GEOPHYSICS, VOL. 74, NO. 6 (NOVEMBER-DECEMBER 2009); P. WCC105–WCC118, 19 FIGS., 3 TABLES.
10.1190/1.3215771

# Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion

Romain Brossier[1], Stéphane Operto[2], and Jean Virieux[3]

## ABSTRACT

Quantitative imaging of the elastic properties of the subsurface at depth is essential for civil engineering applications and oil- and gas-reservoir characterization. A realistic synthetic example provides for an assessment of the potential and limits of 2D elastic full-waveform inversion (FWI) of wide-aperture seismic data for recovering high-resolution P- and S-wave velocity models of complex onshore structures. FWI of land data is challenging because of the increased nonlinearity introduced by free-surface effects such as the propagation of surface waves in the heterogeneous near-surface. Moreover, the short wavelengths of the shear wavefield require an accurate S-wave velocity starting model if low frequencies are unavailable in the data. We evaluated different multiscale strategies with the aim of mitigating the nonlinearities. Massively parallel full-waveform inversion was implemented in the frequency domain. The numerical optimization relies on a limited-memory quasi-Newton algorithm that outperforms the more classic preconditioned conjugate-gradient algorithm. The forward problem is based upon a discontinuous Galerkin (DG) method on triangular mesh, which allows accurate modeling of free-surface effects. Sequential inversions of increasing frequencies define the most natural level of hierarchy in multiscale imaging. In the case of land data involving surface waves, the regularization introduced by hierarchical frequency inversions is not enough for adequate convergence of the inversion. A second level of hierarchy implemented with complex-valued frequencies is necessary and provides convergence of the inversion toward acceptable P- and S-wave velocity models. Among the possible strategies for sampling frequencies in the inversion, successive inversions of slightly overlapping frequency groups is the most reliable when compared to the more standard sequential inversion of single frequencies. This suggests that simultaneous inversion of multiple frequencies is critical when considering complex wave phenomena.

## INTRODUCTION

Quantitative imaging of the elastic properties of the subsurface is essential for oil- and gas-reservoir characterization and for monitoring carbon dioxide sequestration with time-lapse acquisitions. Indeed, fluids and gases have significant effects on the elastic properties of the subsurface in terms of Poisson's ratio anomalies. This quantitative imaging is also required for near-surface imaging in the framework of civil engineering applications because the shear properties of the shallow weathered layers strongly impact the elastic wavefield. Moreover, at the near-surface scale, short propagation times do not allow for easy separation in time of body waves and surface waves. In this case, filtering or muting surface waves is not easy; both types of waves need to be involved in the imaging, which requires solving the elastic wave equation. Evolution of acquisition systems toward wide-aperture/wide-azimuth geometries and multicomponent recordings is another motivation behind the development of elastic imaging methods because the occurrence of P-S-mode conversion is dominant at wide-aperture angles.

Classically, reservoir characterization is performed by amplitude-variation-with-offset (AVO) analysis in the prestack domain (e.g., Jin et al., 2000; Buland and Omre, 2003). An alternative is full-waveform inversion (FWI) of the elastic wavefield recorded by multiple components for reconstructing the P- and S-wave velocity ($V_P$ and $V_S$) models of the subsurface (or related parameters such as impedances if density is involved in the inversion) with a resolution limit on the order of a half-wavelength. The misfit between the recorded and modeled wavefields is minimized through the resolution of a numerical optimization problem (Tarantola, 1987; Nocedal and

Manuscript received by the Editor 30 December 2008; revised manuscript received 7 April 2009; published online 3 December 2009; corrected version published online 17 December 2009.
[1]Université Nice-Sophia Antipolis, Géoazur, Observatoire de la Côte d'Azur, Valbonne, France. E-mail: brossier@geoazur.unice.fr.
[2]Université Nice-Sophia Antipolis, Géoazur, Observatoire de la Côte d'Azur, Villefranche-sur-mer, France. E-mail: operto@geoazur.obs-vlfr.fr.
[3]Université Joseph Fourier, Laboratoire de Géophysique Interne et Tectonophysique, Grenoble, France. E-mail: jean.virieux@obs.ujf-grenoble.fr.

Wright, 1999). The FWI forward problem is based on the complete solution of the full (two-way) wave equation.

One drawback is that elastic FWI is a computationally challenging problem. Recent advances in high-performance computing on large-scale distributed-memory platforms allow 2D problems of representative sizes to be tackled; whereas 3D problems start to be investigated only in the acoustic case (Sirgue et al., 2007; Ben-Hadj-Ali et al., 2008; Vigh and Starr, 2008; Warner et al., 2008). Moreover, the complexity of the elastic wavefield at wide aperture makes the inverse problem highly nonlinear and ill posed. This is even more dramatic for onshore applications where surface waves with high energy and body waves can be used in the optimization process. Full-waveform inversion is solved conventionally with local optimization (linearized) approaches, which makes the inversion sensitive to the limited accuracy of the starting model. As such, it is essential to use realistic synthetic case studies to investigate under which conditions the nonlinearity of the elastic FWI can be mitigated.

Most recent applications of FWI to real data have been performed under acoustic approximation (Pratt and Shipp, 1999; Hicks and Pratt, 2001; Ravaut et al., 2004; Gao et al., 2006; Operto et al., 2006; Bleibinhaus et al., 2007). Although reliable results can be obtained with acoustic approximation if judicious data preprocessing and inversion preconditioning are applied (Brenders and Pratt, 2007b), elastic FWI is desirable for applications that detect fluids and gases and sequester carbon dioxide ($CO_2$). Moreover, acoustic FWI can lead to erroneous models if applied to elastic data when the velocity models show high velocity contrasts and when specific FWI preprocessing and tuning are not applied to the data (Barnes and Charara, 2008).

Only a few applications of elastic FWI have been presented. In early applications, Crase et al. (1990, 1992) apply elastic FWI to real land and marine reflection data using limited offsets. In this framework, FWI is applied as quantitative migration processing for imaging P- and S-wave impedances. With the benefit provided by wide-aperture data to build the large and intermediate wavelengths of the subsurface recognized by Mora (1987, 1988), acoustic and elastic FWI have evolved as an attempt to build high-resolution velocity models.

Shipp and Singh (2002) perform 2D time-domain FWI of a small, wide-aperture marine data subset recorded by a long streamer (12 km long). Although the forward problem is solved using the elastic-wave equation, only $V_P$ is reconstructed during FWI, assuming an empirical relationship between $V_P$ and $V_S$ and between $V_P$ and density. They design a hierarchical multistep approach based on layer stripping and offset and time windowing, where the goal is to mitigate the nonlinearity of the inverse problem.

Sears et al. (2008) design a similar multistep strategy to perform elastic time-domain FWI of multicomponent ocean-bottom-cable (OBC) data. Their strategy is based on the selection of the data component, parameter class, and arrival type (by time windowing). It is useful especially when the P-to-S conversion is small at the sea bottom, which makes reconstructing the $V_S$ model particularly ill posed.

Choi and Shin (2008) and Choi et al. (2008) apply elastic frequency-domain FWI to onshore and offshore versions of the synthetic Marmousi2 model (Martin et al., 2006), respectively. They successfully image the model using a velocity-gradient starting model and a very low starting frequency (0.16 Hz).

Shi et al. (2007) apply elastic time-domain FWI to marine data collected from a gas field in western China. They successfully image a zone of Poisson's ratio anomalies associated with gas layers. Accurate starting $V_P$ and $V_S$ models are built from the P- and P-SV-wave velocity analysis and from a priori information of several well logs along the profile.

Gelis et al. (2007) implement a 2D elastic frequency-domain FWI using Born and Rytov approximations to linearize the inverse problem. They highlight the dramatic footprint of the surface waves on imaging small inclusions in homogeneous background models. To mitigate this footprint, they only involve body waves during the early stages of the inversion by selecting short-offset traces.

Our study presents a 2D massively parallel elastic frequency-domain FWI algorithm based on a discontinuous Galerkin (DG) forward problem (Brossier et al., 2008), applied to a realistic synthetic onshore case study. We want to assess whether surface waves and body waves recorded by wide-aperture acquisition geometries can be inverted jointly to build high-resolution $V_P$ and $V_S$ of complex onshore structures.

We implement FWI in the frequency domain, which presents some distinct advantages with respect to the time-domain formulation (Pratt and Worthington, 1990; Pratt, 1999; Sirgue and Pratt, 2004). The inverse problem can be solved with a local optimization approach using a conjugate gradient or a quasi-Newton method (Nocedal and Wright, 1999). The gradient of the objective function is computed with the adjoint-state technique (Plessix, 2006). Successive inversions of increasing frequency provide a natural framework for multiscale imaging algorithms. Moreover, by sacrificing the data redundancy of multifold acquisitions, inverting a limited number of frequencies can be enough to build reliable velocity models, provided the acquisition geometry spans over sufficiently long offsets. These few frequencies can be modeled efficiently in the 2D case for multiple shots once the impedance matrix that results from discretizing the frequency-domain wave equation is factorized through lower-and-upper-matrices (LU) decomposition (Marfurt, 1984; Nihei and Li, 2007). Finally, attenuation can be implemented in the forward problem straightforwardly and without extra computational cost by using complex velocities.

The main drawback of the frequency-domain FWI formulation arises from the difficulty of time windowing the modeled data when inverting one or a few sparsely sampled frequencies at a time. Time windowing allows selection of specific arrivals during the various stages of the inversion. A last resort is the use of complex-valued frequencies, which damp arrivals starting at a given traveltime (Shin et al., 2002).

In the next section, we review the theory of frequency-domain elastic full-wavefield modeling and inversion. Following, we review several multiscale strategies to mitigate the nonlinearity of the elastic inverse problem. These strategies involve two nested levels of hierarchy over frequencies and aperture angles in the inversion algorithm. The effectiveness of these strategies is first illustrated with a simple two-parameter elastic problem with a free surface. Then we apply the elastic frequency-domain FWI algorithm to a realistic synthetic example to reconstruct a dip section of the SEG/EAGE overthrust model, assuming a constant Poisson's ratio. The results of the analyses show that simultaneous inversions of multiple frequencies and data preconditioning by time damping are critical to obtaining reliable results when surface waves propagating in a heterogeneous near surface are present in the elastic wavefield. We also illustrate the improvements provided by quasi-Newton algorithms compared to more conventional conjugate-gradient approaches.

## METHOD AND ALGORITHM

### Forward problem

The 2D elastic frequency-domain FWI requires computing the frequency solution of the elastic P-SV equations in heterogeneous media. We present a short review of the $P_0$ DG method used in this study (see Brossier et al. [2008] for details).

We consider the first-order hyperbolic system, where particle velocities $(V_x,V_z)$ and stresses $(\sigma_{xx},\sigma_{zz},\sigma_{xz})$ are unknown quantities, as described by the system

$$
\begin{aligned}
-i\omega V_x &= \frac{1}{\rho(\mathbf{x})}\left\{\frac{\partial \sigma_{xx}}{\partial x}+\frac{\partial \sigma_{xz}}{\partial z}\right\}+f_x\\
-i\omega V_z &= \frac{1}{\rho(\mathbf{x})}\left\{\frac{\partial \sigma_{xz}}{\partial x}+\frac{\partial \sigma_{zz}}{\partial z}\right\}+f_z\\
-i\omega\sigma_{xx} &= (\lambda(\mathbf{x})+2\mu(\mathbf{x}))\frac{\partial V_x}{\partial x}+\lambda(\mathbf{x})\frac{\partial V_z}{\partial z}\\
&\quad -i\omega\sigma_{xx_0}\\
-i\omega\sigma_{zz} &= \lambda(\mathbf{x})\frac{\partial V_x}{\partial x}+(\lambda(\mathbf{x})+2\mu(\mathbf{x}))\frac{\partial V_z}{\partial z}\\
&\quad -i\omega\sigma_{zz_0}\\
-i\omega\sigma_{xz} &= \mu(\mathbf{x})\left\{\frac{\partial V_x}{\partial z}+\frac{\partial V_z}{\partial x}\right\}-i\omega\sigma_{xz_0},
\end{aligned}
\tag{1}
$$

where the Lamé coefficients that describe the medium are denoted by $\lambda$ and $\mu$, the density by $\rho$, and the angular frequency by $\omega$. Source terms are point forces $(f_x,f_z)$ or applied stresses $(\sigma_{xx_0},\sigma_{zz_0},\sigma_{xz_0})$. The purely imaginary term $i$ is defined by $i=\sqrt{-1}$. Only isotropic media are considered in this study.

Equation 1 is discretized and solved with the DG method (Käser and Dumbser, 2006). We use low-order $P_0$ interpolation, which corresponds to piecewise constant-velocity and stress fields, and physical parameters in each cell. The DG method is applied to the weak formulation of equation 1; the partial derivatives are computed through numerical fluxes exchanged at interfaces between cells. The perfectly matched layers (PML) method is used for absorbing boundary conditions along the edges of the model (Berenger, 1994). An explicit free-surface boundary condition for arbitrary complex topographies is implemented on top of the model by canceling fluxes of normal stresses along the boundary that consists of edges of triangles.

A sufficient level of accuracy for FWI can be obtained using the DG $P_0$ method, with 10–15 cells per minimum shear wavelength in regular equilateral meshes (Brossier et al., 2008). These regular meshes are used for the simulations presented later. Note that extending the DG method to higher-order interpolations is required for modeling in unstructured meshes and for combining different interpolations arbitrarily during one simulation (e.g., Dumbser and Käser, 2006).

The discretization of equation 1 leads to solving the linear system

$$\mathbf{A}\mathbf{v}=\mathbf{s}, \tag{2}$$

where the coefficients of the impedance matrix $\mathbf{A}$ — namely, the forward-modeling operator — depend on the modeled frequency and the medium properties. Vector $\mathbf{s}$ represents the source term; $\mathbf{v}$ represents the unknowns for particle velocities and stresses in each cell. Note that only the vertical and horizontal particle velocity wavefields are inverted in this study.

### Inverse problem

Full-waveform inversion is an optimization problem that can be recast as a linearized least-squares problem which attempts to minimize the misfit between the recorded and the modeled wavefields (Tarantola, 1987). The inverse problem can be formulated in the frequency domain (Pratt and Worthington, 1990). The associated objective function to be minimized is defined by

$$\mathcal{C}^{(k)}=\frac{1}{2}\Delta\mathbf{d}^{\dagger}\mathbf{W}_d\Delta\mathbf{d}=\frac{1}{2}\Delta\mathbf{d}^{\dagger}\mathbf{S}_d^{\dagger}\mathbf{S}_d\Delta\mathbf{d}, \tag{3}$$

where $\Delta\mathbf{d}=d_{\text{obs}}-d_{\text{calc}}^{(k)}$ is the data misfit vector, the difference between the observed data $d_{\text{obs}}$, and the modeled data $d_{\text{calc}}^{(k)}$ computed in the model $\mathbf{m}^{(k)}$. Superscript † indicates the adjoint (transposed conjugate), and $\mathbf{S}_d$ is a diagonal weighting matrix applied to the misfit vector to scale the relative contributions of each component. The iteration number is given by $k$.

The gradient $\mathcal{G}^{(k)}$ of the objective function is given by

$$\mathcal{G}^{(k)}=\mathcal{R}\{\mathbf{J}^t\mathbf{W}_d\Delta\mathbf{d}^*\}, \tag{4}$$

where $\mathbf{J}$ is the Fréchet derivative matrix. Adjoint-state formalism allows efficient computation of the gradient, without explicitly computing $\mathbf{J}$ (Plessix, 2006). This leads to the following expression of the gradient with respect to the parameter $m_i$ (Pratt et al., 1998; Gelis et al., 2007):

$$\mathcal{G}_{m_i}^{(k)}=\mathcal{R}\left\{\mathbf{v}^t\left[\frac{\partial\mathbf{A}^t}{\partial m_i}\right]\mathbf{A}^{-1}\mathbf{W}_d\Delta\mathbf{d}^*\right\}. \tag{5}$$

This shows that the gradient can be recast as a product between the incident wavefields $\mathbf{v}$ and the back-propagated wavefields $\mathbf{A}^{-1}\mathbf{W}_d\Delta\mathbf{d}^*$, using residuals at receiver positions as a composite source. Therefore, only two forward problems per shot are required for gradient building. In equation 5, we exploit the reciprocity of the Green's functions to remove the transpose operator in the expression of the back-propagated residuals $(\mathbf{A}^{-t}\mathbf{W}_d\Delta\mathbf{d}^*=\mathbf{A}^{-1}\mathbf{W}_d\Delta\mathbf{d}^*)$.

The radiation pattern of the diffraction by $m_i$ is denoted by $\partial\mathbf{A}/\partial m_i$. Analysis of these radiation patterns suggests that $V_P$ and $V_S$ parameterization is optimal for large diffraction angles (i.e., for wide-angle reflections), whereas $I_P$ and $I_S$ impedances should provide better decoupling between the two classes of parameters for small diffraction angles (i.e., short-angle reflections; Tarantola, 1986).

A second-order Taylor expansion of the objective function provides the perturbation model $\delta\mathbf{m}$, which minimizes the objective function assumed to be locally parabolic, expressed as

$$\mathbf{B}^{(k)}\delta\mathbf{m}=-\mathcal{G}^{(\mathbf{k})}, \tag{6}$$

where $\mathbf{B}^{(k)}$ is the Hessian of the objective function. Because of high cost to compute $\mathbf{B}^{(k)}$, Newton and Gauss-Newton methods generally are not considered for realistic problems (Pratt et al., 1998). Steepest-descent or conjugate-gradient methods preconditioned by the diagonal terms of an approximate Hessian are used more conventionally (Pratt et al., 1998; Operto et al., 2006). Shin et al. (2001) use the diagonal part of the pseudo-Hessian, a less computationally demanding approximation of the truncated Hessian. Accounting for

the Hessian accelerates the convergence of the inversion and improves the resolution of the imaging by correctly scaling and deconvolving the gradient by geometric amplitude and limited-bandwidth effects.

We use a quasi-Newton method for the FWI problem. The limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method is used commonly in numerical optimization to solve large-scale nonlinear problems (Nocedal, 1980). As indicated, L-BFGS is a limited-memory version of the BFGS method; it appears to be one of the most robust and efficient limited-memory quasi-Newton algorithms (Nocedal and Wright, 1999). The quasi-Newton L-BFGS ($n$) method estimates curvature information contained in the Hessian matrix from a limited number ($n$) of gradient-difference vectors and model difference vectors associated with the $n$ most recent iterations ($n$ is usually between 3 and 20).

The iterative process is preconditioned by an initial guess of the Hessian, typically the diagonal terms of an approximate Hessian. Using the Sherman-Morrison-Woodburry formula (Nocedal and Wright, 1999), at each iteration the L-BFGS algorithm computes an improved estimate of the inverse $\mathbf{H}^{(k)}$ of the Hessian matrix $\mathbf{B}^{(k)}$. Therefore, resolution of equation 6 is avoided and the perturbation model is inferred directly from

$$\delta\mathbf{m} = -\mathbf{H}^{(k)}\mathcal{G}^{(k)}. \tag{7}$$

The double-loop recursive algorithm designed by Nocedal (1980) does not build and store $\mathbf{H}^{(k)}$ explicitly but computes the right-hand side of equation 7 directly with additions, differences, and inner products of the vectors.

Finally, the model is updated with the perturbation vector:

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \alpha\delta\mathbf{m}, \tag{8}$$

where $\alpha$ denotes the step length, which minimizes the objective function. In this study, we estimate $\alpha$ by parabola fitting. When the Hessian matrix is taken into account in the inversion, perturbation models associated with each parameter class computed with equation 7 are scaled correctly with respect to each other (Nocedal and Wright, 1999). Therefore, an estimation of only one step length $\alpha$ is necessary to minimize the objective function for the multiparameter classes. A subspace method (Sambridge et al., 1991) that leads to similar results also has been tested; however, it is more computationally intensive because it requires extra forward-problem simulations.

### Parallel implementation

The massively parallel direct solver MUMPS (Amestoy et al., 2006) is based on a multifrontal method (Duff and Reid, 1983) and it is used to solve the linear system that results from discretizing the forward problem (equation 2). Parallel LU factorization of $\mathbf{A}$ speeds up factorization by more than one order of magnitude compared to sequential execution. Moreover, LU factors are stored in a distributed form over the in-core memory of the processors (Sourbier et al., 2009), making this quite efficient for solving larger problems without intensive input/output (I/O) resources. After the substitution phase, multiple solutions are distributed over the processors following the domain decomposition driven by the distribution of the LU factors.

Equation 5 shows that the gradient computation is a weighted product of the forward problem solutions, i.e., the incident wavefields and the back-propagated residual wavefields. This product can be easily performed in parallel by assigning one processor to each subdomain. To improve the load balancing over the processors, MUMPS-distributed solutions are reordered with message-passing interface (MPI) point-to-point communications before gradient computation, using a well-balanced mesh partitioning that is performed with METIS software (Karypis and Kumar, 1999). The gradient and the initial estimation of the Hessian are therefore efficiently computed in parallel before being centralized on the master processor for perturbation model building.

## FWI DATA PRECONDITIONING AND MULTISCALE STRATEGIES

CPU-efficient frequency-domain FWI is generally conducted by successive inversions of single frequencies, proceeding from low frequencies to higher ones (Pratt and Worthington, 1990; Sirgue and Pratt, 2004). This defines a multiresolution framework that helps mitigate the nonlinearity of the inverse problem associated with high-frequency cycle-skipping artifacts. The CPU-efficient algorithms can be designed by selecting a few coarsely sampled frequencies so that the wavenumber redundancy which results from a dense sampling of frequencies and aperture angles is decimated. This strategy, referred to as the sequential inversion approach, has proven to be effective for several applications of acoustic FWI (Ravaut et al., 2004; Operto et al., 2006; Brenders and Pratt, 2007a). However, it might lack robustness when complex wave phenomena are present. For example, reconstruction of a low-velocity layer in the overthrust model was improved when several frequencies were inverted simultaneously, rather than successively, during acoustic FWI (Sourbier et al., 2009). More significant wave-propagation effects are expected in elastic FWI because of conversions, dispersive surface waves, and frequency-dependent attenuation.

More robust but more computationally expensive multiscale FWI schemes should be designed by preserving partially the redundancy of multifold seismic data. The first scheme, which we refer to as the Bunks approach, is an adaptation in the frequency domain of the time-domain approach of Bunks et al. (1995). It consists of successive inversions of overlapping frequency groups. The first group contains only the starting frequency; one higher frequency is added from one group to the next. The second approach, which we call the simultaneous inversion approach, consists of successive inversions of slightly overlapping frequency groups. It differs from the Bunks approach in that several frequencies are inverted simultaneously at each inversion stage and overlapping between the two next-frequency groups is minimized. Definition of the frequency bandwidth of each group should be driven by the best compromise between the need to avoid high-frequency cycle-skipping artifacts and the need to invert simultaneously as many frequencies as possible to stabilize the inversion.

The nonlinearity of FWI can also be mitigated efficiently by selecting a subset of specific arrivals (i.e., early arrivals, reflected phases) in the data by time windowing (e.g., Sheng et al., 2006; Sears et al., 2008). Frequency-domain wave modeling is less flexible than the time-domain system for preconditioning the data by time windowing, because a limited number of frequencies are processed conventionally at a given step of the inversion. However, data preconditioning can be applied in the frequency domain by complex frequencies ($\omega + i\gamma$), equivalent to damp seismograms in the time domain

(Shin et al., 2002; Brenders and Pratt, 2007b). The Fourier transform of a signal $f(t)$ damped in time by $\exp^{-\gamma(t-t_0)}$ is given by

$$F(\omega + i\gamma)\exp^{\gamma t_0} = \int_{-\infty}^{+\infty} f(t)\exp^{-\gamma(t-t_0)}\exp^{-i\omega t}dt, \quad (9)$$

where the damping can be applied from an arbitrary arrival time $t_0$, which usually corresponds to the first-arrival time.

Time damping applied from the first-arrival time can be viewed as a heuristic way to select aperture angles of P-waves in the data. Arrivals located in time close to the first-arrival times correspond to wide-aperture P-wave events, whereas the later-arriving phase from the first-arrival traveltime corresponds to shorter-aperture P-wave events and converted waves. Because aperture angle is an additional parameter to frequency in controlling the spatial resolution of FWI (Miller et al., 1987; Wu and Toksöz, 1987; Sirgue and Pratt, 2004), aperture selection can be exploited to implement a second level of hierarchy in FWI in addition to that naturally introduced by frequency selection.

Another benefit expected from complex-valued frequencies is the damping during the early FWI iterations of converted P-S waves, free-surface multiples, and surface waves, which introduces additional nonlinearities into the inversion. In practice, this second level of hierarchy can be implemented by progressively relaxing the time damping during each frequency-group inversion. When complex-valued frequencies are used in FWI, the amplitude term $\exp^{\gamma t_0}$ must be introduced in the weighting matrix $\mathbf{S}_d$ (equation 3) to apply the same damping to the partial-derivative wavefields and to the data residuals, which are crosscorrelated during gradient building.

We designed the elastic frequency-domain FWI algorithm so that each of these strategies could be tested easily. The FWI algorithm implements the two nested hierarchical levels through an outer loop over the frequency groups and an inner loop over the damping terms (i.e., the imaginary part of the complex-valued frequencies). Here, a frequency group is a set of real frequencies inverted simultaneously. A third loop is over inversion iteration, and a fourth is over the frequencies of the group. An updated model is produced after one iteration of the inversion involving one frequency group and one damping term.

The major steps of our FWI algorithm are:

**for** *frequency_group* = *group_1* to *group_n* **do**
  **for** *data_damping* = $high_{damping}$ to $low_{damping}$ **do**
    **while** (NOT *convergence* AND *iter* < $niter_{max}$) **do**
      **for** *frequency* = *frequency_1* to *frequency_n* **do**
        Propagate wavefields from sources
        Compute Residuals $\Delta\mathbf{d}$ and Cost Function $\mathcal{C}_{\mathbf{m}}^{(\mathbf{k})}$
        Backpropagate Residuals from receivers
      **end for**
      Build gradient vector $\mathcal{G}_{\mathbf{m}}^{(\mathbf{k})}$
      **if** *iter* = 1 **then**
        Compute diagonal of Pseudo-Hessian
      **end if**
      Compute perturbation vector $\delta\mathbf{m}$ with L-BFGS
      Define optimal step length $\alpha$ by parabola fitting
      Update model $\mathbf{m}^{(\mathbf{k}+1)} = \mathbf{m}^{(\mathbf{k})} + \alpha\delta\mathbf{m}$
    **end while**
  **end for**
**end for**

## APPLICATION TO A CANONICAL MODEL

Full-waveform inversion is an ill-posed, nonlinear problem even for apparently simple models involving few parameters. Mulder and Plessix (2008) analytically illustrate the nonlinearity of 3D acoustic FWI with a 1D velocity-gradient model defined by two parameters. The objective function shows multiple local minima around the true global minimum. Here, we consider a similar two-parameter problem for a 2D elastic model with a free surface on top.

A 1D $V_P$ gradient model defined by $V_P(z) = V_0 + \eta z$ is considered, where $V_0$ is the P-wave velocity at the surface and $\eta$ is the vertical velocity gradient. The S-wave velocity is inferred from the P-wave velocity by considering a constant Poisson's ratio of 0.24. A vertical point force is located just below the free surface; 350 receivers record horizontal and vertical particle velocities on the free surface along a 17-km-long profile.

The objective function is plotted as a function of $V_0$ and $\eta$ for the 5.8-Hz frequency in Figure 1a. The global minimum is located at the coordinates of the true model ($V_0 = 4$ km/s, $\eta = 0.35$ s$^{-1}$). Cross sections along the $\eta$-axis for $V_0 = 4$ km/s and along the $V_0$-axis for $\eta = 0.35$ s$^{-1}$ show the nonconvex shape of the objective function (Figure 1). Multiple local minima are present, particularly on the $\eta$ section, even for this simple gradient model and the low-frequency content in the data.

We repeat the same simulations for damped data using $\gamma = 3.33$ (equation 9). Using this data preconditioning, the objective function is now convex (Figure 1c and d). The convex shape of the objective function should ensure convergence of the inversion toward the global minimum of the objective function for all of the starting models sampled in Figure 1.

## ONSHORE SYNTHETIC CASE STUDY

### SEG/EAGE overthrust model and experimental setup

We considered a $20 \times 4.65$-km dip section of the SEG/EAGE overthrust model to assess the potential of 2D elastic frequency-domain FWI for imaging complex onshore structures (Aminzadeh et al., 1997; Figure 2a). The overthrust model was a $20 \times 20 \times 4.65$-km 3D acoustic model that represents an onshore complex thrusted sedi-

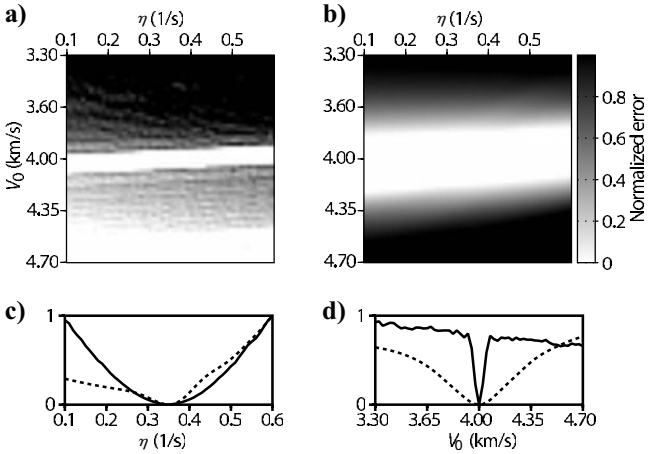


Figure 1. Objective function as a function of the two model parameters for the (a) full and (b) damped data sets. Cross sections for (c) $V_0 = 4$ km/s and (d) $\eta = 0.35$ s$^{-1}$. The solid and dashed lines correspond to the full and the damped data sets, respectively.

mentary succession constructed on top of a basement block. Several faults and channels were present in the model as well as a complex weathering zone on the surface. For elastic FWI, a $V_S$ model was built from the $V_P$ model using a constant Poisson's ratio of 0.24. A uniform density of 1000 kg.m$^{-3}$ was considered and assumed to be known during the inversion. A free surface was set on top of the model.

The onshore, wide-aperture survey consisted of 199 explosive sources spaced every 100 m and located 25 m below the free surface. All of the shots were recorded by 198 vertical and horizontal geophones, spaced every 100 m on the surface. The vertical and horizontal components of particle velocity were used as the data set for the elastic FWI. The data were computed with the same algorithm for observed and computed data in inversion. The source signature was assumed to be known in FWI. An elastic shot gather is shown in Figure 3a and b for the horizontal and vertical components of particle velocity. The corresponding shot gather computed when an absorbing boundary condition is set on top of the model is shown for comparison in Figure 3c and d to highlight the additional wave complexities introduced by free-surface effects (i.e., surface waves and body-wave reflection from the free surface). Of note, the high amplitudes of the surface waves dominate the wavefield, especially on the vertical component (Figure 3a and b).

Starting $V_P$ and $V_S$ models for FWI were computed by smoothing the true velocity models with a 2D Gaussian function of vertical and horizontal correlation ranges of 500 m (Figure 2b). This starting model was proven accurate enough to image the overthrust model by a 2D acoustic frequency-domain FWI using a realistic starting frequency of 3.5 Hz (Sourbier et al., 2009). For the elastic inversions presented here, we used a lower starting frequency of 1.7 Hz. A starting frequency of 3.5 Hz for elastic FWI led to a deficit of long wavelength in the $V_S$ models, which made the inversion converge toward a local minimum.

The different behavior of acoustic and elastic FWI for the overthrust case study highlights the increased sensitivity of elastic FWI with respect to the limited accuracy of the starting models. Five discrete frequencies (1.7, 2.5, 3.5, 4.7, and 7.2 Hz) were used for the elastic FWI. This frequency should allow continuous sampling of the wavenumber spectrum according to the criterion of Sirgue and Pratt (2004).

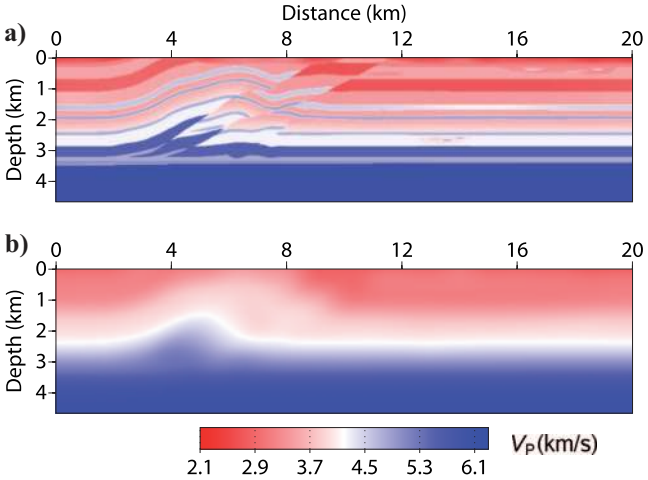


Figure 2. (a) Dip section of the synthetic SEG/EAGE overthrust model. P-wave velocity is depicted. (b) Starting model used for elastic FWI.

We consider three full-waveform inversion data preconditioning and multiscale strategies to manage frequencies: successive inversion of single frequencies, successive inversion of frequency groups of increasing bandwidth, and successive inversion of slightly overlapping frequency groups. For each frequency group, the inversion is subdivided into two steps. In step 1, no offset-dependent gain is applied to the data. Although we scale the gradient by the diagonal terms of the Hessian, we observe some lack of reconstruction in the deep part of the model, suggesting that the near-offset traces have a dominant contribution in the objective function. This layer-stripping effect might provide additional regularization of the inversion, in addition to that provided by the frequency and aperture-angle selections. In step 2, we apply a quadratic gain with offset to the data to strengthen the contribution of long-offset data in the inversion and, hence, to improve the imaging of the deeper part of the model.

During the two-step inversion, the coefficients of the diagonal weighting operator $\mathbf{S}_d$, respectively, are given by

$$\begin{aligned}\mathbf{S}_d &= \exp^{\gamma t_0}\\ \mathbf{S}_d &= \exp^{\gamma t_0}|\text{offset}|^2,\end{aligned} \tag{10}$$

with the reminder that the coefficients $\exp^{\gamma t_0}$ account for offset-dependent time damping (equation 9). For all of the tests presented except the first, we use five damping factors per frequency to precondition the data ($\gamma = 1.5,1.0,0.5,0.1,0.033$). A shot gather computed for the first four damping factors is shown in Figure 4 to illustrate the amount of information preserved in the data. Note how the high damping limits the offset range over which surface waves are seen.

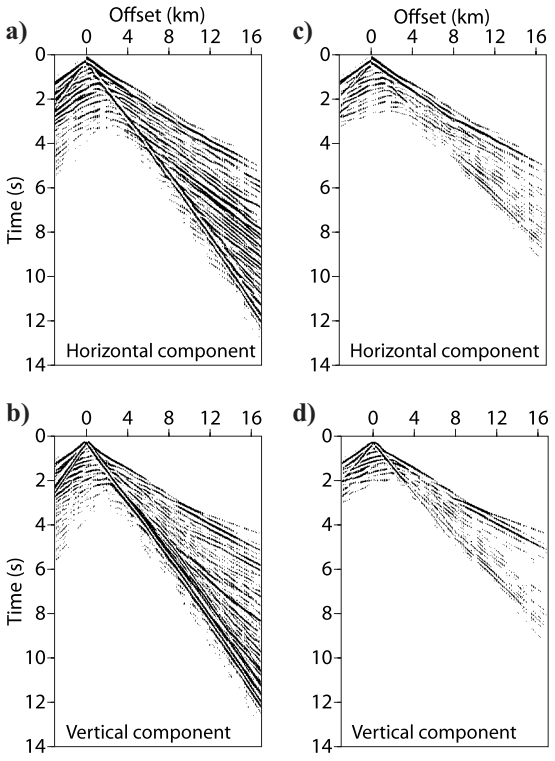


Figure 3. Seismograms computed in the overthrust model for (a) horizontal and (b) vertical components of particle velocity. The shot is located at a horizontal distance of 3 km. A free surface was set on top of the model. (c, d) The same as for (a) and (b) except that an absorbing boundary condition was implemented on top of the model.

Inversion is also regularized by Gaussian smoothing of the perturbation model to cancel high-frequency artifacts in the gradient.

The diagonal terms of the pseudo-Hessian matrix (Shin et al., 2001) provide an initial guess of the Hessian for the L-BFGS algorithm without introducing extra computational costs during gradient building. Five differences of gradients and models vectors are used for the L-BFGS algorithm. The model parameters for inversion are $V_P$ and $V_S$, which are suitable for wide-aperture acquisition geometries (Tarantola, 1986). The loop over the inversion iteration of one complex-valued frequency group is stopped when a maximum iteration of 45 is reached or when the convergence criterion is reached (relative decrease of two successive cost functions lower than $5 \times 10^{-5}$). The schedule of the frequencies and damping terms used in the sequential approach, the Bunks approach, and the simultaneous approach are outlined in Table 1.

Here, we quantify the data misfit for each test with the normalized misfit $\bar{\mathcal{C}}$, defined by

$$\bar{\mathcal{C}} = \frac{\Sigma_{i=1}^{5} \|\Delta \mathbf{d_i}(\mathbf{m_f})\|_2}{\Sigma_{i=1}^{5} \|\Delta \mathbf{d_i}(\mathbf{m_0})\|_2}, \tag{11}$$

where $\Delta \mathbf{d_i}(\mathbf{m_f})$ denotes the data misfit vector for the $i$th frequency and for the final FWI model $\mathbf{m_f}$ and $\mathbf{m}_0$ denotes the starting model shown in Figure 2b.

The FWI model's quality is quantified by

$$mq = \frac{1}{N} \left\| \frac{\mathbf{m_f} - \mathbf{m}_{\text{true}}}{\mathbf{m}_{\text{true}}} \right\|_2, \tag{12}$$

where $\mathbf{m}_{\text{true}}$ denotes the exact model for $V_P$ or $V_S$ and where $N$ is the number of grid points in the computational domain. The normalized misfit and the model quality for the different tests are outlined in Table 2.

## Raw data inversion

A first inversion test was performed without data damping (i.e., without using complex-valued frequencies), which implies that all of the arrivals were involved in the inversion. The five frequencies (Table 1) were inverted successively with the sequential approach. The FWI $V_P$ and $V_S$ models after inversion are shown in Figure 5. The inversion clearly failed to converge toward the true models for the $V_P$ and $V_S$ parameters, even at low frequencies.

## Successive single-frequency inversions of damped data

We repeated the experiment, except that the five damping terms ($\gamma$ = 1.5,1.0,0.5,0.1,0.033) were used to stabilize the inversion (Table 1). The final FWI $V_P$ and $V_S$ models are shown in Figure 6. Contrary to the previous experiment, most of the layers are now reconstructed successfully. A comparison between 1D vertical profiles extracted from the true model, the starting model, and the FWI models shows a reliable estimate of velocity amplitudes despite a low maximum frequency of 7.2 Hz (Figure 7).

To account for the limited bandwidth effect of the source in the FWI model appraisal, we also plotted the vertical profiles of the true models after low-pass filtering at the theoretical resolution of FWI for a maximum frequency of 7.2 Hz. The true models were convert-ed from depth to time using the velocities of the starting model and low-pass filtered with a cut-off frequency of 7 Hz before conversion back to the depth domain. The $V_S$ model is affected more by spurious artifacts than the $V_P$ model, especially in the deep part of the model. This could be caused by a deficit of small wavenumbers in the $V_S$ models resulting from the shorter propagated wavelengths, which makes reconstruction of $V_S$ more nonlinear.

We saw some inaccuracies in the reconstruction of $V_P$ and $V_S$ in the shallowest parts of the models (Figure 7). The resulting residuals of the surface waves and reflections from the free surface might have

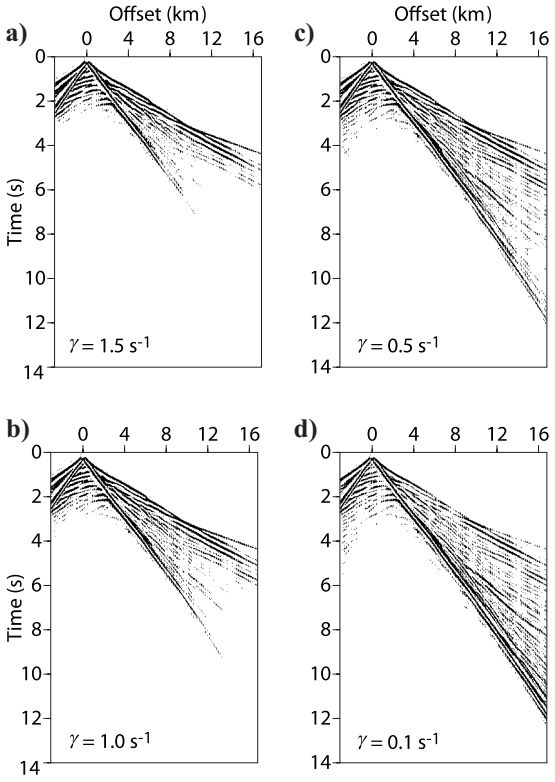


Figure 4. Seismograms for vertical component of particle velocity computed in the dip section of the overthrust model using four values of imaginary frequency: (a) $\gamma$ = 1.5 s$^{-1}$, (b) $\gamma$ = 1.0 s$^{-1}$, (c) $\gamma$ = 0.5 s$^{-1}$, (d) $\gamma$ = 0.1 s$^{-1}$. Time damping was applied from the first-arrival traveltime to preserve long-offset information.

**Table 1. Inversion parameters for the sequential, Bunks, and simultaneous approaches for frequencies within a frequency group (FG).**

| FG | Sequential frequencies (Hz) | Bunks frequencies (Hz) | Simultaneous frequencies (Hz) | Damping factors (1/s) |
|---|---|---|---|---|
| 1 | 1.7 | 1.7 | 1.7, 2.5, 3.5 | 1.5, 1, 0.5, 0.1, 0.033 |
| 2 | 2.5 | 1.7, 2.5 | 3.5, 4.7, 7.2 | 1.5, 1, 0.5, 0.1, 0.033 |
| 3 | 3.5 | 1.7, 2.5, 3.5 | — | 1.5, 1, 0.5, 0.1, 0.033 |
| 4 | 4.7 | 1.7, 2.5, 3.5, 4.7 | — | 1.5, 1, 0.5, 0.1, 0.033 |
| 5 | 7.2 | 1.7, 2.5, 3.5, 4.7, 7.2 | — | 1.5, 1, 0.5, 0.1, 0.033 |

been back-projected erroneously in the deeper part of the model, leading to the above-mentioned inaccuracies (Figure 8). The final normalized $L2$ misfit computed for the five frequencies is $4.12 \times 10^{-1}$. The $V_P$ and $V_S$ model qualities are $5.54 \times 10^{-2}$ and $6.47 \times 10^{-2}$, respectively.

## FWI without free-surface effects

**Table 2. Final $L2$ misfit and model quality for reconstructed models.**

| Approach | Data misfit $\bar{C}$ | $V_P$ | $V_S$ |
|---|---|---|---|
| Sequential | $4.12 \times 10^{-1}$ | $5.54 \times 10^{-2}$ | $6.47 \times 10^{-2}$ |
| Bunks | $1.54 \times 10^{-1}$ | $5.22 \times 10^{-2}$ | $5.33 \times 10^{-2}$ |
| Simultaneous | $1.46 \times 10^{-1}$ | $5.03 \times 10^{-2}$ | $5.39 \times 10^{-2}$ |
| Sequential, no free-surface effects | $9.03 \times 10^{-2}$ | $4.09 \times 10^{-2}$ | $3.89 \times 10^{-2}$ |
| Sequential, PCG optimization | $6.71 \times 10^{-1}$ | $5.56 \times 10^{-2}$ | $6.89 \times 10^{-2}$ |

To assess the footprint of surface waves and free-surface reflections on elastic FWI, we inverted the data computed in the overthrust model with an absorbing boundary condition on top instead of a free surface. The inversion process from the previous section was used (sequential approach with damped data). The final FWI $V_P$ and $V_S$ models were very close to the low-pass-filtered versions of the true models, and they were unaffected by spurious artifacts (Figure 9). The $V_P$ and $V_S$ model qualities are $4.09 \times 10^{-2}$ and $3.89 \times 10^{-2}$, respectively. A comparison with previous results (Figure 6) illustrates the substantial increase of nonlinearity introduced by surface waves and free-surface reflections in elastic FWI.

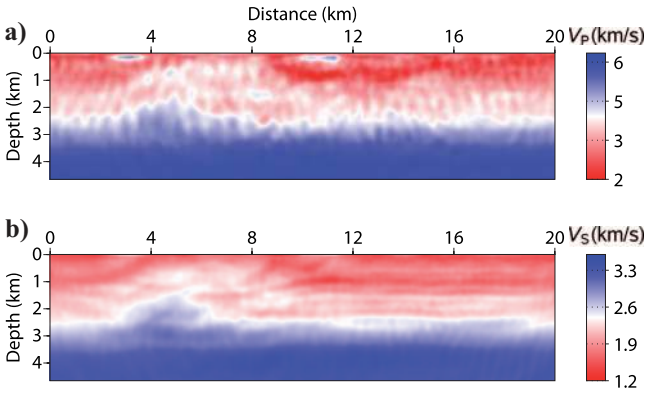


Figure 5. Sequential inversion of raw data for the (a) $V_P$ and (b) $V_S$ models after a frequency of 7.2 Hz.

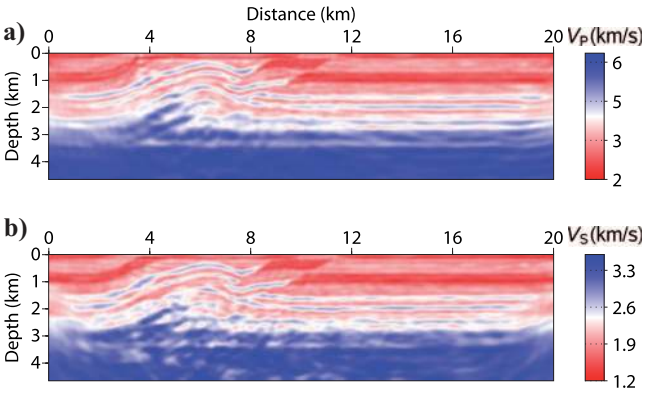


Figure 6. Sequential inversion of damped data for the (a) $V_P$ and (b) $V_S$ models after inversion. The L-BFGS algorithm was used for optimization. Five frequency components were inverted successively. Five damping coefficients were used successively for data preconditioning during each monofrequency inversion.

## Simultaneous multifrequency inversion of damped data

Then we investigated the influence of simultaneous multiple-frequency inversion strategies for FWI improvement. First, we considered the Bunks approach. The different frequency groups and damping terms are outlined in Table 1. Each monofrequency data set belonging to a frequency group was computed with a unit Dirac source wavelet, which implies that each frequency of a group has a similar weight in the inversion. This is equivalent to an inversion of deconvolved data (i.e., data with a flat-amplitude spectrum).

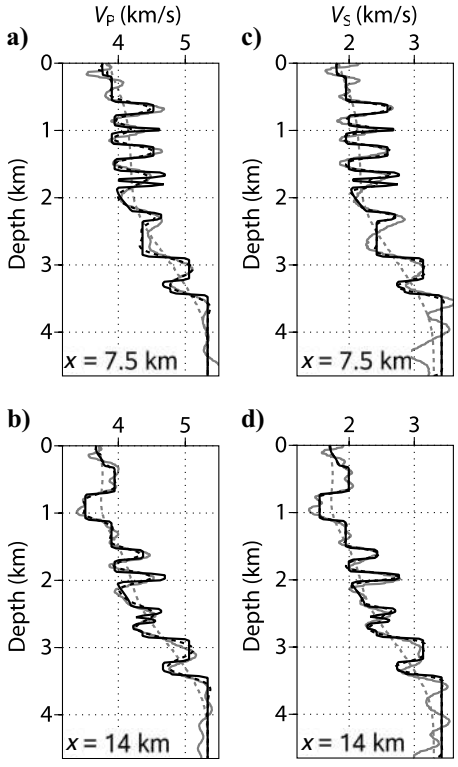


Figure 7. Sequential inversion of damped data — vertical profiles for the (a, b) $V_P$ and (c, d) $V_S$ parameters. Profiles (a) and (c) and profiles (b and d) are at horizontal distances of 7.5 and 14 km, respectively. Profiles of the starting and the true models are plotted with dashed gray and solid black lines, respectively. A low-pass-filtered version of the true model at the theoretical resolution of FWI is plotted with a dashed black line for comparison with the FWI results. The profiles of the FWI models of Figure 6 are plotted with solid gray lines.

The final FWI $V_P$ and $V_S$ models (Figure 10) are slightly better than those of the sequential approach (Figure 6). The improvements are more obvious on the vertical profiles (Figure 11). Near-surface instabilities in the models were mitigated although not fully removed, and estimates of the velocity amplitudes improved in most parts of the model (compare Figures 7 and 11). Mitigation of the near-surface instabilities translated into a significant misfit reduction for the surface waves and free-surface reflections (Figure 12). The final normalized $L2$ misfit decreased to $1.54 \times 10^{-1}$. The $V_P$ and $V_S$ qualities were $5.22 \times 10^{-2}$ and $5.33 \times 10^{-2}$, respectively.

In a second step, we considered the simultaneous approach implemented by successive inversions of two overlapping frequency groups (Table 1). The frequency bandwidth of each group was chosen to avoid cycle-skipping artifacts. For example, we tried to invert the five frequencies listed in Table 1 simultaneously. In this case, the inversion failed to converge toward acceptable models. The computing cost of the simultaneous approach is similar to the sequential approach if the same convergence rate for the two approaches is assumed. The total number of iterations in the simultaneous approach is less important because the iterations are performed over fewer frequency groups at the expense of more factorization and substitution phases per frequency group. The extra cost of the simultaneous approach results only from the overlap between the frequency groups.

The final FWI $V_P$ and $V_S$ velocity models (Figure 13) are improved with respect to the velocity models produced by the sequential approach (Figure 6) and the Bunks approach (Figure 10), especially for the $V_S$ velocity model in the thrust zone. The vertical profiles extracted from the final FWI models no longer show near-surface instabilities (Figure 14), significantly reducing data misfit for the surface waves and free-surface reflections (Figure 15). A significant misfit reduction of the wide-aperture arrivals recorded at large offsets also is seen. The final $L2$ misfit is $1.46 \times 10^{-1}$. The $V_P$ and $V_S$ model qualities are $5.03 \times 10^{-2}$ and $5.39 \times 10^{-2}$, respectively.

We note, however, slightly underestimated velocity amplitudes in the deep part of the $V_P$ and $V_S$ models at the thrust location (see below 3 km depth in Figure 14a and c). We attribute this amplitude deficit to a slower convergence of the simultaneous approach when compared to that of the sequential approach, which results from the fact that more information is inverted simultaneously in the simultaneous approach. The imaging was further improved by decreasing the frequency interval by a factor of two within each frequency group (five frequencies instead of three frequencies per group). This resampling strengthens the spectral redundancy in the imaging. Close-ups of $V_P$ and $V_S$ models centered on the thrust zone show how the resolution and the signal-to-noise ratio of the velocity models still improve by involving more frequencies in one inversion iteration (Figure 16). Using five frequencies instead of three in each group leads to a factor of 5/3 in computing costs.

## L-BFGS versus preconditioned conjugate-gradient optimizations

The sequential approach is applied using a preconditioned conjugate-gradient (PCG) algorithm for numerical optimization. Preconditioning is performed by scaling the gradient by the diagonal terms of the pseudo-Hessian matrix. The same frequencies and damping terms are used as for the L-BFGS run based on the sequential approach (Table 1). The velocity models recovered with the PCG algorithm are shown in Figure 17 and can be compared with the corresponding L-BFGS algorithms (Figure 6). Amplitude estimations and structure focusing are improved by the L-BFGS algorithm (Figure 18), leading to sharper models.

The improvements in model resolution and quantitative estimate of model parameters can be attributed to the approximate estimate of the off-diagonal terms of the Hessian performed by the L-BFGS algorithm. These off-diagonal elements help to deconvolve the models from limited-bandwidth effects resulting from the limited-source bandwidth and the limited extent of the acquisition geometry (e.g., Pratt et al., 1998).

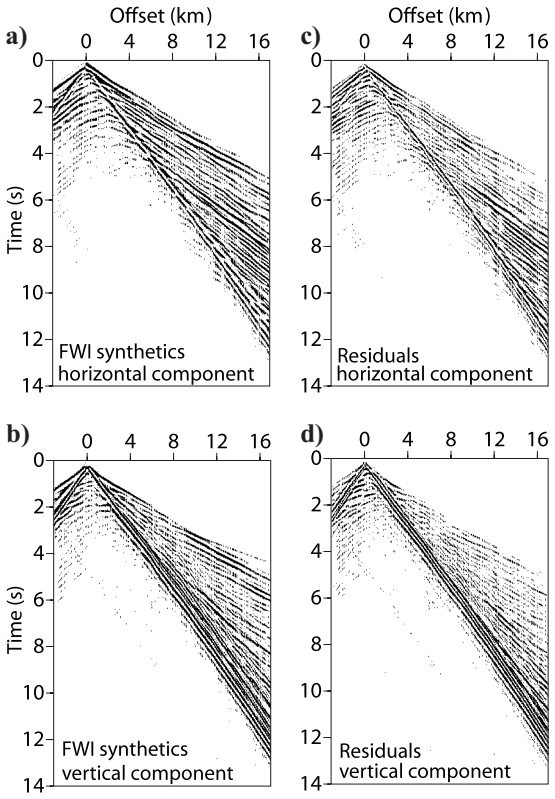


Figure 8. Sequential inversion of damped data. Seismograms computed in the FWI models of Figure 6 for a shot located at a horizontal distance of 3 km. (a) Horizontal component. (b) Vertical component. (c, d) Residuals between seismograms computed in the true $V_P$ and $V_S$ models (Figure 3a and b) and in the FWI models of Figure 6.

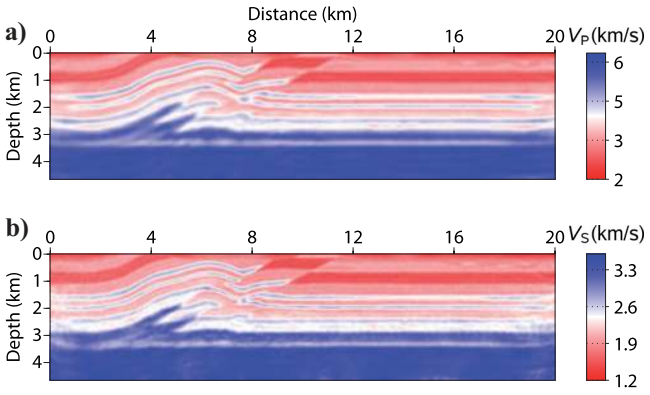


Figure 9. Sequential inversion without free-surface effects for the (a) $V_P$ and (b) $V_S$ models. The models can be compared with those of Figure 6 to assess the footprint of free-surface effects on elastic FWI.

Figure 19 shows the objective functions as a function of iteration number for the L-BFGS and PCG algorithms. The L-BFGS algorithm provides accelerated and improved convergence when compared to PCG. The PCG convergence level (i.e., the minimum value of the objective function reached during optimization) cannot match that of L-BFGS because the off-diagonal information of Hessian estimated by L-BFGS cannot be retrieved by more iterations of PCG. The final $L2$ misfit of PCG is $6.71 \times 10^{-1}$, whereas that of L-BFGS is $4.12 \times 10^{-1}$. These two issues open promising applications of L-BFGS for computationally challenging problems, such as for 3D FWI.

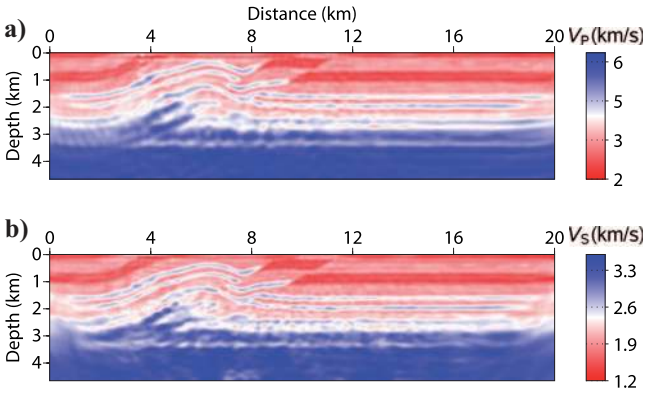


Figure 10. Bunks inversion of the (a) $V_P$ and (b) $V_S$ models obtained with the frequency-domain adaptation of the multiscale approach of Bunks et al. (1995).

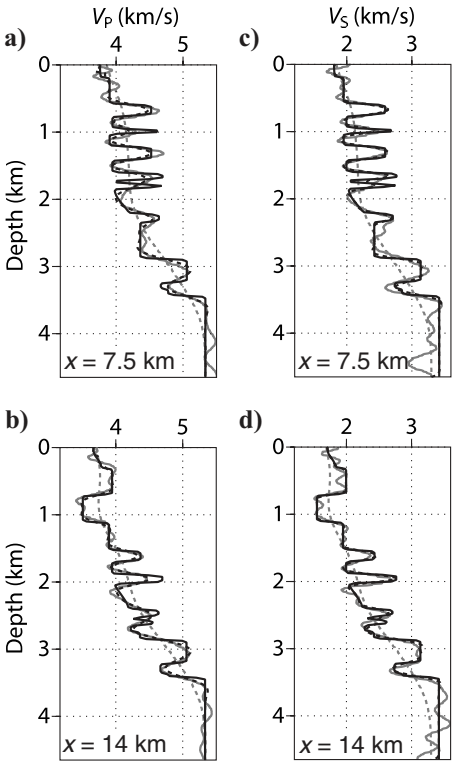


Figure 11. Same as Figure 7 but for the profiles extracted from the models recovered by the Bunks approach (Figure 10).

## Computational aspect

All of the simulations were performed on the cluster at the Simulations Interactives en Géophysique, Astronomie, Mécanique et Mathématiques (SIGAMM) computer center, which is a 48-node cluster with each node comprising two dual-core 2.4-GHz Opteron processors, providing 19.2 Gflops peak performance per node. This computer has a distributed-memory architecture, where each node has 8 GB of RAM. The interconnection network between processors is Infiniband 4X. Twenty-four processors were used for each simulation, leading to the best compromise between execution time and numerical resources used. A single regular equilateral mesh composed of 265,675 cells was designed for the simulations. Although the mesh could have been adapted to the inverted frequency we did not consider this strategy here and the mesh was kept constant whatever the inverted frequency.

Table 3 outlines the memory requirements and computing time of the major tasks performed by the parallel FWI algorithm. Of note, most of the memory and computational time were dedicated to LU factorization and substitution phases performed during the multisource forward problem. Computing the gradient had a negligible computational cost because of the domain-decomposition parallelism. The L-BFGS algorithm required a negligible extra amount of memory and computing time compared to a steepest-descent or PCG algorithm, suggesting that this optimization scheme can be used efficiently for realistic 2D and 3D FWI applications.

## DISCUSSION

Elastic FWI applied to the overthrust model highlights the strong nonlinearity of the inversion resulting from free-surface effects. The impact of these effects on FWI can be assessed by comparing the FWI results inferred from the data including or without the free-surface effects (compare Figures 6 and 9). The best models are obtained when free-surface effects are not considered, showing for this case study that inversion of the surface waves did not improve the reconstruction of the near-surface structure.

We interpret the failure of the raw-data inversion as the footprint of surface waves, the amplitudes of which dominate the wavefield and carry no information about the deep part of the model (Figure 5). Similar effects of surface waves on elastic FWI are also seen on a smaller scale by Gelis et al. (2007). A comparison between the sequential FWI results obtained with the raw data and the preconditioned data illustrates how time damping helps to mitigate the nonlinearities of FWI by injecting progressively more complex wave phenomena in the inversion (Figures 5 and 6).

The improvements obtained by simultaneous inversion of multiple frequencies combined with hierarchical inversions of damped data show that preserving some wavenumber redundancy in elastic FWI is critical to mitigating the nonlinearity of the inversion associated with propagating surface waves in weathered near-surface layers and free-surface reflections. The more stable results obtained with the simultaneous approach compared to the Bunks approach, especially in the near surface, suggest that several frequencies must be inverted simultaneously from the early stage of the inversion (compare Figures 10 and 13). Strengthening the wavenumber redundancy by decreasing the frequency interval in each frequency group further improves imaging (Figure 16).

Another factor that increases the nonlinearity of elastic FWI is the short S-wavelengths, which could require more accurate starting models or lower frequencies to converge toward an acceptable mod-

el. The maximum frequency of the starting frequency group must be chosen so that it prevents cycle-skipping artifacts that can result from the limited accuracy of the starting S-wave velocity model. Laplace-domain waveform inversion, proposed as a reliable approach to build smooth initial elastic models of the subsurface (Pyun et al., 2008; Shin and Cha, 2008), might represent a way to build the starting model. An alternative approach is PP-PS stereotomography (Alerini et al., 2002), including joint inversion of refraction and reflection traveltimes of wide-aperture data.

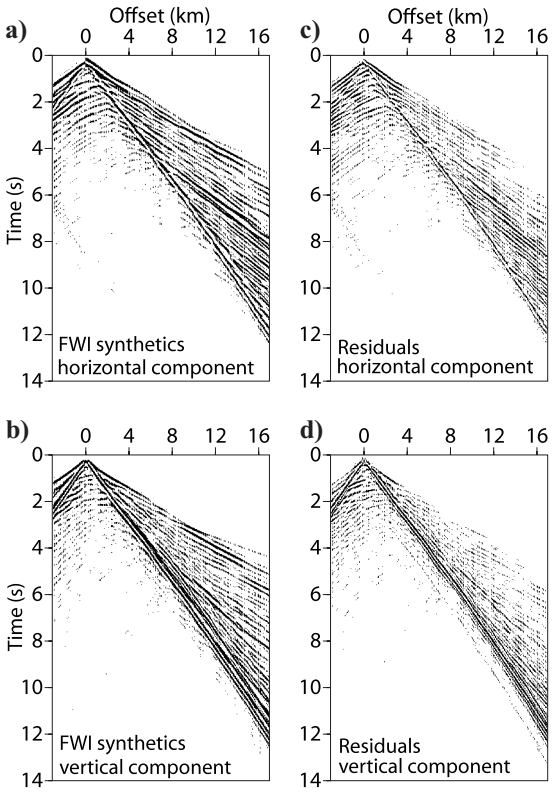


Figure 12. Same as Figure 8 but for seismograms computed in the FWI models recovered by the Bunks approach (Figure 10).

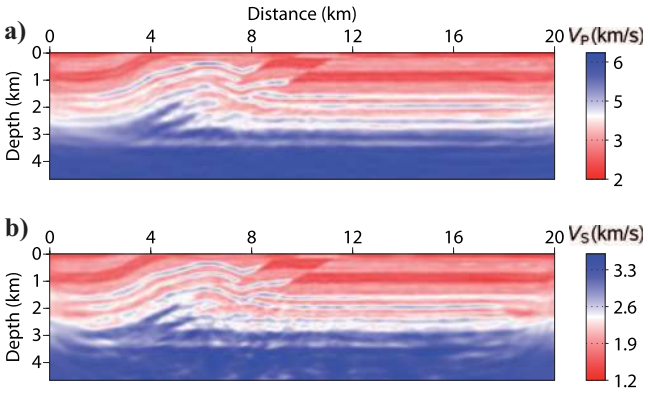


Figure 13. Simultaneous inversion. Final models obtained by successive inversions of two overlapping frequency groups composed of three frequencies each for (a) $V_P$ and (b) $V_S$ models.

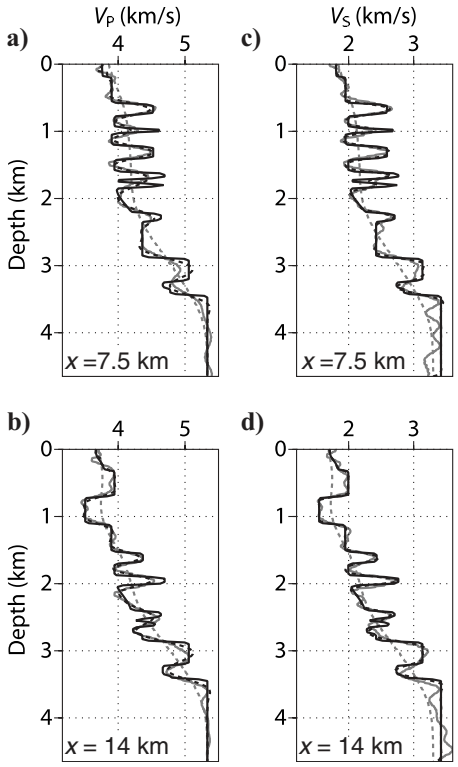


Figure 14. Same as Figure 7 but for the profiles extracted from the models recovered by the simultaneous approach (Figure 13).

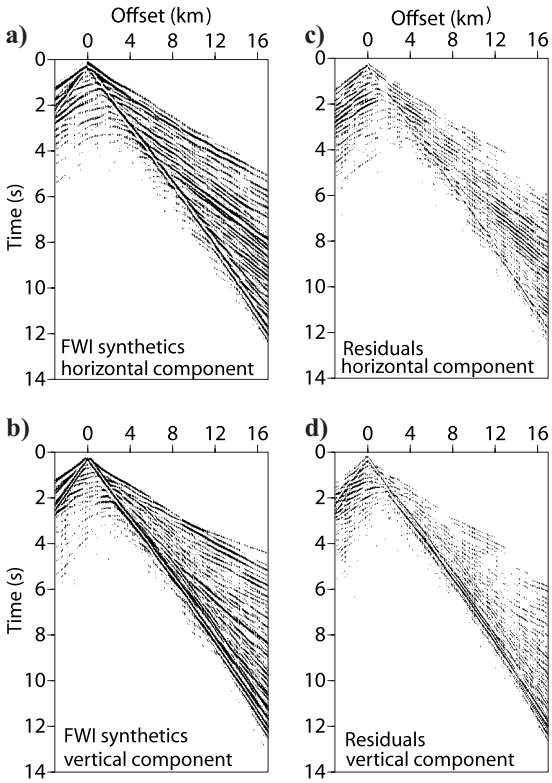


Figure 15. Same as Figure 8 but for seismograms computed in the FWI models recovered by successive inversion of two overlapping frequency groups (Figure 13).

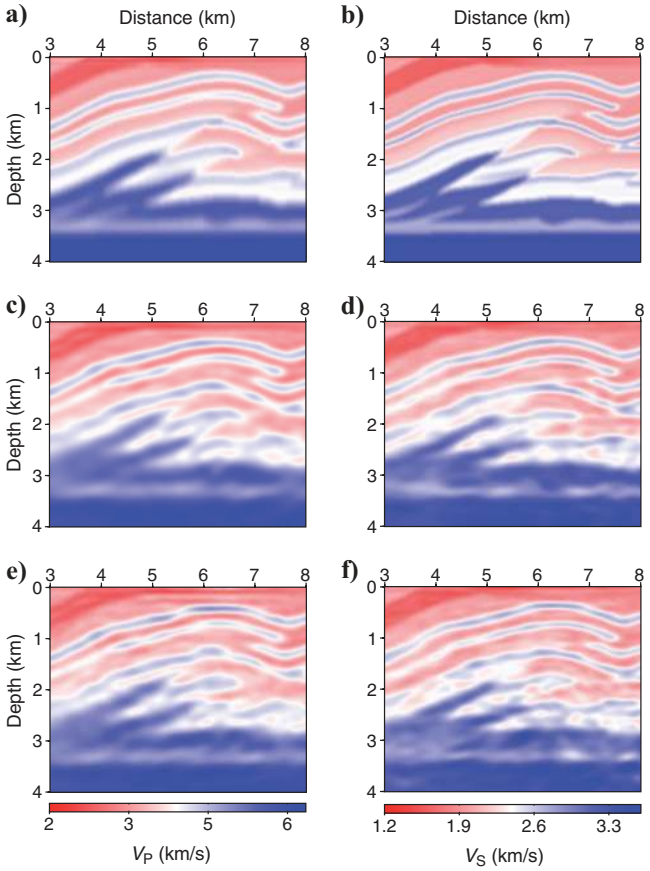


Figure 16. Comparison between FWI models obtained by the successive inversion of two overlapping groups of frequencies (simultaneous approach) when three and five frequencies per group are used in the inversion, respectively. The frequency bandwidth is the same for each experiment, but the frequency interval differs. (a) Close-up of the true $V_P$ model after low-pass filtering at the theoretical resolution of FWI. (c, e) Close-up of the FWI $V_P$ model when five frequencies (c) and three frequencies (e) per group are used, respectively. (b, d, f) Same as (a, c, e) for the $V_S$ model.

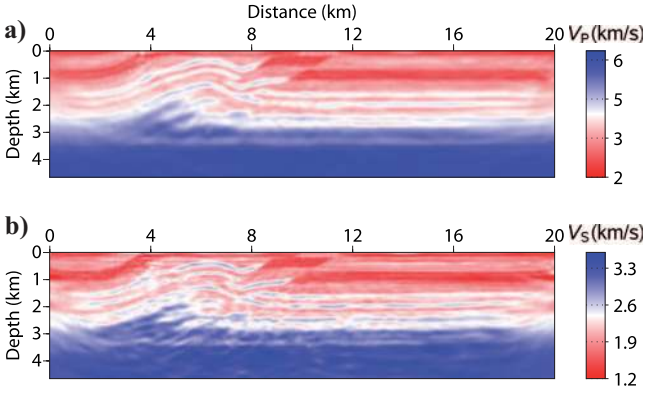


Figure 17. Full-waveform inversion velocity models obtained with the PCG algorithm for (a) $V_P$ and (b) $V_S$ parameters. The sequential approach with five damping terms was used. The velocity models can be compared to those recovered by the L-BFGS algorithm (Figure 6).

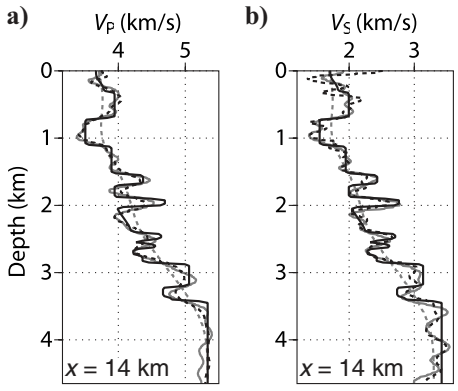


Figure 18. Comparison between velocity profiles extracted from the FWI models recovered by L-BFGS (solid gray line) and the PCG algorithm (dashed black line) for (a) $V_P$ and (b) $V_S$ models. Starting and true models are depicted with dashed gray and solid black lines, respectively.

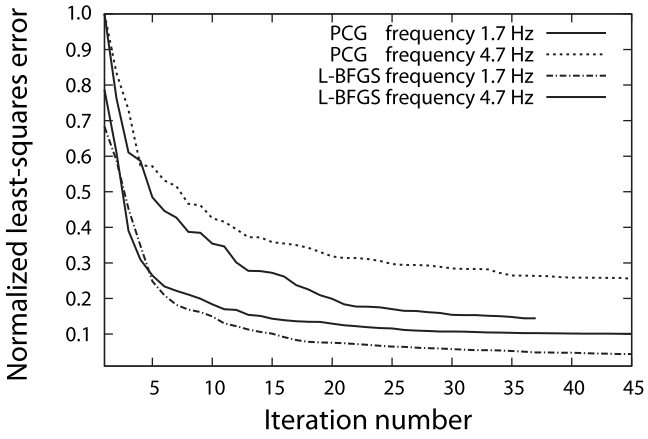


Figure 19. The L-BFGS and PCG objective functions plotted as a function of iteration number for the inversion of the complex frequencies $(1.7 + i1.5)$ Hz and $(4.7 + i1.5)$ Hz. The curves associated with one frequency are normalized by the PCG objective functions at the first iteration. Other frequencies and damping factors show similar trends.

**Table 3. Computing values for the main tasks performed by FWI, averaged over several iterations for 24 processors.**

| Task | Value |
|---|---|
| Time for factorization (s) | 21.3 |
| MUMPS total memory for factorization (Gb) | 4.8 |
| Time for substitutions (199 shots) (s) | 22.9 |
| MUMPS total memory for substitution (Gb) | 20.2 |
| Time for gradient buildup (s) | 13.8 |
| Time for L-BFGS perturbation computation (s) | 1.2 |
| Memory for L-BFGS(5) history (Mb) | 21.2 |

With this case study, we were unable to illustrate the usefulness of the surface waves to reconstruct the near-surface structure because the most accurate FWI models were inferred without involving free-surface effects in the data (Figure 9). Instead, we have shown how to manage the nonlinearities introduced by the surface waves via judicious data preconditioning and FWI tuning. Alternatively, the surface waves in the recorded and modeled data can be filtered out or muted. We did not investigate this approach because efficient filtering of the modeled surface waves in the frequency domain is not straightforward. The surface waves must be filtered out not only at the receiver positions but also at each position in the computational domain, where they have significant amplitudes to remove their footprint from the gradient of the objective function. This investigation requires further work.

Realistic applications of elastic FWI in more complex models still must be investigated. Areas of complex topography such as foothills will lead to conversions from surface waves to body waves and vice versa, which could carry additional information on the near surface. The robustness of elastic FWI for imaging models with heterogeneous Poisson's ratios is a second field of investigation, especially in areas of soft seabeds where the P-S-converted wavefield might have a limited signature in the data (Sears et al., 2008). In our study, we inverted data computed with a constant density that was assumed to be known. For real data inversion, a density estimate is required for a more reliable amplitude match. A reliable density estimate by FWI is difficult because the P-wave velocity and density have similar radiation patterns at short apertures (Forgues and Lambaré, 1997). The benefit provided by wide apertures to uncouple these two parameters should be investigated.

Other extensions of isotropic elastic FWI could relate to reconstructing attenuation factors and some anisotropic parameters. Vertical transversely isotropic elastic FWI should be implemented easily from isotropic elastic FWI because only the expression of the coefficients of the P-SV elastodynamic system needs to be modified compared to the isotropic case (e.g., Carcione et al., 1988).

Application of 2D elastic FWI to real data will require additional data preprocessing that is not addressed in this study, such as source estimation (Pratt, 1999) and 3D to 2D amplitude corrections (Bleinstein, 1986; Williamson and Pratt, 1995). The sensitivity of the elastic FWI to the approximations and errors underlying this processing will need to be determined.

## CONCLUSIONS

Our study presents a new, massively parallel 2D elastic frequency-domain FWI algorithm applied to a dip section of the SEG/EAGE onshore overthrust model. Strong nonlinearities of elastic FWI arise from the presence of converted and surface waves and from the limited accuracy of the $V_S$ starting model. These factors prevent the convergence of FWI on the global minimum of the objective function if specific preconditioning is not applied to the data and no low starting frequency is available. Data preconditioning performed by time damping is necessary to converge toward acceptable velocity models, whatever the frequency sampling strategy.

Successive inversions of overlapping frequency groups outperform successive inversions of single frequencies for removing instabilities in the near surface of the FWI models. The bandwidth of the frequency groups must be chosen so that cycle-skipping artifacts are avoided while injecting a maximum amount of redundant information into the frequency groups. The quasi-Newton algorithm of L-BFGS outperforms the most popular PCG algorithm in terms of convergence rate and convergence level without significant extra computational costs and, hence, is very useful for this application.

## ACKNOWLEDGMENTS

This study was funded by the Seiscope consortium, sponsored by BP, CGG-Veritas, Exxon-Mobil, Shell, and TOTAL, and by Agence Nationale de la Recherche (ANR) under project ANR-05-NT05-2-42427. The linear systems were solved with MUMPS package, available at http://graal.ens-lyon.fr/MUMPS/index.html. The mesh generation was performed with help of Triangle, available at http://www.cs.cmu.edu/~quake/triangle.html. Access to the high-performance computing facilities of the SIGAMM (Observatoire de la Côte d'Azur) and the IDRIS computer center (project 082280) provided the required computer resources, and we gratefully acknowledge both of these facilities and the support of their staff. Many thanks go to Serge Gratton (CERFACS and CNES, Toulouse, France) for interesting and stimulating discussions on quasi-Newton optimization methods. We would like to thank Isabelle Lecomte, Christophe Barnes, Hansruedi Maurer, and Changsoo Shin for their fruitful comments on the manuscript.

## REFERENCES

Alerini, M., S. L. Bégat, G. Lambaré, and R. Baina, 2002, 2D PP- and PS-stereotomography for a multicomponent data set: 72nd Annual International Meeting, SEG, Expanded Abstracts, 838–841.

Amestoy, P. R., A. Guermouche, J. Y. L'Excellent, and S. Pralet, 2006, Hybrid scheduling for the parallel solution of linear systems: Parallel Computing, **32**, 136–156.

Aminzadeh, F., J. Brac, and T. Kunz, 1997, 3-D salt and overthrust models: SEG.

Barnes, C., and M. Charara, 2008, Full-waveform inversion results when using acoustic approximation instead of elastic medium: 78th Annual International Meeting, SEG, Expanded Abstracts, 1895–1899.

Ben-Hadj-Ali, H., S. Operto, and J. Virieux, 2008, Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data: Geophysics, **73**, no. 5, VE101–VE117.

Berenger, J.-P., 1994, A perfectly matched layer for absorption of electromagnetic waves: Journal of Computational Physics, **114**, 185–200.

Bleibinhaus, F., J. A. Hole, T. Ryberg, and G. S. Fuis, 2007, Structure of the California Coast ranges and San Andreas fault at SAFOD from seismic waveform inversion and reflection imaging: Journal of Geophysical Research, **112**, B06315, doi: 10.1029/2006JB004611.

Bleinstein, N., 1986, Two-and-one-half dimensional in-plane wave-propagation: Geophysical Prospecting, **34**, 686–703.

Brenders, A. J., and R. G. Pratt, 2007a, Efficient waveform tomography for lithospheric imaging: Implications for realistic 2D acquisition geometries and low frequency data: Geophysical Journal International, **168**, 152–170.

——, 2007b, Full waveform tomography for lithospheric imaging: Results from a blind test in a realistic crustal model: Geophysical Journal International, **168**, 133–151.

Brossier, R., J. Virieux, and S. Operto, 2008, Parsimonious finite-volume frequency-domain method for 2-D P-SV-wave modelling: Geophysical Journal International, **175**, 541–559.

Buland, A., and H. Omre, 2003, Bayesian linearized AVO inversion: Geophysics, **68**, 185–198.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.

Carcione, J. M., D. Kosloff, and R. Kosloff, 1988, Wave-propagation simulation in an elastic anisotropic (transversely isotropic) solid: Quarterly Journal of Mechanics and Applied Mathematics, **41**, 319–345.

Choi, Y., D. Min, and C. Shin, 2008, Two-dimensional waveform inversion of multi-component data in acoustic-elastic coupled media: Geophysical Prospecting, **56**, 863–881.

Choi, Y., and C. Shin, 2008, Frequency-domain elastic full waveform inversion using the new pseudo-Hessian matrix: Experience of elastic Marmousi 2 synthetic data: Bulletin of the Seismological Society of America, **98**, 2402–2415.

Crase, E., A. Pica, M. Noble, J. McDonald, and A. Tarantola, 1990, Robust elastic nonlinear waveform inversion: Application to real data: Geophysics, **55**, 527–538.

Crase, E., C. Wideman, M. Noble, and A. Tarantola, 1992, Nonlinear elastic inversion of land seismic reflection data: Journal of Geophysical Research, **97**, 4685–4705.
Duff, I. S., and J. K. Reid, 1983, The multifrontal solution of indefinite sparse symmetric linear systems: ACM Transactions on Mathematical Software, **9**, 302–325.
Dumbser, M., and M. Käser, 2006, An arbitrary high order discontinuous Galerkin method for elastic waves on unstructured meshes II: The three-dimensional isotropic case: Geophysical Journal International, **167**, 319–336.
Forgues, E., and G. Lambaré, 1997, Parameterization study for acoustic and elastic ray + Born inversion: Journal of Seismic Exploration, **6**, 253–278.
Gao, F., A. R. Levander, R. G. Pratt, C. A. Zelt, and G. L. Fradelizio, 2006, Waveform tomography at a groundwater contamination site: $V_{sp}$-surface data set: Geophysics, **71**, no. 1, H1–H11.
Gelis, C., J. Virieux, and G. Grandjean, 2007, 2D elastic waveform inversion using Born and Rytov approximations in the frequency domain: Geophysical Journal International, **168**, 605–633.
Hicks, G. J., and R. G. Pratt, 2001, Reflection waveform inversion using local descent methods: Estimating attenuation and velocity over a gas-sand deposit: Geophysics, **66**, 598–612.
Jin, S., G. Cambois, and C. Vuillermoz, 2000, Shear-wave velocity and density estimation from PS-wave AVO analysis: Application to an OBS data set from the North Sea: Geophysics, **65**, 1446–1454.
Karypis, G., and V. Kumar, 1999, A fast and high quality multilevel scheme for partitioning irregular graphs: SIAM Journal on Scientific Computing, **20**, 359–392.
Käser, M., and M. Dumbser, 2006, An arbitrary high order discontinuous Galerkin method for elastic waves on unstructured meshes I: The two-dimensional isotropic case with external source terms: Geophysical Journal International, **166**, 855–877.
Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, **49**, 533–549.
Martin, G. S., R. Wiley, and K. J. Marfurt, 2006, Marmousi2: An elastic upgrade for Marmousi: The Leading Edge, **25**, 156–166.
Miller, D., M. Oristaglio, and G. Beylkin, 1987, A new slant on seismic imaging: Migration and integral geometry: Geophysics, **52**, 943–964.
Mora, P. R., 1987, Nonlinear two-dimensional elastic inversion of multi-offset seismic data: Geophysics, **52**, 1211–1228.
———, 1988, Elastic wavefield inversion of reflection and transmission data: Geophysics, **53**, 750–759.
Mulder, W., and R.-E. Plessix, 2008, Exploring some issues in acoustic full waveform inversion: Geophysical Prospecting, **56**, 827–841.
Nihei, K. T., and X. Li, 2007, Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD): Geophysical Journal International, **169**, 1069–1078.
Nocedal, J., 1980, Updating quasi-Newton matrices with limited storage: Mathematics of Computation, **35**, 773–782.
Nocedal, J., and S. J. Wright, 1999, Numerical optimization: Springer-Verlag.
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.
Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.
Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part I: Theory and verification in a physic scale model: Geophysical Journal International, **64**, 888–901.
Pratt, R. G., C. Shin, and G. J. Hicks, 1998, Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion: Geophysical Journal International, **133**, 341–362.
Pratt, R. G., and R. M. Shipp, 1999, Seismic waveform inversion in the frequency domain, Part II: Fault delineation in sediments using crosshole data: Geophysics, **64**, 902–914.
Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography, Part 1: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310.
Pyun, S., C. Shin, H. Lee, and D. Yang, 2008, 3D elastic full waveform inversion in the Laplace domain: 78th Annual International Meeting, SEG, Expanded Abstracts, 1976–1980.
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.
Sambridge, M. S., A. Tarantola, and B. L. Kennett, 1991, An alternative strategy for non-linear inversion of seismic waveforms: Geophysical Prospecting, **39**, 723–736.
Sears, T., S. Singh, and P. Barton, 2008, Elastic full waveform inversion of multi-component OBC seismic data: Geophysical Prospecting, **56**, 843–862.
Sheng, J., A. Leeds, M. Buddensiek, and G. T. Schuster, 2006, Early arrival waveform tomography on near-surface refraction data: Geophysics, **71**, no. 4, U47–U57.
Shi, Y., W. Zhao, and H. Cao, 2007, Nonlinear process control of wave-equation inversion and its application in the detection of gas: Geophysics, **72**, no. 1, R9–R18.
Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931.
Shin, C., S. Jang, and D. J. Min, 2001, Improved amplitude preservation for prestack depth migration by inverse scattering theory: Geophysical Prospecting, **49**, 592–606.
Shin, C., D.-J. Min, K. J. Marfurt, H. Y. Lim, D. Yang, Y. Cha, S. Ko, K. Yoon, T. Ha, and S. Hong, 2002, Traveltime and amplitude calculations using the damped wave solution: Geophysics, **67**, 1637–1647.
Shipp, R. M., and S. C. Singh, 2002, Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data: Geophysical Journal International, **151**, 325–344.
Sirgue, L., J. Etgen, and U. Albertin, 2007, 3D full-waveform inversion: Wide-versus narrow-azimuth acquisitions: 77th Annual International Meeting, SEG, Expanded Abstracts, 1760–1764.
Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248.
Sourbier, F., S. Operto, J. Virieux, P. Amestoy, and J.-Y. L'Excellent, 2009, FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data, Part 2: Numerical examples and scalability analysis: Computers & Geosciences, **35**, 496–514.
Tarantola, A., 1986, A strategy for nonlinear inversion of seismic reflection data: Geophysics, **51**, 1893–1903.
———, 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier.
Vigh, D., and E. W. Starr, 2008, 3D prestack plane-wave, full-waveform inversion: Geophysics, **73**, no. 5, VE135–VE144.
Warner, M., I. Stekl, and A. Umpleby, 2008, 3D wavefield tomography: Synthetic and field data examples: 78th Annual International Meeting, SEG, Expanded Abstracts, 3330–3334.
Williamson, P., and G. Pratt, 1995, A critical review of 2.5D acoustic wave modeling procedures: Geophysics, **60**, 591–595.
Wu, R.-S., and M. N. Toksöz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11–25.