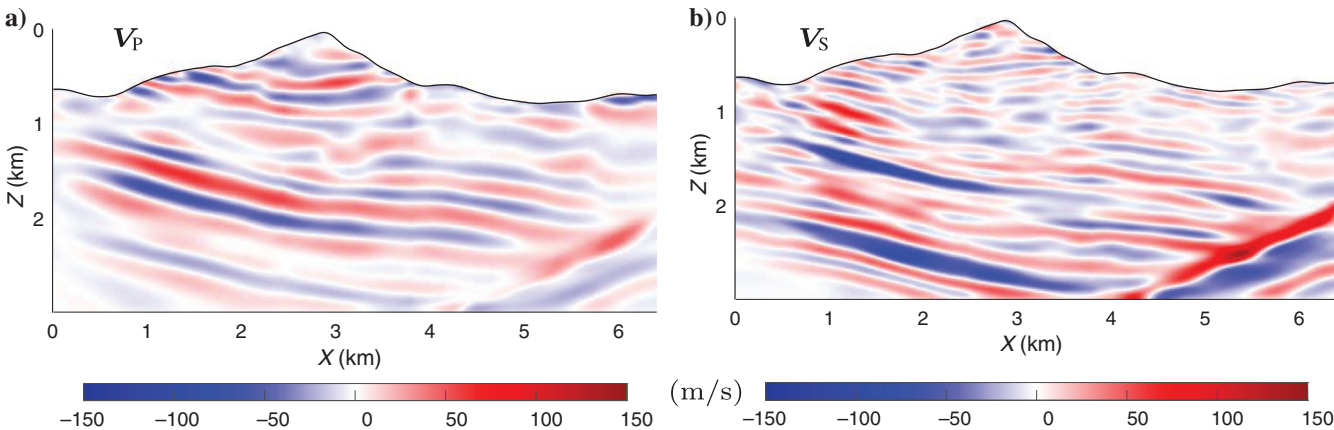

Figure 14. Difference between inverted results with structure-oriented preconditioning (Figure 11d) and the inversion results with stationary smoothing $L_v = 25$ m, $L_u = L_w = 100$ m, 0° dip and azimuth angles. (a and b) V_P and V_S cross sections at $Y = 750$ m. The rotational gradient preconditioning mainly affects the amplitude estimation.

Table 1. Top — memory requirement and computational cost for the elastic FWI test on the 3D subset of the SEAM Phase II Foothills benchmark. Bottom — estimation for 3D viscoelastic test with a similar setting.

	Total memory per shot	First gradient estimation	FWI (60 iterations)
3D elastic FWI example ($3 \times 7 \times 2$ km) — 1600 cores: 3.98×10^6 dof; 82.6×10^3 receivers per source; 10^4 time steps; decimation ratio = 5; mesh with all deformed elements. Inverting for V_P, V_S parameters	44 Gb	20 min	20.8 h
3D viscoelastic FWI test — 1600 cores: quality factors $Q_p = Q_s = 50$; incident fields reconstruction: 80 checkpoints, decimation ratio = 5, energy tolerance = 5% leading to recomputation ratio ≈3.2.	70 Gb	1.36 h	84.8 h (estimation)

in parallel, providing an estimation of 85 h for 60 iterations. In this setting, we obtain a factor 4 between elastic and viscoelastic inversion. It should be noted that storing the incident fields at each time step would require 3.5 Tb per shot, due to the fact that the C_{IJ} gradient requires access to strain and memory-variable incident fields (equation 20).

It should be noted that Table 1 provides estimations for a relatively small example. In realistic applications, even with variable-element mesh design, we expect to consider 10^8 dof with a longer recording time. In such applications, storing a large number or checkpoints or incident wavefields even at the Nyquist frequency in core memory is infeasible. One might use the out-of-core memory at the expense of the I/O cost. We rather focus on storing a small number of checkpoints in core memory with limited I/O request but higher recomputation ratio of the incident fields.

CONCLUSION

We provide a complete review of the 3D multiparameter FWI in viscoelastic medium, using SLS mechanisms. We separate purposely the elasticity with the attenuation mechanisms, which offers an easy and explicit coupling between the anisotropic elasticity and isotropic attenuation. The attenuation parameters are explicitly incorporated in the wave equations and therefore can be naturally reconstructed in the FWI framework. Even if the second-order viscoelastic wave equation is not self-adjoint, we can develop the adjoint system with similar equations as for the incident fields, implying that they can be propagated under the same numerical scheme. Following the forward and adjoint formulation using SLS mechanisms, the gradient expressions for density, attenuation parameters, and stiffness coefficients C_{IJ} are simply zero-lag crosscorrelations in time between incident and adjoint fields, weighted by the associated spatial radiation pattern. The gradient for any other parameter can be computed by chain rule from these elements. We report a two-term expression for the C_{IJ} gradient: The first term, related to the zero-lag crosscorrelation of the adjoint and incident strain fields, is commonly used in practical application. The second term can be interpreted as the time dependency of the perturbation of the relaxation rate. The influence of the second term is shown to have an accumulative effect on the amplitude of the model parameter estimation, depending on the physical domain and acquisition settings.

We present an integrated SEM-based workflow, capable of efficiently performing 3D modeling and FWI in the time domain for elastic and viscoelastic exploration-scale targets. Our approach relies on the trade-off between computational cost, accuracy and memory requirement. The simulation accuracy is ensured by the use of a Cartesian-based deformed mesh with high-order geometry interpolation to capture rough topography variations. The element size can be adapted to the variation of the local shortest wavelength to reduce the numerical cost. The Cartesian-based mesh also allows to access the spatial position and the neighbors of each element without any extra cost. Two MPI-based parallelism levels are considered for tackling large-scale and multiple-shot experiments, associated with an efficient computation and low-memory implementation of incident and adjoint fields. The simulation accuracy of our approach is illustrated through the comparison with elastic and viscoelastic semianalytical solutions via the boundary integral method. For the same physical model and similar numerical settings, we achieve comparable computational cost as two open-source codes: SW4 V1.1 and SPECFEM V2.0 (CPU version) for elastic and viscoelastic simulations.

For the gradient calculation, the incident field is reconstructed from the stored wavefield in the boundaries, synchronously with the propagation of the adjoint field. The gradient is directly accumulated during this process, resulting in a cheap operation. To reduce the memory requirement, the boundary wavefield is saved at decimated positions in time and reinterpolated when needed. In a viscoelastic medium, the instability in the reconstruction of the incident wavefield can be mitigated by using the elastic version of CARFS algorithm. This strategy is critical in realistic applications, normally associated with a large number of dof and time steps, in which massive storage and intensive I/O might not be the optimal choice. The structurally based nonstationary and anisotropic Bessel smoothing is considered for an efficient gradient preconditioning. The filter is implemented as a PDE solved with SEM, directly on the modeling mesh, following the same high-performance-computing structure as the one for the wave equation.

The complexity of the data and elastic multiparameter FWI problem in complex land areas are illustrated through a 3D subset of SEAM Phase II Foothill model. The two-step data-windowing hierarchy makes possible for simultaneous estimation of V_P and V_S. Surface waves can be treated as additional sources of information, to provide better constraints on V_P and V_S estimation, at near-surface and deeper parts. In this model, the smoothing process costs only 0.4% the computational cost of one inversion iteration, but it significantly enhances the structure and model parameter estimation. Perspectives include the applications on a real foothill data set, in which rough topography and near-surface heterogeneities make the subsurface imaging very challenging.

ACKNOWLEDGMENTS

DATA AND MATERIALS AVAILABILITY

Data associated with this research belongs to the SEISCOPE consortium. Any request related to the materials availability should be addressed to Romain Brossier (romain.brossier@univ-grenoble-alpes.fr).

APPENDIX A

STRESS AND STRAIN RELATIONSHIP IN THE TIME DOMAIN

According to equation 1, in anelastic medium, the stress σ_{ij} is the convolution of the relaxation rate M_{ijkl} with the strain ε_{kl} as

$$\sigma_{ij}(\mathbf{x}, t) = M_{ijkl}(\mathbf{x}, t) *_t \varepsilon_{kl}(\mathbf{x}, t) + \mathcal{T}_{ij}(\mathbf{x}, t). \qquad \text{(A-1)}$$

In the frequency domain, this relationship becomes

$$\hat{\sigma}_{ij}(\mathbf{x}, \omega) = \hat{M}_{ijkl}(\mathbf{x}, \omega)\hat{\varepsilon}_{kl}(\mathbf{x}, \omega) + \hat{\mathcal{T}}_{ij}(\mathbf{x}, \omega). \qquad \text{(A-2)}$$

Under the assumption that L SLSs at specific reference frequencies ω_ν are used to fit a constant Q parameter over the frequency band $[\omega_{\min}, \omega_{\max}]$, the associated expression for the complex relaxation rate $\hat{M}_{ijkl}(\mathbf{x}, \omega)$ is represented in equation 7. Equation A-2 now becomes

$$\hat{\sigma}_{ij}(\mathbf{x}, \omega) = \left[c_{ijkl}(\mathbf{x}) - c_{ijkl}(\mathbf{x})Q_{ijkl}^{-1}(\mathbf{x}) \sum_{\nu=1}^{L} y_\nu \frac{\omega_\nu}{\omega_\nu + i\omega} \right] \hat{\varepsilon}_{kl}(\mathbf{x}, \omega)$$
$$+ \hat{\mathcal{T}}_{ij}(\mathbf{x}, \omega). \qquad \text{(A-3)}$$

By using the definition of the attenuative stiffness coefficients c_{ijkl}^a given in equation 9, we can rewrite equation A-3 as

$$\hat{\sigma}_{ij}(\mathbf{x}, \omega) = c_{ijkl}(\mathbf{x})\hat{\varepsilon}_{kl}(\mathbf{x}, \omega)$$
$$- c_{ijkl}^a(\mathbf{x}) \sum_{\nu=1}^{L} y_\nu \frac{\omega_\nu}{\omega_\nu + i\omega} \hat{\varepsilon}_{kl}(\mathbf{x}, \omega) + \hat{\mathcal{T}}_{ij}(\mathbf{x}, \omega). \qquad \text{(A-4)}$$

Introducing the complex memory variable $\hat{\psi}_{\nu;kl}(\mathbf{x}, \omega)$ such that

$$\hat{\psi}_{\nu;kl}(\mathbf{x}, \omega) = y_\nu \frac{\omega_\nu}{\omega_\nu + i\omega} \hat{\varepsilon}_{kl}(\mathbf{x}, \omega), \qquad \text{(A-5)}$$

leads to the simplification of equation A-4 into

$$\hat{\sigma}_{ij}(\mathbf{x}, \omega) = c_{ijkl}(\mathbf{x})\hat{\varepsilon}_{kl}(\mathbf{x}, \omega) - c_{ijkl}^a(\mathbf{x}) \sum_{\nu=1}^{L} \hat{\psi}_{\nu;kl}(\mathbf{x}, \omega)$$
$$+ \hat{\mathcal{T}}_{ij}(\mathbf{x}, \omega). \qquad \text{(A-6)}$$

In the time domain, equation A-6 is equivalent to the relationship between stress and strain mentioned in equation 10.

Equation A-5 can be developed as

$$(i\omega + \omega_\nu)\hat{\psi}_{\nu;kl}(\mathbf{x}, \omega) = \omega_\nu y_\nu \hat{\varepsilon}_{kl}(\mathbf{x}, \omega). \qquad \text{(A-7)}$$

By applying the inverse Fourier transform into this equation, we obtain

$$\partial_t \psi_{\nu;kl}(\mathbf{x}, t) + \omega_\nu \psi_{\nu;kl}(\mathbf{x}, t) = \omega_\nu y_\nu \varepsilon_{kl}(\mathbf{x}, t), \qquad \text{(A-8)}$$

which is the ODE governing the propagation of the memory variables, mentioned in equation 11.

APPENDIX B

ADJOINT FIELDS DEFINITION AND GRADIENTS ESTIMATION

The misfit function in equation 15 between the synthetic and the observed data $\mathbf{d}_{\text{obs}}$ can be written as

$$\chi(\mathbf{m}) = \frac{1}{2} \|R\mathcal{W}(\mathbf{m}) - \mathbf{d}_{\text{obs}}\|^2, \qquad \text{(B-1)}$$

where R is a restriction operator extracting the full wavefield $\mathcal{W}$ at receiver positions for each source. We define the incident wavefield from the displacement and memory variable vectors as

$$\mathcal{W} = (\underbrace{u_1, u_2, u_3}_{\mathbf{u}}, \underbrace{\psi_{1;1}, \ldots, \psi_{1;6}}_{\boldsymbol{\psi}_1}, \ldots, \underbrace{\psi_{L;1}, \ldots, \psi_{L;6}}_{\boldsymbol{\psi}_L})^T.$$
$$\text{(B-2)}$$

The forward problem (equation 12) can now be written as

$$\overbrace{\begin{pmatrix} \rho I_3 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 0 \end{pmatrix}}^{\mathbf{B}_2(\mathbf{m})} \partial_{tt}\mathcal{W} + \overbrace{\begin{pmatrix} 0 & 0 & \ldots & 0 \\ 0 & I_6 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & I_6 \end{pmatrix}}^{\mathbf{B}_1} \partial_t\mathcal{W}$$
$$+ \underbrace{\begin{pmatrix} -DCD^T & DC^a & \ldots & DC^a \\ -y_1\omega_1 D^T & \omega_1 I_6 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ -y_L\omega_L D^T & 0 & \ldots & \omega_L I_6 \end{pmatrix}}_{\mathbf{B}_0(\mathbf{m})} \mathcal{W} = \mathbf{S}, \qquad \text{(B-3)}$$

or

$$F(\mathbf{m}, \mathcal{W}) = \mathbf{B}_2(\mathbf{m})\partial_{tt}\mathcal{W} + \mathbf{B}_1\partial_t\mathcal{W} + \mathbf{B}_0(\mathbf{m})\mathcal{W} - \mathbf{S} = 0, \qquad \text{(B-4)}$$

where the model $\mathbf{m}$ represents some physical parameters of the subsurface.

Adjoint system

Based on the second-order forward problem and the misfit function in equation 15, we introduce the associated Lagrangian functional with the Lagrangian multiplier $\bar{\mathcal{W}}$

$$L(\mathbf{m}, \mathcal{W}, \bar{\mathcal{W}}) = \frac{1}{2} \|R\mathcal{W} - \mathbf{d}_{\text{obs}}\|^2 + (\bar{\mathcal{W}}, F(\mathbf{m}, \mathcal{W})). \qquad \text{(B-5)}$$

Similar to the incident full wavefield $\mathcal{W}$, the adjoint multiplier consists of adjoint displacement and memory variable fields

$$\bar{\mathcal{W}} = (\bar{\mathbf{u}}, \bar{\boldsymbol{\phi}}_1, \ldots, \bar{\boldsymbol{\phi}}_L)^T. \qquad \text{(B-6)}$$

Following the standard development of the adjoint-state method (Plessix, 2006), the Lagrangian multiplier should satisfy the following equation:

$$\mathbf{B}_2^{\dagger}\partial_{tt}\bar{\mathcal{W}} - \mathbf{B}_1^{\dagger}\partial_t\bar{\mathcal{W}} + \mathbf{B}_0(\mathbf{m})^{\dagger}\bar{\mathcal{W}} = -R^{\dagger}(R\mathcal{W} - \mathbf{d}_{\text{obs}}),$$
(B-7)

resulting in the second-order adjoint system:

$$\begin{cases} \rho\partial_{tt}\bar{\mathbf{u}} = DCD^T\bar{\mathbf{u}} - D\sum_{\nu=1}^{L} y_\nu\omega_\nu\bar{\boldsymbol{\phi}}_\nu - R^{\dagger}\Delta d_{\mathbf{u}}, \\ \partial_t\bar{\boldsymbol{\phi}}_\nu - \omega_\nu\bar{\boldsymbol{\phi}}_\nu = -C^a D^T\bar{\mathbf{u}}, \end{cases}$$
(B-8)

where $\Delta d_{\mathbf{u}}$ is the displacement misfit at receiver positions, acting as a source term. We introduce the modified adjoint memory variables $\bar{\boldsymbol{\psi}}_\nu$ such that

$$y_\nu\omega_\nu\bar{\boldsymbol{\phi}}_\nu = C^a\bar{\boldsymbol{\psi}}_\nu,$$
(B-9)

which transforms the adjoint system into

$$\begin{cases} \rho\partial_{tt}\bar{\mathbf{u}} = DCD^T\bar{\mathbf{u}} - DC^a\sum_{\nu=1}^{L}\bar{\boldsymbol{\psi}}_\nu - R^{\dagger}\Delta d_{\mathbf{u}}, \\ \partial_t\bar{\boldsymbol{\psi}}_\nu - \omega_\nu\bar{\boldsymbol{\psi}}_\nu = -y_\nu\omega_\nu D^T\bar{\mathbf{u}}. \end{cases}$$
(B-10)

The adjoint system now has nearly identical equations as the incident problem in equation 12, except the sign "−" in the memory variable ODE.

Gradient expressions

The gradient of the misfit function on any model parameter can be computed through

$$\frac{\partial\chi}{\partial\mathbf{m}} = \left(\bar{\mathcal{W}}, \frac{\partial F(\mathbf{m}, \mathcal{W})}{\partial\mathbf{m}}\right).$$
(B-11)

The gradient on density ρ is given by

$$\frac{\partial\chi}{\partial\rho} = \left(\bar{\mathcal{W}}, \frac{\partial\mathbf{B}_2}{\partial\rho}\right) = (\bar{\mathbf{u}}, \partial_{tt}\mathbf{u}).$$
(B-12)

The gradient on stiffness coefficient C_{IJ} is given by

$$\frac{\partial\chi}{\partial C_{IJ}} = \left(\bar{\mathcal{W}}, \frac{\partial\mathbf{B}_0}{\partial C_{IJ}}\right)$$
$$= \left(\bar{\boldsymbol{\varepsilon}}, \frac{\partial C}{\partial C_{IJ}}\boldsymbol{\varepsilon}\right)_{\Omega,t} - \left(\bar{\boldsymbol{\varepsilon}}, \sum_{\nu=1}^{L}\frac{\partial C^a}{\partial C_{IJ}}\boldsymbol{\psi}_\nu\right)_{\Omega,t}.$$
(B-13)

The gradient on attenuation parameters $Q_{p,s}^{-1}$ is given by

$$\frac{\partial\chi}{\partial Q_{p,s}^{-1}} = \left(\bar{\mathcal{W}}, \frac{\partial\mathbf{B}_0}{\partial Q_{p,s}^{-1}}\right) = -\left(\bar{\boldsymbol{\varepsilon}}, \sum_{\nu=1}^{L}\frac{\partial C^a}{\partial Q_{p,s}^{-1}}\boldsymbol{\psi}_\nu\right).$$
(B-14)

The C_{IJ} gradient in equation B-13 involves the multiplication of the incident strain field with the quantity $\partial C/\partial C_{IJ}$, which is a matrix with zero or one coefficients. An example is given in the equation below:

$$\frac{\partial C}{\partial C_{13}} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$
(B-15)

The C_{IJ} and $Q_{p,s}^{-1}$ gradient expressions (Equation B-14) also require the matrices $\partial C^a/\partial C_{IJ}$ and $\partial C^a/\partial Q_{p,s}^{-1}$. Under the isotropic attenuation regime, our approximation in equations 13 and 14 provides simple expressions of the derivatives of the attenuative stiffness tensor C^a with respect to elastic stiffness coefficients and attenuation parameters as

$$\left.\frac{\partial C^a}{\partial C_{IJ}}\right|_{I\neq J} = 0,$$
(B-16)

$$\left.\frac{\partial C^a}{\partial C_{II}}\right|_{I=1,2,3} = \frac{1}{3}Q_p^{-1}\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$
(B-17)

$$\left.\frac{\partial C^a}{\partial C_{II}}\right|_{I=4,5,6} = \frac{1}{3}Q_s^{-1}\begin{bmatrix} 0 & -2 & -2 & 0 & 0 & 0 \\ -2 & 0 & -2 & 0 & 0 & 0 \\ -2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$
(B-18)

$$\frac{\partial C^a}{\partial Q_p^{-1}} = \frac{1}{3}\left(\sum_{I=1}^{3}C_{II}\right)\begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$
(B-19)

$$\frac{\partial C^a}{\partial Q_s^{-1}} = \frac{1}{3}\left(\sum_{I=4}^{6}C_{II}\right)\begin{bmatrix} 0 & -2 & -2 & 0 & 0 & 0 \\ -2 & 0 & -2 & 0 & 0 & 0 \\ -2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$
(B-20)

These expressions confirm that the attenuation coefficient Q_p has an isotropic impact on compressional components, thus having an isotropic impact on V_P estimation (a similar interpretation for the attenuation coefficient Q_s). Equations B-17 and B-18 also highlight that, in a weakly attenuative medium, the second term in the viscoelastic C_{IJ} gradient (Equation 20) can be neglected and tends to zero in elastic medium.

Table C-1. Medium properties of the LOH benchmark.

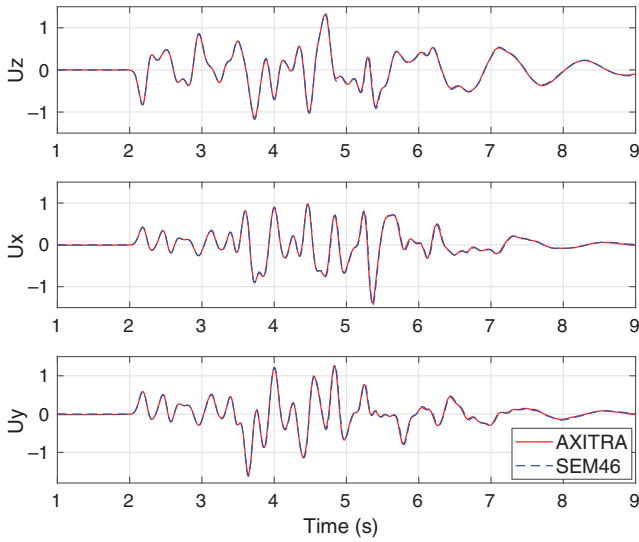|	V_P(m/s)	V_S(m/s)	ρ(m/s)	Q_p	Q_s
Top layer	4000	2000	2600	120	40
Half-space	6000	3464	2700	155.9	69.3

Figure C-1. Excellent agreement between the elastic SEM46 simulation result and the semianalytic solution obtained from AXITRA on LOH1 configuration. Blue, SEM46 solution; red, AXITRA solution.

APPENDIX C

SEMIANALYTICAL SOLUTIONS COMPARISON — LOH BENCHMARKS

The LOH benchmark has a layered material model, in which the top 1000 m have different properties than the rest of the domain (see Table C-1). The problem is driven by a single-point moment source $\mathcal{M}_{xy} = 10^{18}$ N.m, positioned in the lower half-space at depth $z_s = 2$ km. The receiver is located at the free surface with the distance approximately 10 km from the source ($\Delta x_{sr} = 6$ km, $\Delta y_{sr} = 8$ km).

We are going to compare the SEM46 solutions with the AXITRA open-source package solutions, which offers a highly accurate semianalytical solution on a layered medium via the boundary integral method (Coutant, 1989). For viscoelastic solution, only the Kjartansson's constant Q model is originally available in AXITRA (Kjartansson, 1979). We thus introduce the approximation by three predefined SLS mechanisms over the modeling frequency range of the benchmark.

LOH1 comparison

LOH1 solutions come from an elastic simulation with a Gaussian source time function centered at 2.65 Hz. The SEM mesh consists of a set of regular cubic elements, with the element size calculated from the volume condition: $dx = dy = 200$ m. The element size in the vertical direction is a factor of the thickness of the top layer $dz = 100$ m to better represent the sharp interface.

In the LOH configuration, the wavefield is complex with dispersive surface waves and multiple body-wave reflections between the free surface and the sharp contrast. However, Figure C-1 shows an excellent agreement between the elastic SEM46 simulation result and the semianalytic solution obtained from AXITRA. Because SEM is a continuous FE method, it has some intrinsic limitation in the representation of the sharp contrast. We thus do not push the comparison further with any time-frequency analysis.

a) Comparison with SLSs mechanism
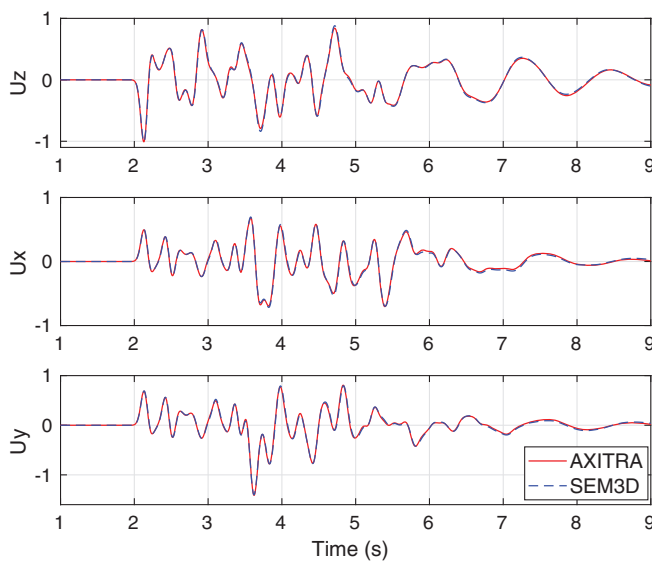
b) Comparison with Kjartasson mechanism
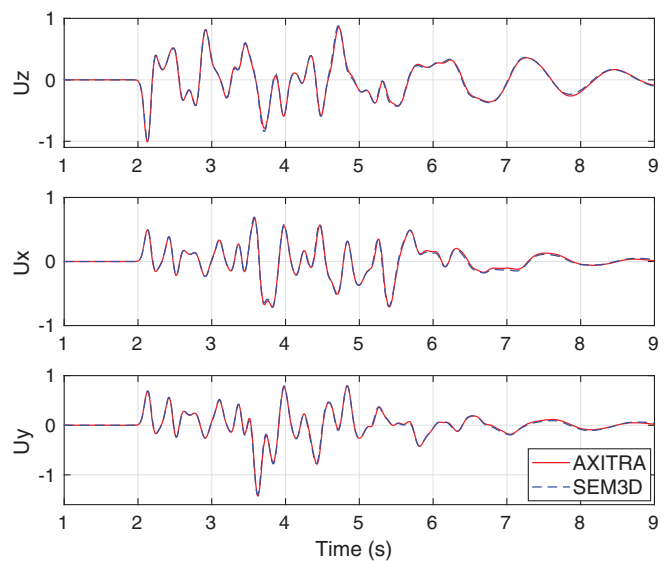
Figure C-2. Excellent agreement between the viscoelastic SEM46 simulation result and the semianalytic solution obtained from AXITRA on LOH3 configuration: (a) using three SLSs mechanism and (b) Kjartansson model with reference frequency $f_r = 100$ Hz. Blue, SEM46 solution; red, AXITRA solution.

LOH3 comparison

LOH3 solutions come from a viscoelastic simulation with a Gaussian source time function at a 3.18 Hz center frequency. The medium is meshed with regular cubic elements $dx = dy = 160$ m and $dz = 100$ m to better represent the sharp contrast at $z = 1$ km. SEM46 uses three SLSs to approximate the constant quality factors. By using the same SLS mechanisms (similar dimensionless anelastic coefficients y_ν and reference frequencies ω_ν), we obtain an excellent agreement between the SEM46 and modified-AXITRA solutions in Figure C-2a. Figure C-2b also shows very good fit between SEM46 and AXITRA solutions with Kjartansson model, at reference frequency $f_r = 100$ Hz. It has been highlighted by Wang (2009) that the SLS model has a finite phase velocity at infinite frequency ($V_\infty < \infty$), whereas Kjartansson's constant-Q model assumes $V_\infty = \infty$. The use of a "tuning" frequency is thus necessary to match solutions obtained by different attenuation models. It should be remarked that a different attenuation model provides a different expression of the medium parameters, leading to difficulties for assessing the validation of the attenuation propagation.

REFERENCES

Abramowitz, M., and I. Stegun, 1972, Handbook of mathematical functions with formulas, graphs, and mathematical tables: Dover publications Inc.

Amestoy, P., R. Brossier, A. Buttari, J.-Y. L'Excellent, T. Mary, L. Métivier, A. Miniussi, and S. Operto, 2016, Fast 3D frequency-domain full waveform inversion with a parallel block low-rank multifrontal direct solver: Application to OBC data from the North Sea: Geophysics, **81**, no. 6, R363–R383, doi: 10.1190/geo2016-0052.1.

Anderson, J. E., L. Tan, and D. Wang, 2012, Time-reversal checkpointing methods for RTM and FWI: Geophysics, **77**, no. 4, S93–S103, doi: 10.1190/geo2011-0114.1.

Blanch, J., J. O. A. Robertson, and W. W. Symes, 1995, Modeling of a constant Q: Methodology and algorithm for an efficient and optimally inexpensive viscoelastic technique: Geophysics, **60**, 176–184, doi: 10.1190/1.1443744.

Boehm, C., M. Hanzich, J. de la Puente, and A. Fichtner, 2016, Wavefield compression in full-waveform inversion for adjoint methods: Geophysics, **81**, no. 6, R385–R397, doi: 10.1190/geo2015-0653.1.

Bohlen, T., 2002, Parallel 3-D viscoelastic finite-difference seismic modeling: Computers and Geosciences, **28**, 887–899, doi: 10.1016/S0098-3004(02)00006-7.

Bohlen, T., and E. H. Saenger, 2006, Accuracy of heterogeneous staggered-grid finite-difference modeling of Rayleigh waves: Geophysics, **71**, no. 4, T109–T115, doi: 10.1190/1.2213051.

Borisov, D., and S. C. Singh, 2015, Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: A synthetic study: Geophysical Journal International, **201**, 1215–1234, doi: 10.1093/gji/ggv048.

Brenders, A. J., and R. G. Pratt, 2007, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151, doi: 10.1111/j.1365-246X.2006.03156.x.

Brossier, R., S. Operto, and J. Virieux, 2009, 2D elastic frequency-domain full-waveform inversion for imaging complex onshore structures: 71st Annual International Conference and Exhibition, EAGE, Extended Abstracts, U019, doi: 10.3997/2214-4609.201400376.

Capdeville, Y., L. Guillot, and J.-J. Marigo, 2010, 2-D non-periodic homogenization to upscale elastic media for P-SV waves: Geophysical Journal International, **182**, 903–922, doi: 10.1111/j.1365-246X.2010.04636.x.

Carcione, J., D. Kosloff, and R. Kosloff, 1988, Wave propagation simulation in a linear viscoacoustic medium: Geophysical Journal International, **93**, 393–401, doi: 10.1111/j.1365-246X.1988.tb02010.x.

Coutant, O., 1989, Program of numerical simulation AXITRA: Research report, LGIT, Grenoble.

Cupillard, P., E. Delavaud, G. Burgos, G. Festa, J. Vilotte, Y. Capdeville, and J. Montagner, 2011, RegSEM: A versatile code based on the spectral element method to compute seismic wave propagation at the regional scale: Geophysical Journal International, **188**, 1203–1220, doi: 10.1111/j.1365-246X.2011.05311.x.

de la Puente, J., M. Ferrer, M. Hanzich, J. E. Castillo, and J. M. Cela, 2014, Mimetic seismic wave modeling including topography on deformed staggered grids: Geophysics, **79**, no. 3, T125–T141, doi: 10.1190/geo2013-0371.1.

Deville, M., P. Fischer, and E. Mund, 2002, High order methods for incompressible fluid flow: Cambridge University Press.

Dussaud, E., W. W. Symes, P. Williamson, L. Lemaistre, P. Singer, B. Denel, and A. Cherrett, 2008, Computational strategies for reverse-time migration: 78th Annual International Meeting, SEG, Expanded Abstracts, 2267–2271, doi: 10.1190/1.3059336.

Emmerich, H., and M. Korn, 1987, Incorporation of attenuation into time-domain computation of seismic wavefield: Geophysics, **52**, 1252–1264, doi: 10.1190/1.1442386.

Faccioli, E. F., R. Paolucci, and A. Quarteroni, 1997, 2D and 3D elastic wave propagation by a pseudo-spectral domain decomposition method: Journal of Seismology, **1**, 237–251, doi: 10.1023/A:1009758820546.

Fichtner, A., J. Trampert, P. Cupillard, E. Saygin, T. Taymaz, Y. Capdeville, and A. V. Nor, 2013, Multiscale full waveform inversion: Geophysical Journal International, **194**, 534–556, doi: 10.1093/gji/ggt118.

Fichtner, A., and M. van Driel, 2014, Models and Fréchet kernels for frequency-(in)dependent Q: Geophysical Journal International, **198**, 1878–1889, doi: 10.1093/gji/ggu228.

Gao, L., R. Brossier, B. Pajot, J. Tago, and J. Virieux, 2015, An immersed free surface boundary treatment for seismic wave simulation: Geophysics, **80**, no. 5, T193–T209, doi: 10.1190/geo2014-0609.1.

Gokhberg, A., and A. Fichtner, 2016, Full-waveform inversion on heterogeneous HPC systems: Computers and Geosciences, **89**, 260–268, doi: 10.1016/j.cageo.2015.12.013.

Gosselin-Cliche, B., and B. Giroux, 2014, 3D frequency-domain finite-difference viscoelastic-wave modeling using weighted average 27-point operators with optimal coefficients: Geophysics, **79**, no. 3, T169–T188, doi: 10.1190/geo2013-0368.1.

Griewank, A., and A. Walther, 2000, Algorithm 799: Revolve: An implementation of checkpointing for the reverse or adjoint mode of computational differentiation: ACM Transactions on Mathematical Software, **26**, 19–45, doi: 10.1145/347837.347846.

Guitton, A., G. Ayeni, and E. Díaz, 2012, Constrained full-waveform inversion by model reparameterization: Geophysics, **77**, no. 2, R117–R127, doi: 10.1190/geo2011-0196.1.

Huiskes, M., R. Plessix, and W. Mulder, 2017, Acoustic VTI full-waveform inversion with 3-D free-surface topography: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.201700503.

Kjartansson, E., 1979, Constant Q-wave propagation and attenuation: Journal of Geophysical Research, **84**, 4737–4748, doi: 10.1029/JB084iB09p04737.

Komatitsch, D., 1997, Méthodes spectrales et éléments spectraux pour l'équation de l'élastodynamique 2D et 3D en milieu hétérogène: Ph.D. thesis, Institut de Géophysique du Globe de Paris.

Komatitsch, D., G. Erlebacher, D. Göddeke, and D. Michéa, 2010, High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster: Journal of Computational Physics, **229**, 7692–7714, doi: 10.1016/j.jcp.2010.06.024.

Komatitsch, D., and J. Tromp, 1999, Introduction to the spectral element method for three-dimensional seismic wave propagation: Geophysical Journal International, **139**, 806–822, doi: 10.1046/j.1365-246x.1999.00967.x.

Komatitsch, D., and J. P. Vilotte, 1998, The spectral element method: An efficient tool to simulate the seismic response of 2D and 3D geological structures: Bulletin of the Seismological Society of America, **88**, 368–392.

Komatitsch, D., Z. Xie, E. Bozdag, E. S. de Andrade, D. Peter, Q. Liu, and J. Tromp, 2016, Anelastic sensitivity kernels with parsimonious storage for adjoint tomography and full waveform inversion: Geophysical Journal International, **206**, 1467–1478, doi: 10.1093/gji/ggw224.

Kostin, V., D. Neklyudov, V. Tcheverda, M. Belonosov, and M. Dmitriev, 2016, 3D elastic frequency-domain iterative solver for full-waveform inversion: 86th Annual International Meeting, SEG, Expanded Abstracts, 3825–3829, doi: 10.1190/segam2016-13611428.1.

Kristeková, M., J. Kristek, P. Moczo, and S. Day, 2006, Misfit criteria for quantitative comparison of seismograms: Bulletin of the Seismological Society of America, **96**, 1836–1850, doi: 10.1785/0120060012.

Lailly, P., 1983, The seismic problem as a sequence of before-stack migrations: Presented at the Conference on Inverse Scattering: Theory and Applications, SIAM.

LeVeque, R. J., 2007, Finite difference methods for ordinary and partial differential equations, steady state and time dependent problems: SIAM.

Li, Y., L. Métivier, R. Brossier, B. Han, and J. Virieux, 2014, 2D and 3D frequency-domain elastic wave modeling in complex media with a parallel iterative solver: Geophysics, **80**, no. 3, T101–T118, doi: 10.1190/geo2014-0480.1.

Liu, Q., and J. Tromp, 2008, Finite-frequency sensitivity kernels for global seismic wave propagation based upon adjoint methods: Geophysical Journal International, **174**, 265–286, doi: 10.1111/j.1365-246X.2008.03798.x.

Lombard, B., J. Piraux, C. Gelis, and J. Virieux, 2008, Free and smooth boundaries in 2-D finite-difference schemes for transient elastic waves:

Geophysical Journal International, **172**, 252–261, doi: 10.1111/j.1365-246X.2007.03620.x.

Maggi, A., C. Tape, M. Chen, D. Chao, and J. Tromp, 2009, An automated time-window selection algorithm for seismic tomography: Geophysical Journal International, **178**, 257–281, doi: 10.1111/j.1365-246X.2009.04099.x.

Marfurt, K., 1984, Accuracy of finite-difference and finite-element modeling of the scalar and elastic wave equations: Geophysics, **49**, 533–549, doi: 10.1190/1.1441689.

Métivier, L., and R. Brossier, 2016, The seiscope optimization toolbox: A large-scale nonlinear optimization library based on reverse communication: Geophysics, **81**, no. 2, F11–F25, doi: 10.1190/geo2015-0031.1.

Moczo, P., 1989, Finite-difference technique for SH waves in 2-D media using irregular grids, application to the seismic response problem: Geophysical Journal International, **99**, 321–329, doi: 10.1111/j.1365-246X.1989.tb01691.x.

Moczo, P., J. Ampuero, J. Kristek, S. Day, M. Kristekova, P. Pazak, M. Galis, and H. Igel, 2006, Comparison of Numerical Methods for Seismic Wave Propagation and Source Dynamics — The SPICE Code Validation: Presented at the International Symposium on the Effects of Surface Geology on Seismic Motion.

Moczo, P., E. Bystricky, J. M. Carcione, and M. Bouchon, 1997, Hybrid modeling of P-SV seismic motion at inhomogeneous viscoelastic topographic structures: Bulletin of the Seismological Society of America, **87**, 1305–1323.

Moczo, P., and J. Kristek, 2005, On the rheological models used for time-domain methods of seismic wave propagation: Geophysical Research Letters, **32**, doi: 10.1029/2004GL021598.

Nocedal, J., and S. J. Wright, 2006, Numerical optimization, 2nd ed.: Springer.

Operto, S., and A. Miniussi, 2017, Vertical wavespeed, density and attenuation imaging by 3D efficient frequency-domain FWI of wide-azimuth OBC data — A North Sea case study: 79th Annual International Conference and Exhibition, EAGE, Extended Abstracts, doi: 10.3997/2214-4609.201701159.

Operto, S., A. Miniussi, R. Brossier, L. Combe, L. Métivier, V. Monteiller, A. Ribodetti, and J. Virieux, 2015, Efficient 3-D frequency-domain mono-parameter full-waveform inversion of ocean-bottom cable data: Application to Valhall in the visco-acoustic vertical transverse isotropic approximation: Geophysical Journal International, **202**, 1362–1391, doi: 10.1093/gji/ggv226.

Operto, S., J. Virieux, J. X. Dessa, and G. Pascal, 2006, Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: Application to the eastern Nankai trough: Journal of Geophysical Research, **111**, doi: 10.1029/2005JB003835.

Oristaglio, M., 2012, SEAM phase II — Land seismic challenges: The Leading Edge, **31**, 264–266, doi: 10.1190/1.3694893.

Oristaglio, M., 2016, Land seismic challenges wraps up with foothills model: The Leading Edge, **35**, 292–293, doi: 10.1190/tle35030292.1.

Peter, D., D. Komatitsch, Y. Luo, R. Martin, N. Le Goff, E. Casarotti, P. Le Loher, F. Magnoni, Q. Liu, C. Blitz, T. Nissen-Meyer, P. Basini, and J. Tromp, 2011, Forward and adjoint simulations of seismic wave propagation on fully unstructured hexahedral meshes: Geophysical Journal International, **186**, 721–739, doi: 10.1111/j.1365-246X.2011.05044.x.

Petersson, N. A., and B. Sjögreen, 2012, Stable and efficient modeling of anelastic attenuation in seismic wave propagation: Communications in Computational Physics, **12**, 193–225, doi: 10.4208/cicp.201010.090611a.

Petersson, N. A., and B. Sjögreen, 2013, User's guide to SW4, version 1.1. Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550, USA, llnl-sm-662014 edition: Computational Infrastructure for geodynamics (CIG), www.geodynamics.org.

Petersson, N. A., and B. Sjögreen, 2015, Wave propagation in anisotropic elastic materials and curvilinear coordinates using a summation-by-parts finite difference method: Journal of Computational Physics, **299**, 820–841, doi: 10.1016/j.jcp.2015.07.023.

Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503, doi: 10.1111/j.1365-246X.2006.02978.x.

Plessix, R. E., and C. Perkins, 2010, Full waveform inversion of a deep water ocean bottom seismometer dataset: First Break, **28**, 71–78, doi: 10.3997/1365-2397.2010013.

Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. dell'Aversana, 2004, Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt: Geophysical Journal International, **159**, 1032–1056, doi: 10.1111/j.1365-246X.2004.02442.x.

Regone, C., J. Stefani, P. Wang, C. Gerea, G. Gonzalez, and M. Oristaglio, 2017, Geologic model building in SEAM Phase II — Land seismic challenges: The Leading Edge, **36**, 738–749, doi: 10.1190/tle36090738.1.

Robertsson, J. O. A., 1996, A numerical free-surface condition for elastic/viscoelastic finite-difference modeling in the presence of topography: Geophysics, **61**, 1921–1934, doi: 10.1190/1.1444107.

Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931, doi: 10.1111/j.1365-246X.2008.03768.x.

Shragge, J., 2016, Acoustic wave propagation in tilted transversely isotropic media: Incorporating topography: Geophysics, **81**, no. 5, C265–C278, doi: 10.1190/geo2015-0311.1.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, 65–70, doi: 10.3997/1365-2397.2010012.

Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248, doi: 10.1190/1.1649391.

Sjögreen, B., and N. A. Petersson, 2012, A fourth order accurate finite difference scheme for the elastic wave equation in second order formulation: Journal of Scientific Computing, **52**, 17–48, doi: 10.1007/s10915-011-9531-1.

Symes, W. W., 2007, Reverse time migration with optimal checkpointing: Geophysics, **72**, no. 5, SM213–SM221, doi: 10.1190/1.2742686.

Tape, C., Q. Liu, A. Maggi, and J. Tromp, 2010, Seismic tomography of the southern California crust based on spectral-element and adjoint methods: Geophysical Journal International, **180**, 433–462, doi: 10.1111/j.1365-246X.2009.04429.x.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259–1266, doi: 10.1190/1.1441754.

Tarantola, A., 1988, Theoretical background for the inversion of seismic waveforms including elasticity and attenuation: Pure and Applied Geophysics, **128**, 365–399, doi: 10.1007/BF01772605.

Trinh, P. T., R. Brossier, L. Métivier, L. Tavard, and J. Virieux, 2017a, Efficient 3D elastic FWI using a spectral-element method: 87th Annual International Meeting, SEG, Expanded Abstracts, 1533–1538, doi: 10.1190/segam2017-17587770.1.

Trinh, P. T., R. Brossier, L. Métivier, J. Virieux, and P. Wellington, 2017b, Bessel smoothing filter for spectral element mesh: Geophysical Journal International, **209**, 1489–1512, doi: 10.1093/gji/ggx103.

Tromp, J., C. Tape, and Q. Liu, 2005, Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels: Geophysical Journal International, **160**, 195–216, doi: 10.1111/j.1365-246X.2004.02453.x.

van Driel, M., and T. Nissen-Meyer, 2014, Optimized viscoelastic wave propagation for weakly dissipative media: Geophysical Journal International, **199**, 1078–1093, doi: 10.1093/gji/ggu314.

Vigh, D., K. Jiao, D. Watts, and D. Sun, 2014, Elastic full-waveform inversion application using multicomponent measurements of seismic data collection: Geophysics, **79**, no. 2, R63–R77, doi: 10.1190/geo2013-0055.1.

Vigh, D., N. Moldoveanu, K. Jiao, W. Huang, and J. Kapoor, 2013, Ultra-long-offset data acquisition can complement full-waveform inversion and lead to improved subsalt imaging: The Leading Edge, **32**, 1116–1122, doi: 10.1190/tle32091116.1.

Virieux, J., A. Asnaashari, R. Brossier, L. Métivier, A. Ribodetti, and W. Zhou, 2017, An introduction to full waveform inversion, *in* V. Grechka and K. Wapenaar, eds., Encyclopedia of exploration geophysics: SEG, R1-1–R1-40.

Virieux, J., and S. Operto, 2009, An overview of full waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.

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

Warner, M., A. Ratcliffe, T. Nangoo, J. Morgan, A. Umpleby, N. Shah, V. Vinje, I. Stekl, L. Guasch, C. Win, G. Conroy, and A. Bertrand, 2013, Anisotropic 3D full-waveform inversion: Geophysics, **78**, no. 2, R59–R80, doi: 10.1190/geo2012-0338.1.

Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016a, A review on the systematic formulation of 3D multiparameter full waveform inversion in viscoelastic medium: Geophysical Journal International, **207**, 129–149, doi: 10.1093/gji/ggw262.

Yang, P., R. Brossier, L. Métivier, and J. Virieux, 2016b, Wavefield reconstruction in attenuating media: A checkpointing-assisted reverse-forward simulation method: Geophysics, **81**, no. 6, R349–R362, doi: 10.1190/geo2016-0082.1.

Yang, P., R. Brossier, and J. Virieux, 2016c, Wavefield reconstruction from significantly decimated boundaries: Geophysics, **81**, no. 5, T197–T209, doi: 10.1190/geo2015-0711.1.

Zhu, H., E. Bozdağ, and J. Tromp, 2015, Seismic structure of the European upper mantle based on adjoint tomography: Geophysical Journal International, **201**, 18–52, doi: 10.1093/gji/ggu492.