Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data

Hafedh Ben-Hadj-Ali[1], Stéphane Operto[1], and Jean Virieux[2]

ABSTRACT

We assessed 3D frequency-domain (FD) acoustic full-waveform inversion (FWI) data as a tool to develop high-resolution velocity models from low-frequency global-offset data. The inverse problem was posed as a classic least-squares optimization problem solved with a steepest-descent method. Inversion was applied to a few discrete frequencies, allowing management of a limited subset of the 3D data volume. The forward problem was solved with a finite-difference frequency-domain method based on a massively parallel direct solver, allowing efficient multiple-shot simulations. The inversion code was fully parallelized for distributed-memory platforms, taking advantage of a domain decomposition of the modeled wavefields performed by the direct solver. After validation on simple synthetic tests, FWI was applied to two targets (channel and thrust system) of the 3D SEG/EAGE overthrust model, corresponding to 3D domains of $7 \times 8.75 \times 2.25$ km and $13.5 \times 13.5 \times 4.65$ km, respectively. The maximum inverted frequencies are 15 and 7 Hz for the two applications. A maximum of 30 dual-core biprocessor nodes with 8 GB of shared memory per node were used for the second target. The main structures were imaged successfully at a resolution scale consistent with the inverted frequencies. Our study confirms the feasibility of 3D frequency-domain FWI of global-offset data on large distributed-memory platforms to develop high-resolution velocity models. These high-velocity models may provide accurate macromodels for wave-equation prestack depth migration.

INTRODUCTION

Three-dimensional quantitative seismic imaging in complex environments (e.g., deep water, thrust belts, subsalt and subbasalt structures) is a primary challenge of seismic exploration for hydrocarbon exploitation. In the depth domain, the imaging flowchart for multichannel seismic reflection data is subdivided into two main steps: velocity macromodel estimation and prestack depth migration (PSDM). These steps typically are performed iteratively until flattening of reflectors in common image gathers (CIGs) is optimized. The human interactions during several tasks related to velocity model-building, such as CIG flattening, layer interpretation, and quality control of picking, makes the PSDM workflow time consuming and potentially subjective. Therefore, any approach that helps to automate and optimize velocity model-building will speed up the output of the final PSDM image.

Estimating the velocity macromodel is critical because it has a strong impact on the accuracy of the migrated images in terms of focusing and positioning in depth of the reflectors. The criteria that the velocity macromodel must verify to provide accurate migrated images are still unclear (for illustrations of the sensitivity of 2D and 3D true-amplitude PSDM to the accuracy of the velocity macromodel, see Operto et al., 2000, 2003). Estimating a reliable velocity macromodel for PSDM from conventional multichannel seismic reflection data is a difficult task — one that becomes even more dramatic in complex environments because of the velocity-depth ambiguity at significant depths (Bickel, 1990; Pon and Lines, 2005).

The most common approaches for building a PSDM velocity model rely on reflection traveltime tomography (e.g., Bishop et al., 1985; Stork, 1992) or migration velocity analysis (e.g., Biondi and Symes, 2004; Sava and Biondi, 2005). Both approaches result in approximations for modeling wave propagation, such as the high-frequency approximation or the one-way approximation of the wave equation. This incomplete modeling of wave propagation, together with the limitations imposed by narrow-aperture acquisition geometries, can prevent imaging of steeply dipping reflectors. For example, Zhang et al. (2006) illustrate that turning waves and multireflected arrivals can improve images of the flank of salt bodies. As a result, research has been dedicated to extending the one-way propagator to exploit these arrivals in PSDM (e.g., Zhang et al., 2007).

In this paper, we investigate frequency-domain (FD) full-waveform inversion (FWI) of wide-aperture data as a tool to build 3D

high-resolution velocity models in complex environments (Pratt, 2004). By a wide-aperture acquisition survey, referred to as *global-offset acquisition*, we mean any acquisition geometry with sufficiently long offset coverage to record diving waves whose refraction depths cover the zone of interest. Wide-aperture arrivals such as diving waves and supercritical reflections are sensitive to the large and intermediate wavelengths of a medium (Pratt and Worthington, 1990; Sirgue and Pratt, 2004), which are difficult to image from multichannel seismic reflection acquisition and limited-bandwidth sources. Moreover, multifold wide-aperture surveys lead to a redundant control of frequency and aperture angle on the wavenumber illumination in the model space. This redundancy can be decimated to design efficient numerical approaches for seismic imaging in the frequency domain (Pratt and Worthington, 1990; Pratt, 1999). Global-offset acquisition surveys can be conducted at sea or on land with a network of stations (see Clarke et al. [2007] for a recent 3D wide-azimuth node survey).

FWI refers to a quantitative imaging method based on a complete solution of the full (two-way) wave equation for the forward problem and on inverse problem theory for the imaging problem (Tarantola, 1987). An improved model is built by minimizing the misfit between the recorded data and the data computed in the model. FWI shares some similarities with generalized diffraction tomography (Wu and Töksoz, 1987; Pratt et al., 1998). The partial-derivative and misfit wavefields can be interpreted as the wavefields emitted by the shots and scattered by secondary sources (virtual sources in Pratt et al. [1998, their equation 16]), triggered at the position of the heterogeneities lacking in the starting model. Zero-lag correlation between the misfit and the partial-derivative wavefields at the receiver locations provides an unscaled image of the missing heterogeneities (the so-called perturbation model) in the opposite direction of the gradient of the least-squares objective function.

The heterogeneities can be represented by a series of closely spaced diffractors. By virtue of the Huygens' principle, an image of the perturbation model is built by summing the elementary images of each diffractor. The gradient of the objective function can be computed more efficiently by zero-lag convolution of the incident wavefields with the backpropagated residual wavefields, thanks to the spatial reciprocity of the Green's function. The zero-lag convolution between the incident wavefields and the backpropagated residuals is similar to the imaging principle of reverse time migration originally proposed by Claerbout (1971) and recast in the framework of inverse problem theory by Lailly (1984) and Tarantola (1984).

There are two main drawbacks of FWIs. First, they are very expensive computationally because of the complete resolution of the wave equation for a large number of sources. Second, they lack robustness as a result of the complexity of the global-offset wavefields and their sensitivity to noise and to the inaccuracies of the starting model. In the 2D case, the FD formulation of FWI applied to global-offset acquisition provides a promising approach to mitigate these difficulties (Pratt, 1999; Sirgue and Pratt, 2004; Brenders and Pratt, 2007a). The extension of this approach to three dimensions is investigated in this paper.

The FD formulation of FWI was developed originally for 2D crosshole acquisition surveys, which involve wide-aperture recording (Song et al., 1995; Pratt, 1999). Because of the wavenumber redundancy provided by multifold wide-aperture geometries, only a few discrete frequencies are required to develop a reliable image of the medium. Some guidelines to define the optimal frequency interval for FWI are given in Sirgue and Pratt (2004). This FD decimation

leads to a very compact volume of data to be managed, which may be critical for 3D applications.

The presence of many local minima in the least-squares objective function can prevent convergence of FWI based on local optimization toward the global minimum of the objective function. This originally motivated development of multiscale strategies in the time domain through successive inversions of subdata sets of increasing frequency bandwidth (Bunks et al., 1995). The FD formulation of FWI provides a more natural and flexible framework with which to design a hierarchical multiresolution imaging strategy, helping to manage the inherent nonlinearity of the inverse problem.

The multiscale approach in the frequency domain is generally implemented by successive inversions of single frequencies of increasing value (Pratt and Worthington, 1990; Pratt, 1999). This strategy differs from that of Bunks et al. (1995), which does not take advantage of the redundant control of frequency and aperture on wavenumber coverage. Indeed, in the approach of Bunks et al. (1995), higher frequencies are injected at a given step of the multiscale approach while keeping the frequencies from the previous steps involved in the inversion. Although this approach may be more robust in the sense that redundant information is involved simultaneously in the inversion during the last stages of the multiscale approach, it may also be prohibitively expensive for 3D FWI. When global-offset recording is available, a starting model for FWI typically is built by first-arrival traveltime tomography (for recent applications at subsurface and crustal scales, see Ravaut et al., 2004; Operto et al., 2006; Brenders and Pratt, 2007a, 2007b), which provides large-scale velocity models whose resolution limit is on the order of the first Fresnel zone width (Williamson, 1991; Williamson and Worthington, 1993).

Application of FWI to 2D real data case studies has been limited to frequencies less than 20 Hz (Hicks and Pratt, 2001; Shipp and Singh, 2002; Ravaut et al., 2004; Operto et al., 2006). In three dimensions, the computational cost of the forward problem suggests that it is difficult to handle frequencies greater than 10 Hz for representative problems (Operto et al., 2007). At this scale, the resulting velocity models might be used as macromodels for PSDM. However, the relevance of FWI velocity models as reference models for PSDM requires further demonstration. An illustration with a real data case study is provided in Operto et al. (2004, 2005) in the frame of imaging a thrust belt in the southern Apennines by combined first-arrival traveltime tomography, FWI, and true-amplitude ray-based PSDM.

Full-wave-propagation modeling is a critical issue in FWI methods because it is the most computationally expensive task in the process. In the frequency domain, the forward problem reduces to resolving a large, sparse system of linear equations for each frequency considered. In two dimensions, the few frequencies involved in the inversion can be modeled efficiently for multiple shots using a direct solver (Marfurt, 1984). Since the original work of Pratt and Worthington (1990), optimal finite-difference stencils have been designed for the FD method based on direct solvers (Jo et al., 1996; Stekl and Pratt, 1998; Hustedt et al., 2004). The extension to three dimensions of this modeling approach is addressed in Operto et al. (2007), who show that problems of representative size (e.g., the 3D SEG/EAGE overthrust model) can be addressed at low frequencies (<10 Hz) on currently available distributed-memory platforms.

However, it remains unclear which approach (time versus FD based on direct or iterative solvers) is the most efficient for 3D FWI (Nihei and Li, 2007; Plessix, 2007). This may depend on several pa-

rameters related to the experimental setup, such as dimensions of the model, frequency bandwidth, number of traces in the acquisition, and acquisition geometry. Three-dimensional FD modeling methods based on iterative solvers are presented by Plessix (2007), Riyanti et al. (2007), and Warner et al. (2007). These approaches demand far less memory and have better scalability than approaches based on direct solvers, but their run time depends linearly on the number of sources, which may be a significant drawback with 3D surveys involving several thousands of shots or receivers. Alternatively, time-domain modeling methods were proposed to extract the frequency response of the wavefields by discrete Fourier transform (Sirgue et al., 2007b) or phase-sensitive detection (Nihei and Li, 2007) and subsequently invert in the frequency domain. The time-domain approach shares with the iterative FD approach similar advantages and drawbacks with respect to memory complexity, scalability, and computational burden for multishot simulations.

A few applications of 3D frequency-domain FWI to synthetic models are presented by Stekl et al. (2007) and Sirgue et al. (2007a). Sirgue et al. (2007a) apply frequency-domain FWI at low frequencies (3–5 Hz) to the 3D SEG/EAGE overthrust model. Their scheme is based on a finite-difference time-domain approach (Sirgue et al., 2007b). They illustrate the footprint of narrow- and wide-azimuth multichannel seismic-reflection acquisitions on the reconstructed velocity models and compare the convergence rate of 2D and 3D FWI. Stekl et al. (2007) apply 3D frequency-domain FWI to a channel model. Their scheme is based on an iterative solver (Warner et al., 2007). To overcome the computational burden of multishot simulations, several shots are triggered simultaneously, following an approach proposed by Capdeville (2005).

Our paper presents a massively parallel algorithm for a distributed-memory platform that performs 3D frequency-domain FWI using FD modeling based on a direct solver. We provide insight into the relevance of 3D frequency-domain FWI for building high-resolution velocity models of isotropic acoustic media and quantify the associated computational requirement thanks to realistic synthetic case studies. Numerical examples focus on surface wide-aperture/wide-azimuth surveys conducted with networks of sources and receivers on the surface. The most representative example, performed in a 12- × 12- × 4.5-km target of the overthrust model, uses a receiver spacing of 300 m, like the order used during a node survey in the Gulf of Mexico (426 m) (Clarke et al., 2007). The source-receiver patch was deployed above the 12- × 12-km area of the target. Therefore, the maximum far-inline and far-crossline offsets are 12 km for receivers located near the ends of the target area and 6 km for receivers located in the middle.

We do not address building a reliable starting model for FWI that is carried out conventionally by first-arrival traveltime tomography when global-offset acquisitions are considered. The maximum offset and the coarsest source and receiver spacings required to build a reliable large-scale model of the FWI target by first-arrival traveltime tomography must be clarified to assess the feasibility of this tomographic approach. We assume that a starting model describing the long wavelengths of the true medium is available. The second issue is to verify that the velocity models inferred from FWI can be used as a macromodel for 3D wave-equation PSDM.

First, we briefly review the theory of FD full-waveform modeling and inversion. Second, we discuss the parallel implementation of frequency-domain FWI for distributed-memory platforms. Third, we present several numerical examples of increasing complexity whose aim is to validate the algorithm, to illustrate the sensitivity of the imaging resolution to the acquisition geometry, and to provide some insight on the computational complexity of the approach for representative studies.

THEORY

Frequency-domain full-waveform modeling and inversion is a well-established method for imaging 2D media. Extension to the 3D case closely follows 2D strategies. Therefore, only a brief review of FD modeling and inversion is given here. The reader is referred to Operto et al. (2007) for the method used in this paper for FD wave-propagation modeling, to Pratt et al. (1998) for theoretical aspects of frequency-domain FWI, and to Pratt (1999) for practical aspects such as waveform-inversion data preprocessing and source estimation.

3D acoustic finite-difference FD modeling

The 3D viscoacoustic wave equation in the frequency domain is given by

$$\frac{\omega^2}{\kappa(x,y,z)}P(x,y,z,\omega) + \frac{\partial}{\partial x}\left(\frac{1}{\rho(x,y,z)}\frac{\partial P(x,y,z,\omega)}{\partial x}\right)$$
$$+ \frac{\partial}{\partial y}\left(\frac{1}{\rho(x,y,z)}\frac{\partial P(x,y,z,\omega)}{\partial y}\right)$$
$$+ \frac{\partial}{\partial z}\left(\frac{1}{\rho(x,y,z)}\frac{\partial P(x,y,z,\omega)}{\partial z}\right) = -S(x,y,z,\omega),$$

(1)

where $\rho(x,y,z)$ is density, $\kappa(x,y,z)$ is the complex bulk modulus, ω is frequency, $P(x,y,z,\omega)$ is the pressure field, and $S(x,y,z,\omega)$ is the source. Various attenuation models can be implemented easily in equation 1 using complex velocities in the expression of the bulk modulus (Toksöz and Johnston, 1981). Sponge-like perfectly matched layer (PML) boundary conditions can be implemented easily in the frequency domain to absorb outgoing energy (Berenger, 1994; Operto et al., 2007).

The relationship between the pressure wavefield and the source is linear, so the discrete acoustic wave equation 1 can be recast in a matrix form as

$$\mathbf{Ap} = \mathbf{s},$$

(2)

where the complex-valued impedance matrix $\mathbf{A}$ depends on the frequency and the medium's properties. System equation 1 can be discretized with the so-called parsimonious mixed-grid finite-difference method (Jo et al., 1996; Hustedt et al., 2004; Operto et al., 2007). The mixed-grid discretization, which uses multiple rotated coordinate systems, is complemented by a mass-term distribution (an antilumped mass) that significantly improves the accuracy of the stencil (Marfurt, 1984). The combined use of the mixed coordinate systems and mass distribution allows one to design both accurate and spatially compact stencils.

Dispersion analysis demonstrates that only four grid points per wavelength are needed to obtain accurate simulations in homogeneous media. This discretization rule is optimal for FWI, whose resolution limit is $\lambda/2$, where λ is the wavelength. A compact stencil is critical if a direct method is used to solve the system resulting from discretizing the Helmholtz equation because compact stencils limit the numerical bandwidth of the matrix and hence its fill-in during LU factorization. The use of a direct solver is interesting in the case of multishot simulations as required by tomographic applications be-

cause LU factorization is independent of the right-hand side terms in equation 2. To solve system 2, we use the MUMPS massively parallel direct solver, developed for distributed-memory platforms (Amestoy et al., 2007). A detailed complexity analysis of this approach is provided in Operto et al. (2007).

FD full-waveform inversion

The inverse problem is posed as a classic weighted least-squares optimization problem and is solved by the steepest-descent method (Tarantola, 1987). Newton and quasi-Newton (Gauss-Newton) methods were rejected because of the computer cost of calculating either the Hessian or approximate Hessian (Pratt et al., 1998).

The weighted least-squares objective function is given by

$$C(\mathbf{m}) = \Delta \mathbf{d}^\dagger \mathbf{W}_d \Delta \mathbf{d}, \tag{3}$$

where $\Delta \mathbf{d}$ is the misfit (the difference between the observed data and the data computed with model $\mathbf{m}$) and the superscript † indicates the adjoint (transpose conjugate). The value $\mathbf{W}_d$ is a weighting operator applied to the data; it scales the relative contribution of each component of the vector $\Delta \mathbf{d}$ in the inversion. Minimizing the objective function leads to the following solution for the model perturbation $\Delta \mathbf{m}$ after scaling and smoothing the gradient (Pratt et al., 1998; Ravaut et al., 2004; Operto et al., 2006):

$$\Delta m_i = -\alpha(\text{diag }\mathbf{H}_a + \varepsilon)^{-1}$$
$$\times G_\mathbf{m} \, \text{Re}\left\{\mathbf{p}^t\left[\frac{\partial \mathbf{A}^t}{\partial m_i}\right]\mathbf{A}^{-1}\mathbf{W}_d\Delta \mathbf{d}^*\right\}, \tag{4}$$

where diag $\mathbf{H}_a = \text{diag Re}\{\mathbf{J}^t\mathbf{W}_d\mathbf{J}^*\}$ denotes the diagonal elements of the weighted approximate Hessian $\mathbf{H}_a$, $\mathbf{J}$ denotes the sensitivity matrix, and $G_\mathbf{m}$ is a smoothing regularization operator.

One element of the sensitivity matrix is given by

$$\mathbf{J}_{k(m,n),i} = \mathbf{p}_m^t\left[\frac{\partial \mathbf{A}^t}{\partial m_i}\right]\mathbf{A}^{-1}\delta_n, \tag{5}$$

where $k(m,n)$ denotes a source-receiver couple of the acquisition geometry; m and n denote a shot and a receiver position, respectively; and δ_n is an impulse source located at receiver position n.

The diagonal of the approximate Hessian provides a preconditioner of the gradient that properly scales the perturbation model (Shin et al., 2001). The damping parameter ε is used to avoid numerical instabilities (i.e., division by zero). The matrix $G_\mathbf{m}$ is implemented in the form of a 3D Gaussian spatial filter whose correlation lengths are adapted to the inverted frequency component (Ravaut et al., 2004). Amplitude gain with offset can be applied to each seismic trace within the operator $\mathbf{W}_d$:

$$w_d(o_{SR}) = |o_{SR}|^g, \tag{6}$$

where the scalar g controls the amplitude of the gain with respect to the source-receiver offset o_{SR}. Originally, this operator was introduced to mitigate the contribution of the high-amplitude direct water wave when inverting long-offset ocean-bottom-seismic data (Operto et al., 2006). In our algorithm, the gradient scaling could be estimated once per frequency before the first iteration and kept constant over iterations or recomputed at each iteration. The term $(\partial \mathbf{A}/\partial m_i)$ is the radiation pattern of the diffraction by the model parameter m_i.

For P-wave velocity, the pattern is an explosion. In other words, the matrix whose number of rows corresponds to the number of diffractor points in the 3D finite-difference grid has only one nonzero element located on the diagonal of the ith row. This can be checked easily by noting that the P-wave velocity only appears in the coefficient $(\omega^2/\kappa(x,y,z))$ of acoustic wave equation 1. The finite-difference discretization without antilumped mass of this term leads to one nonzero coefficient per row on the diagonal of $\mathbf{A}$. Differentiating these diagonal coefficients with respect to the ith model parameter reduces to one nonzero coefficient on the ith row. The same conclusion would apply to the attenuation embedded in the expression of the complex velocity. On the other hand, a more complex radiation pattern would be observed for density whose expression appears in the stiffness matrix of the discrete wave equation (Forgues and Lambaré, 1997).

The source term in the FWI algorithm can be estimated by solving a linear inverse problem (Pratt, 1999). The inversion code can be applied to vertical geophone or hydrophone data generated by explosive sources. Indeed, vertical geophone data can be processed as pressure data thanks to the reciprocity principle (Operto et al., 2006). The inversion is applied in cascade to several groups of discrete frequencies. All frequencies of one group are inverted simultaneously. The final model obtained close to inversion of one group of frequencies is used as a starting model for the next group of frequencies. For each frequency group, several iterations can be computed.

PARALLEL NUMERICAL IMPLEMENTATION

We use the MUMPS massively parallel direct solver (Amestoy et al., 2006, 2007) based on a multifrontal method (Duff and Reid, 1983) to solve the forward problem (system 2). Before LU decomposition, the matrix coefficients are ordered to minimize dependencies in the graph of the matrix. Using nested dissection ordering, the theoretical memory complexity of the factorization for a 3D finite-difference problem is $O(n^4)$ and the number of floating-point operations is $O(n^6)$, where n is the number of grid points along one dimension of the 3D square finite-difference grid (Ashcraft and Liu, 1998). The source vectors for the resolution phase are provided in sparse format on the host processor. After resolution, the multiple solutions are distributed over processors following a domain decomposition driven by the distribution of the LU factors. This means that each processor stores a spatial subdomain of all the solutions. We take advantage of this distributed in-core storage of the forward-problem solutions (FPS) to solve the inverse problem in parallel.

The central component of the FWI algorithm is computing the gradient of the objective function. This operator is computed basically by a weighted summation of the FPS, namely, the incident and the backpropagated residual wavefields computed in the starting model (equation 4). The weights in the summation account for the radiation pattern of the diffraction tomography reconstruction (the operator $\partial \mathbf{A}/\partial m_i$ in equation 5) and for the data residuals. This weighted summation is computed in parallel straightforwardly by taking advantage of the distribution of the FPS: each processor computes the subdomain of the gradient corresponding to the subdomain of the FPS stored on this processor. At the end of the summation, the distributed gradient is gathered on the master processor with a collective communication.

When only the P-wave velocity parameter is involved in the inversion, the matrix $\partial \mathbf{A}/\partial m_i$ reduces to a scalar located on the ith diagonal. This implies that the gradient at position of m_i depends only on the values of the FPS at this same position. In that case, the parallel computation of the gradient does not require any point-to-point

communication, leading to a parallelism efficiency close to one for this task. Also, all FPS remain in core in the algorithm without disk swapping. If not enough memory is available to store in core all FPS values in addition to the LU factors, the FPS, the gradient, and the diagonal Hessian are computed in a sequential loop over partitions of the right-hand-side terms. Each partition loads in core the maximum number of solutions fitting the available memory. The efficiency of the parallel inversion algorithm is controlled mainly by that of the LU factorization. We obtain a maximum speed-up of 13 with MUMPS on our applications (Operto et al., 2007).

The parallel FWI algorithm is summarized in Figure 1. More details on the 2D version of the parallel FWI algorithm can be found in Sourbier et al. (2007).

SYNTHETIC EXAMPLES

In this section, we present several numerical examples of 3D FWI of increasing complexity to validate the algorithm and to give some insight on the computing cost of the approach on realistic cases. All examples were computed on an HP DL 145G2 Beowulf cluster at the SIGAMM computer center, located in the Observatoire de la Côte d'Azur (France). This parallel-distributed computer is a 48-node cluster, each node comprising two dual-core 2.4-GHz processors, giving 19.2 Gflops peak performance per node. The computer has a distributed-memory architecture, wherein each node has 8 GB of RAM. The interconnection network between processors is Infiniband 4X. Data are shared among processors using the MPIHP message-passing library. For the examples presented here, the PML layers spread along five grid points on each side and each direction.

Loop over groups of frequencies [ifreqgroup/Nfreqgroup]

  Loop over iterations [it/nitermax]
  Initialization of gradient, diagonal Hessian and cost function
  Read starting model

    Loop over frequencies in one group [ifreq/nfreq]
    Build impedance matrix
    Parallel factorization with MUMPS

    Diagonal Hessian computation (if it = 1)
      Build Nshot + Nreceiver RHSs on P0
      Parallel multi-RHS resolution with MUMPS
      Compute subdomains of diagonal Hessian on $P_{[i]}, i = 1, Nproc$

    Gradient computation
      Build Nshot RHS for shot positions on P0
      Parallel multi-RHS resolution with MUMPS
      Extract wavefields at receiver positions on $P_i, i = 1, Nproc$
      Compute data residuals and partial RMS on $P_i, i = 1, Nproc$
      Estimate source
      Build Nshot RHS for residual positions on P0
      Parallel multi-RHS resolution with MUMPS
      Compute subdomains of gradient on $P_{[i]}, i = 1, Nproc$
  End of loop over frequencies

  Reduce objective function on P0
  Centralize the gradient and the diagonal Hessian on P0
  Scale the gradient by the diagonal Hessian on P0

  Compute step length α
  Updated the velocity model on P0

  End of loop over iterations

End of loop over groups of frequencies

Figure 1. Outline of the FWI algorithm. Parallel tasks are written in gray. RHS — right-hand-side terms (i.e., sources in wave modeling); Nproc — number of MPI processes in the parallel run; nitermax — maximum number of iterations of one frequency-group inversion; P_i — processor i, where P0 is the master processor. An arbitrary number of frequencies N can be inverted simultaneously (set Nfreqgroup = 1 and nfreq = N) or successively (set Nfreqgroup = N and nfreq = 1).

Therefore, no free-surface multiples are considered in the examples. These PML grid points are not taken into account in the description of the finite-difference grids.

In the following examples, we use the inverse crime, which consists of computing the data with the modeling tool implemented in the FWI code. The whole wavefield, including refractions, turning waves, and reflections, is involved simultaneously in the inversion. All of the inversions were performed with unweighted data, i.e., using $g = 0$ in equation 6.

3D FWI in 2D configuration

In the first step, we validate the 3D FWI algorithm by comparing the results obtained using a 2D FWI code and the 3D code applied in a 2D configuration. Two-dimensional experiments can be designed considering 2.5D velocity models (laterally invariant in the y-direction) and an infinite line source in the y-direction. The infinite line source in the y-direction was implemented on a limited computational domain in the y-direction using periodic boundary conditions on the two faces of the model corresponding to $y = 0$ and $y = y_{max}$. The periodic boundary conditions are

$$\left[\frac{\partial P}{\partial y}\right]_{y = -h/2, y_{max}+h/2} = 0. \qquad (7)$$

They are applied on two virtual ghost faces located outside the computational domain at positions $y = -h/2$ and $y = y_{max} + (h/2)$, where h stands for the grid interval.

We applied 3D and 2D FWI to a dip section of the overthrust model (Aminzadeh et al., 1997) (Figure 2), discretized on an 801×187 grid with a grid spacing $h = 25$ m. For the 3D application, the dip section of the overthrust model was duplicated three times in the y-direction, leading to a 3D $801 \times 3 \times 187$ finite-difference grid. A 2D wavefield computed in this 2.5D model with the above-men-
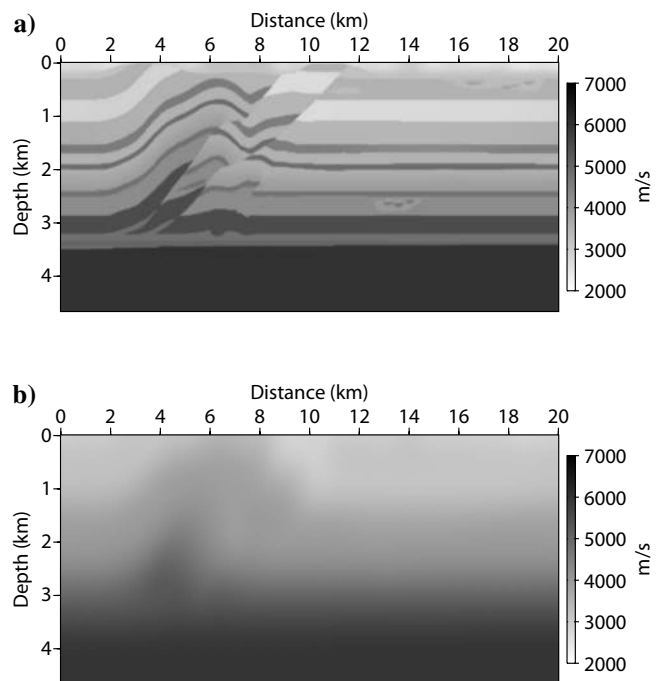

Figure 2. Imaging a dip section of the overthrust model. (a) True velocity model. (b) Starting velocity model.

tioned boundary conditions is shown in Figure 3. PML absorbing boundary conditions are set on the four edges of the 2D model.

The 2D acquisition geometry consists of a line of 200 sources and receivers, equally spaced on the surface. The corresponding 2.5D acquisition geometry consists of duplicating three times the source and receiver lines in the y-direction. The true model was augmented with a 250-m-thick layer on top of it; the sources and receivers were deepened accordingly to avoid having sources and receivers just below the PML-model interface. Velocities in this layer are vertically homogeneous and are equal to that on the surface of the original model. We observed strong instabilities during FWI in the near-surface velocities when this layer was not added to the model (Figure 1c in Ravaut et al., 2004). These instabilities can be removed by setting the true velocities in the first 100 m of the starting model without augmenting the model with an artificial layer (Operto et al., 2008).

The starting model for inversion is obtained by smoothing the true velocity model with a Gaussian function of horizontal and vertical correlation lengths of 500 m (Figure 2).
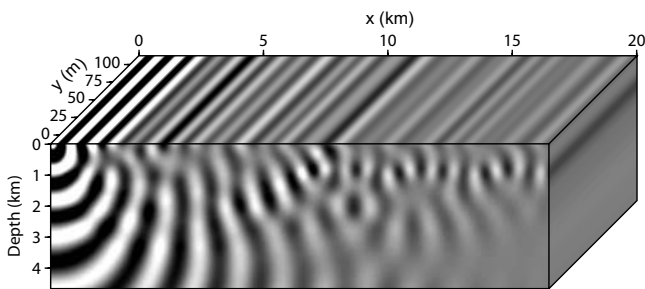
Figure 3. Example of a 2D wavefield computed in a 3D FD grid. Note the limited dimension of the grid in the y-dimension. Five grid points are used in the y-direction.
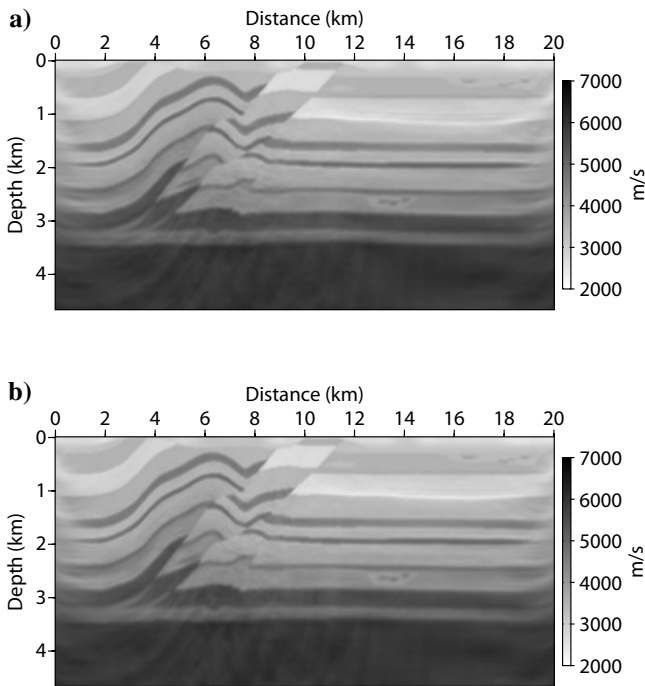
Figure 4. Imaging a dip section of the overthrust model where f = 20.63 Hz. (a) Final velocity model from 2D FWI. (b) Final velocity model from 3D FWI.

We inverted sequentially seven frequencies ranging from 5 to 20 Hz. For each frequency, we computed 40 iterations. The final velocity models inferred from 2D and 3D FWI are shown in Figure 4. Some vertical profiles extracted from these models are compared in Figure 5. They are very similar, providing a first validation of the 3D FWI algorithm. The agreement between the final FWI models and the true model is also quite good, although some discrepancies exist between the true and reconstructed velocities around a low-velocity layer located from 0.7 to 1 km deep (Figure 5b). This discrepancy is not observed when the first 100 m of the true model are set in the starting model (Operto et al., 2008). Some high-amplitude perturbations are still slightly underestimated, mainly because of an insufficient number of iterations. (The profiles in Figure 5 can be compared with those obtained from 2D FWI using 80 iterations per frequency in Operto et al. [2008].)

These results also give some insight on the high spatial resolution that can be achieved in the velocity models at relatively low frequencies (i.e., <15 Hz) by FWI of global-offset data thanks to continuously sampling the wavenumber spectrum up to a maximum wavenumber of $2/\lambda_{15\ Hz}$ m^{-1}. (We define the wavenumber as the inverse of the wavelength λ. The value $\lambda_{15\ Hz}$ is the wavelength corresponding to a frequency of 15 Hz.)

Inclusion models

In this section, we apply 3D FWI for simple velocity models composed of a homogeneous background with one and two inclusions. The models are discretized on a small $31 \times 31 \times 31$ grid with 50-m cubic cells. The velocity in the background medium is 4000 m/s. The inverted frequencies are 3.72, 6.07, 10.00, and 16.27 Hz for the one- and two-inclusion models. To select the inverted frequencies, we chose to remove the wavenumber redundancy in the model space (Sirgue and Pratt, 2004). The frequencies were inferred from the relationship between wavenumber, frequency, and aperture angle provided by the theory of diffraction tomography (Wu and Töksoz, 1987). The linear relation between wavenumber and frequency allows us to increase the frequency interval when the frequency increases (Sirgue and Pratt, 2004). One hundred sources (10×10) and
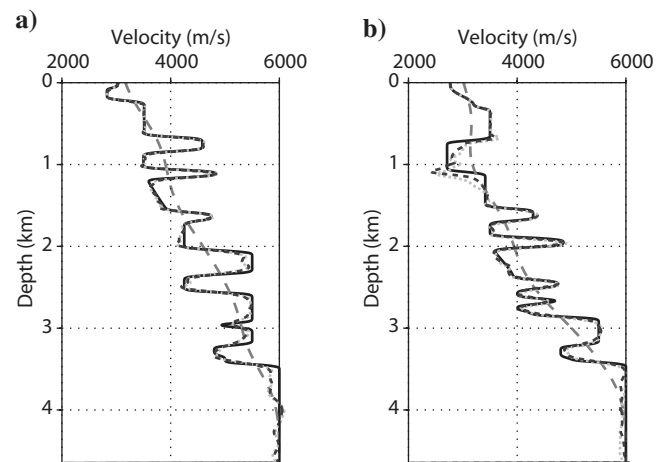

Figure 5. Imaging a dip section of the overthrust model: Comparison between vertical profiles extracted from the true (black solid line), the starting (gray dashed line), and the 2D and 3D FWI models (light and dark gray dotted lines, respectively). The two profile series are located at (a) 4.5 and (b) 13.5 km distance. The 2D and 3D FWI profiles are almost identical.

36 receivers (6 × 6) were distributed uniformly on the top and bottom of the 3D model, respectively. Source and receiver spacings were 150 and 250 m, respectively.

We first consider a velocity model with one inclusion in the homogeneous background. The velocity in the inclusion is 3500 m/s (Figure 6). The inclusion is centered on the 3D grid. The four frequencies were inverted successively. Some horizontal and vertical sections of the inclusion are shown in Figure 6. Note the vertically elongated shape of the inclusion in the vertical cross section and the symmetric
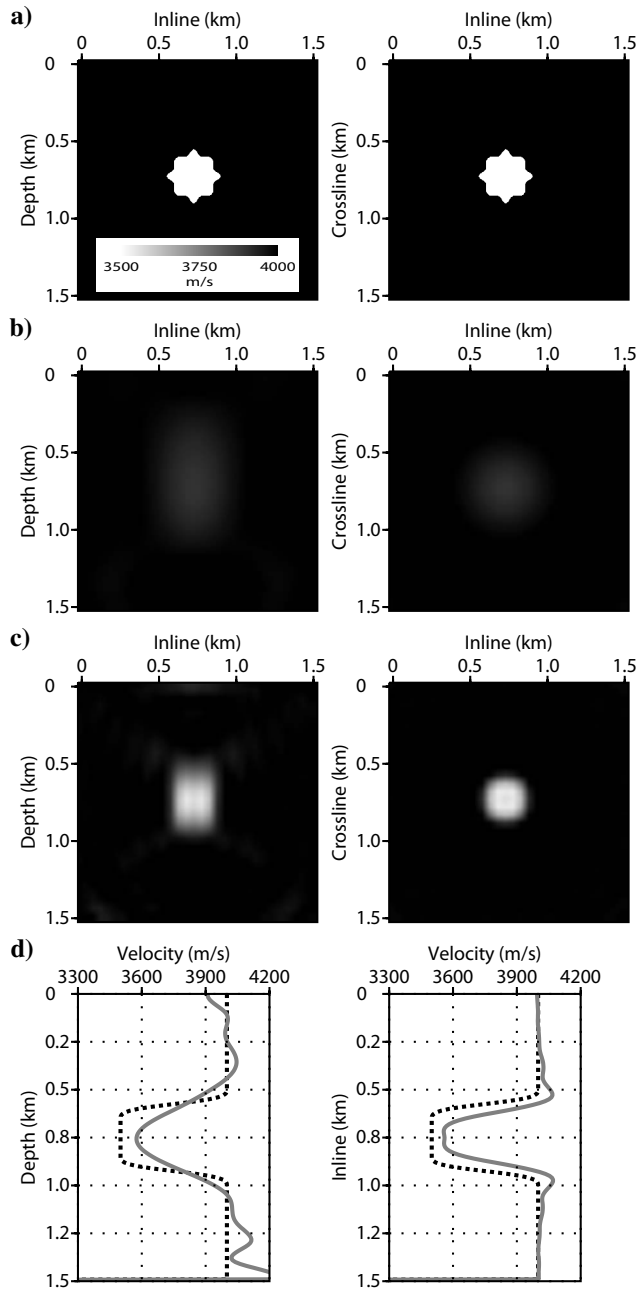
shape of the inclusion in the horizontal slice. The vertical elongation results because the top and bottom of the inclusion are sampled mainly by downgoing transmitted wavepaths (i.e., forward-scattered wavepaths) that have limited resolution power, but the shape of the inclusion in a horizontal plane is mainly controlled by reflections (i.e., backward-scattered wavepaths) associated with shots and receivers located near the same face of the 3D model.

This relationship between aperture illumination and image resolution is also illustrated on the two profiles extracted from a vertical and horizontal section running through the inclusion. The vertical profile exhibits a clear deficit of high wavenumbers as a result of transmission-like reconstruction, but the horizontal profile exhibits only a slight deficit of small wavenumbers as a result of reflection-like reconstruction. The symmetry of the image of the inclusion in the horizontal plane, which results from the symmetry of the inclusion with respect to the acquisition geometry, is additional validation of the 3D FWI algorithm.

The second example contains two spherical inclusions (3500 and 4500 m/s) corresponding to positive and negative perturbations in the homogeneous background (Figure 7a). The center of the inclusions lies on the same vertical plane in the middle of the grid. The

Figure 6. Imaging one inclusion by 3D FWI. Vertical (left) and horizontal (right) sections of (a) the true inclusion, (b) the inclusion after inverting the 3.72-Hz frequency, and (c) the inclusion after inverting the four frequencies successively. (d) Vertical (left) and horizontal (right) profiles extracted from models shown in (a) (dotted lines) and (c) (solid lines).
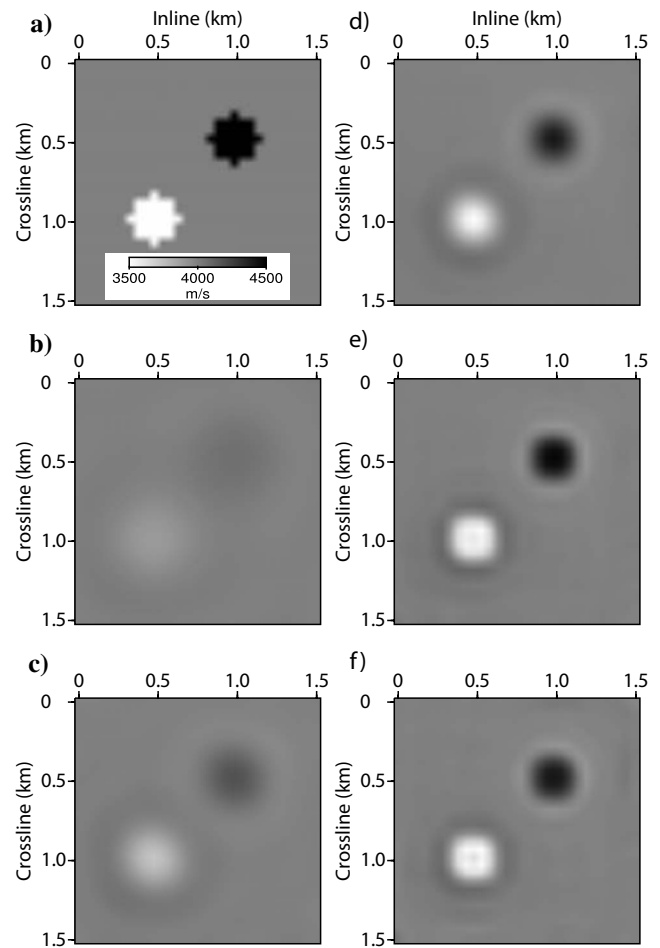

Figure 7. Imaging two inclusions by 3D FWI. (a) True model. Also included are vertical cross sections of the FWI velocity models after successive inversion of frequencies of (b) 3.72, (c) 6.07, (d) 10.00, and (e) 16.27 Hz. (f) Vertical cross section of the 3D FWI velocity model after simultaneous inversion of the four frequencies: 3.72, 6.07, 10.00, and 16.27 Hz.
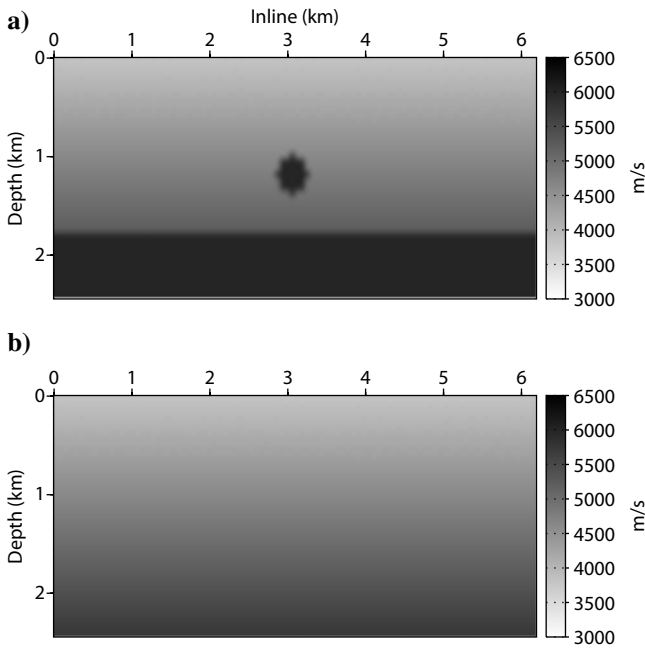
Figure 8. Imaging the inclusion/interface model. (a) True velocity model. (b) Starting velocity model for FWI.
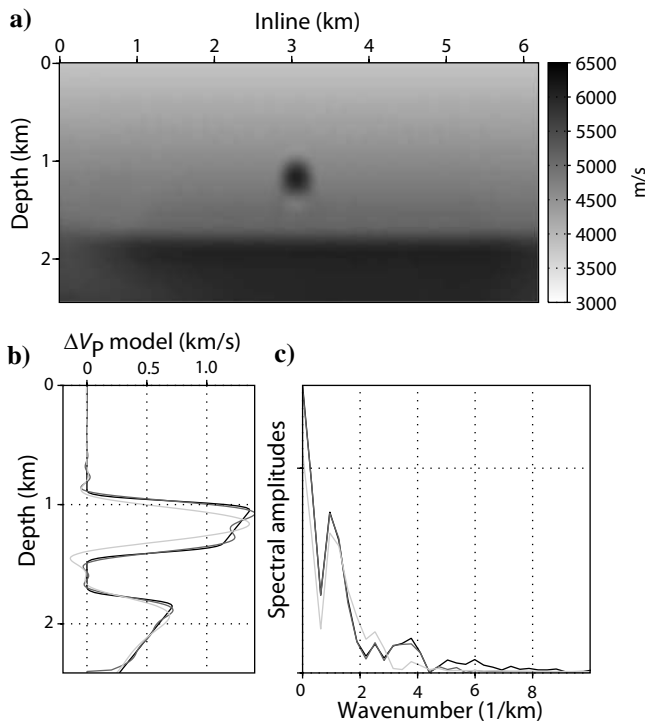

Figure 9. Imaging the inclusion/interface model. (a) Final FWI velocity model. (b) Vertical profile across the inclusion extracted from the true perturbation velocity model (black), from the true perturbation velocity model after low-pass filtering at the theoretical resolution of 12.15-Hz FWI (gray), and from the final FWI perturbation velocity model. (c) Spectral amplitudes as a function of wavenumber of the profiles shown in (b). These spectra suggest that wavenumbers between 0.003 and 0.004 m^{-1} corresponding to wavelengths of 333 and 250 m, respectively, were unsuccessfully imaged.

goal of this test is to verify that the 3D inversion properly handles multiple scattering occurring between the two inclusions. For this case study, the four frequencies were inverted successively and simultaneously (Figure 7b-f). In both cases, the inversion successfully imaged the two inclusions.

Inclusion + interface velocity model

A more realistic example consists of a velocity-gradient layer above a homogeneous layer. A high-velocity inclusion corresponding to a velocity perturbation of >1 km/s was incorporated into the velocity-gradient layer (Figure 8). The minimum and maximum velocities were 3.8 and 6.0 km/s, respectively. It was discretized on a $100 \times 100 \times 40$ grid with a grid spacing of $h = 62.5$ m, which corresponds to a physical domain of $6.25 \times 6.25 \times 2.5$ km. The grid spacing h was kept constant over the successive monofrequency inversions and was set according to the maximum inverted frequency.

The starting model for inversion was the velocity-gradient layer extended down to the bottom of the model (Figure 8). The acquisition geometry consisted of two regular grids of $17 \times 17 = 289$ sources and receivers deployed on the surface. The receiver grid was shifted with respect to the shot grid, such that each receiver was midway between four adjacent shots. This source-receiver configuration was chosen to avoid recording high-amplitude zero-offset data, which degrade the inversion conditioning. The distance between either two sources or receivers was 312.5 m. We sequentially inverted five frequencies, ranging from 1.76 to 12.15 Hz, and computed 20 iterations per frequency.

The final FWI velocity model is shown in Figure 9a. A vertical profile across the inclusion extracted from the final FWI perturbation model (i.e., the difference between the final FWI model and the starting model) is shown in Figure 9b. It is compared with that extracted from the true perturbation model after low-pass filtering at the theoretical resolution of FWI at 12 Hz. The bottom layer is well recovered, thanks to the large offset coverage, allowing it to image a broad range of the layer wavelengths quantitatively. The shape of the inclusion is recovered incompletely with respect to the expected resolution of the imaging at 12 Hz, although the velocity amplitude in the inclusion is recovered fully (Figure 9b and c).

The spectra of the two profiles reveal that the amplitudes of the low wavenumbers were recovered incompletely. This is probably from an insufficient number of iterations, as suggested by the plot of the objective function as a function of iteration number (Figure 10). Moreover, some discrepancies in the shape of the two spectra of Figure 9c for wavenumbers greater than 0.0017 m^{-1} suggest that the inversion may have converged toward a local minimum. This may have been caused by the high amplitudes of the model perturbations and the related complex interactions between waves multiscattered between the bottom of the inclusion and the top of the layer, which make the inverse problem more nonlinear. We speculate that simultaneous inversion of multiple frequencies following the multiscale approach of Bunks et al. (1995) may help to manage this nonlinearity.

Qualitative inspection of the vertical profiles also reveals a slight deficit of small (vertical) wavenumbers in the image of the inclusion and of the bottom layer (suggested by the negative velocity perturbations with respect to the true model). This deficit is again explained by the surface acquisition geometry, which illuminates the vertical components of the wavenumber vector with reflections only.

SEG/EAGE overthrust model

The 3D SEG/EAGE overthrust model is a constant-density acoustic model dimensions are $20 \times 20 \times 4.65$ km (Aminzadeh et al., 1997). It is discretized with 25-m³ cells, representing a uniform mesh of $801 \times 801 \times 187$ nodes. The minimum and maximum velocities in the overthrust model are 2.2 and 6.0 km/s, respectively (Figure 11).

Overthrust model: Channel target

Because of limited computer resources, our first application was restricted to a small section of the overthrust model centered on a channel. The maximum frequency involved in the inversion was 15 Hz. A horizontal slice and a vertical section of the model are shown in Figure 12. The model dimensions are $7 \times 8.75 \times 2.25$ km and is discretized with a grid spacing $h = 50$ m, leading to a $141 \times 176 \times 46$ grid. The minimum and maximum velocities are 3.3 and 6.0 km/s, respectively. The acquisition geometry consists of two regular grids of 44×33 sources and receivers on the surface corresponding to 1452 sources and receivers. The distance between two sources or two receivers is 200 m. The receiver array is shifted according to the source array, following the geometry in the previous example.

We sequentially inverted five frequencies ranging from 5 to 15 Hz. For each frequency, we computed seven iterations. The starting velocity model was obtained by smoothing the true model with a wavenumber filter having a cut-off wavenumber of $1/500$ m^{-1} (Figure 13). The final FWI model provides a low-pass
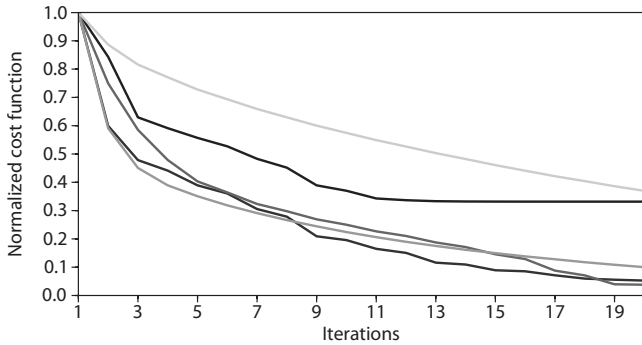

Figure 10. Imaging the inclusion/interface model: cost function versus iteration number for each inverted frequency (dark to light gray for increasing frequency). Convergence was not achieved, which partially explains underestimated velocities.
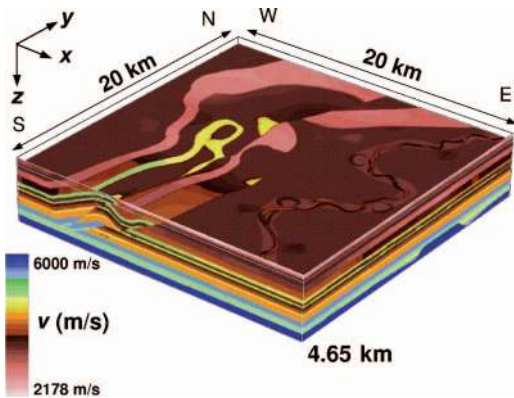

Figure 11. The 3D SEG/EAGE overthrust model (Aminzadeh et al., 1997).

version of the true model (Figure 14). To assess the accuracy of the FWI, we low-pass-filtered the true model in the time domain with a cut-off frequency of 15 Hz to mimic the exact velocity model that would have been inferred by FWI (Figure 15). Qualitative comparison between the final FWI velocity model and the low-pass-filtered true model shows good agreement between the two models.

Figure 16 compares a vertical profile extracted from the starting model, the low-pass true velocity model, and the final FWI model. The agreement is reasonably good with, again, a slight deficit of
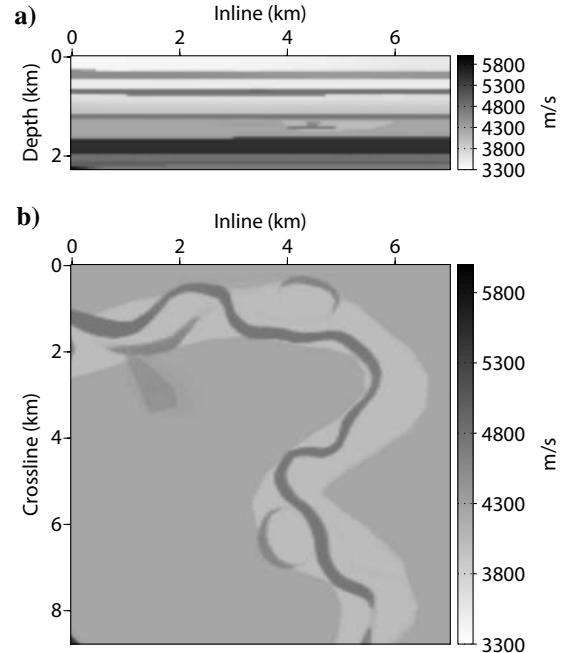

Figure 12. Imaging a channel in the overthrust model, true velocity model. (a) Cross section at $x = 4$ km. (b) Horizontal slice at $z = 1.5$ km.
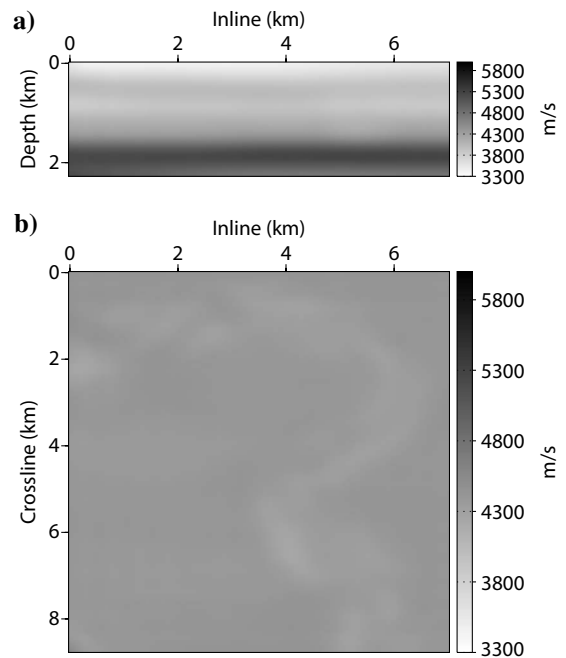

Figure 13. Imaging a channel in the overthrust model, starting velocity model. (a) Cross section at $x = 4$ km. (b) Horizontal slice at $z = 1.5$ km.

small wavenumbers in the FWI profile because of the surface-to-surface illumination. We also note an underestimation of velocities in the deep part of the model (see the high-velocity layer above 2 km in depth). This may result from an insufficient number of iterations. Indeed, the deep structures are mainly constrained by later-arriving re-
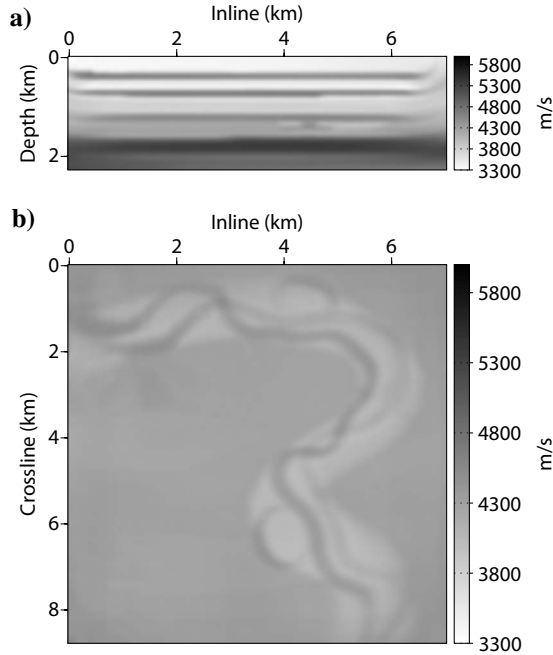

Figure 14. Imaging a channel in the overthrust model, FWI velocity model after successive inversion of the five frequencies, f = 14.9 Hz. (a) Cross section at x = 4 km. (b) Horizontal slice at z = 1.5 km.
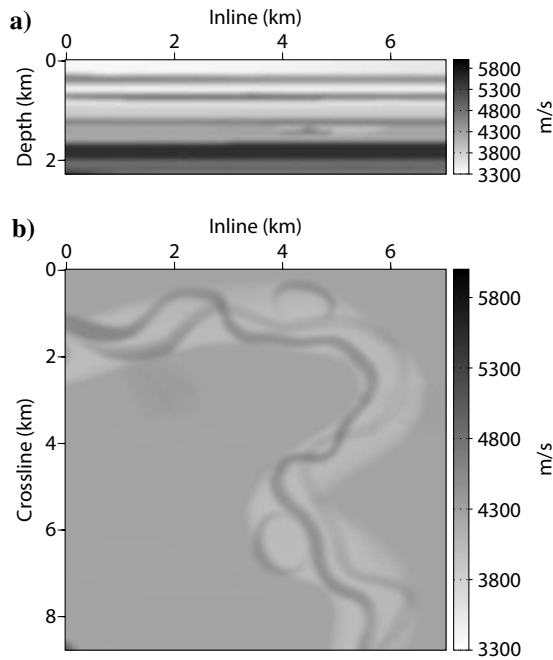

Figure 15. Imaging a channel in the overthrust model, low-pass filtered true velocity model (~15 Hz). (a) Cross section at x = 4 km. (b) Horizontal slice at z = 1.5 km.

flections of smaller amplitude recorded at larger offsets. Misfit reduction may be slower for these arrivals because the value of the objective function is dominated by the residuals of the high-amplitude shallow arrivals during the first iterations.

The weighting operator in the data space corresponding to an amplitude gain with offset (equation 6) may accelerate the reduction of the long-offset residuals at the partial expense of the short-offset ones during late iterations. However, this strategy requires that we ensure short-offset residuals are reduced enough to avoid propagating errors associated with inaccurate shallow structures deeper in the model. This detailed tuning of 3D FWI requires further investigation.

To perform this application, we used 60 MPI processes distributed over 15 dual-core biprocessor nodes. Each MPI process used 1.5 GB of RAM (Table 1). Seven iterations of the inversion of one frequency
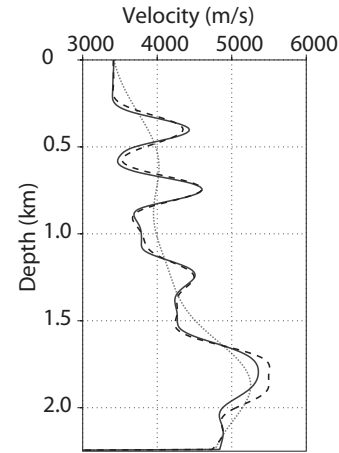

Figure 16. Imaging a channel in the overthrust model — comparison between vertical profiles extracted from the starting model (dotted line), the low-pass filtered true model (dashed line), and the final FWI velocity model (solid line). The profile is located at (x = 3.5 km, y = 4.4 km).

Table 1. Computational cost of imaging the overthrust model (channel system).

Requirement term	Value
MEM$_{\text{FACTO ALL}}$ (GB)	67.0
MEM$_{\text{FACTO PROC}}$ (GB)	1.5
TIME$_{\text{FACTO}}$ (s)	510.0
TIME$_{\text{SOLVE ALL}}$ (s)	1270.0
TIME$_{\text{SOLVE SOURCE}}$ (s)	0.9
TIME$_{\text{GRADIENT}}$ (s)	4.0
TIME$_{\text{diag HESSIAN a}}$ (s)	3093.0
TIME$_{\text{ITERATION}}$ (s)	18,865.0

MEM$_{\text{FACTO ALL}}$ — Total memory allocated during factorization
MEM$_{\text{FACTO PROC}}$ — Average allocated memory per working processor during factorization
TIME$_{\text{FACTO}}$ — Elapsed time for factorization
TIME$_{\text{SOLVE ALL}}$ — Total elapsed time for multishot resolution
TIME$_{\text{SOLVE SOURCE}}$ — Elapsed time for resolution for one source
TIME$_{\text{GRADIENT}}$ — Elapsed time to compute gradient
TIME$_{\text{diag HESSIAN a}}$ — Elapsed time to compute diagonal Hessian
TIME$_{\text{ITERATION}}$ — Average elapsed time to process one iteration

took approximately 45 hours. Table 1 lists information related to run time and memory requirement for LU factorization, multishot resolutions (both tasks being devoted to the forward problem), gradient, and diagonal Hessian computation. Run time for the solution phase is very small (0.9 s per source) and illustrates the main advantage of FD modeling methods based on direct solvers for tomographic applications involving a few thousand sources. Computation of the gradient is also negligible in the frequency domain (4 s) as a result of summation without disk swapping over a very compact volume of data limited to few frequency components. Increasing the number of cores in the inversion would have reduced computational time significantly but would have increased the memory requirement because of overheads during parallel LU factorization (Operto et al., 2007).

Overthrust model: Thrust target

We now consider imaging a significant target of the overthrust model that incorporates the main thrusts of the model (Figure 17). The minimum and maximum velocities are 2.2 and 6.0 km/s, respectively. The model dimensions are $13.425 \times 13.425 \times 4.65$ km.

The acquisition geometry consists of two coincident $43 \times 43 = 1849$ grids of sources and receivers deployed on the surface. The distance between two sources or two receivers is 300 m. A receiver spacing of 300 m is representative of the spacing between two adjacent nodes in a dense 3D wide-azimuth node survey (Clarke et al., 2007). We used the same spacing between two adjacent shots and receivers, although a more representative shot survey could have been designed by using smaller shot and line intervals in the dip-line and crossline directions, respectively. The increased number of shots should not dramatically increase the computational time needed to compute the gradient because the residuals recorded at the shot positions (by virtue of shot-receiver reciprocity) can be propagated at one time for each receiver.

The extra computational cost caused by a denser shot survey would have resulted from building and storing denser residual sources ($\Delta \mathbf{d}*$ in equation 4) and from the more expensive backward/forward substitutions solving $\mathbf{A}^{-1}\mathbf{W}_d\Delta \mathbf{d}*$. On the contrary, the CPU time required to compute the diagonal Hessian would increase dramatically because it requires a forward simulation for nonredundant shot and receiver positions. However, a good approximation of the diagonal Hessian can be computed on a coarser shot grid with a shot interval of the same order as the receiver's (Operto et al., 2006).

We sequentially inverted three frequencies — 3.5, 5, and 7 Hz — and computed 10 iterations per frequency. For this application, we adapted the grid interval to the inverted frequency. Grid intervals were $h = 150, 100$, and 75 m for frequencies of 3.5, 5, and 7 Hz, respectively. The source and receiver positions were chosen to coincide with the position of the nodes of the FD grids associated with the three inverted frequencies. This allowed us to bypass the problem of accurate implementation of point sources in a coarse FD grid, a critical issue in 3D frequency-domain FWI when the grid interval is adapted to the frequency (Hicks, 2002). These discretizations led to grids measuring $90 \times 90 \times 32$, $135 \times 135 \times 47$, and $180 \times 180 \times 63$.

The starting velocity model was obtained by smoothing the true model with a 3D Gaussian function having a correlation length of 1000 m in the three directions (Figure 18).
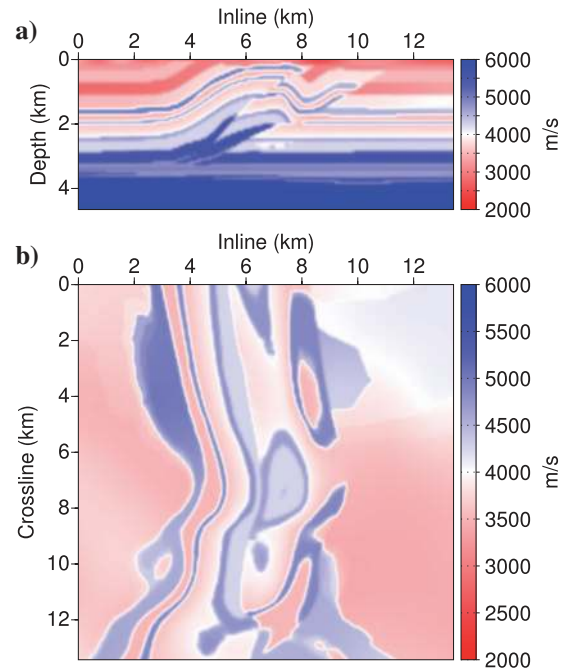

Figure 17. Imaging the thrust system in the overthrust model, true velocity model. (a) Cross section at $x = 3.3$ km. (b) Horizontal slice at $z = 1.5$ km.
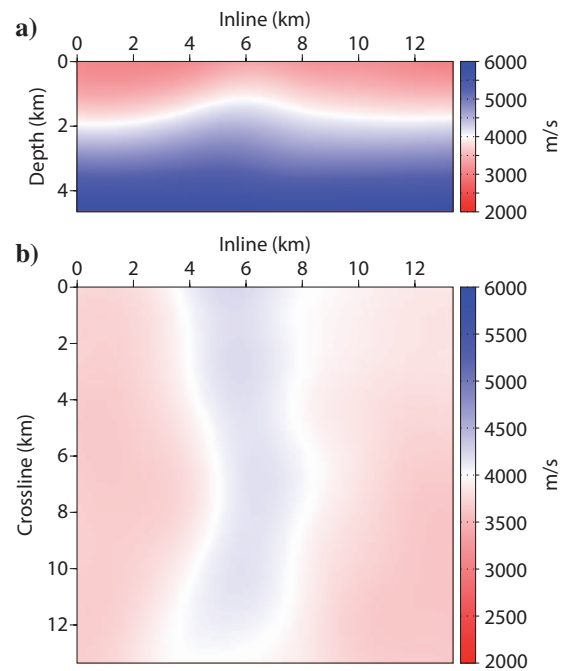

Figure 18. Imaging the thrust system in the overthrust model, starting velocity model. (a) Cross section at $x = 3.3$ km. (b) Horizontal slice at $z = 1.5$ km.
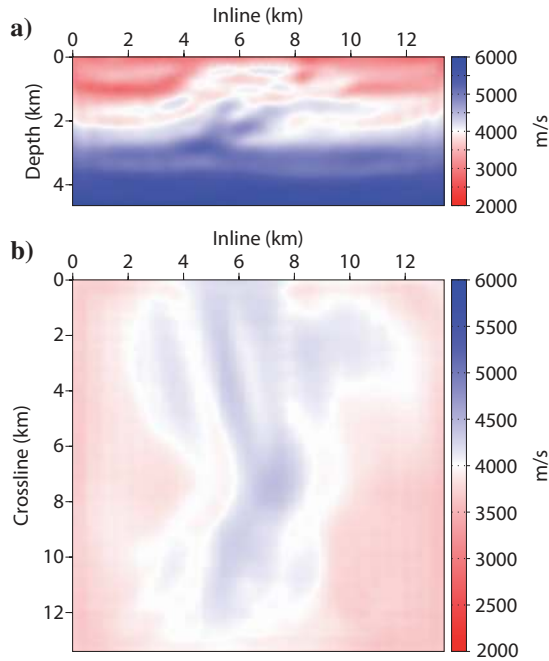
Figure 19. Imaging the thrust system in the overthrust model, 3.5-Hz FWI velocity model. (a) Cross section at $x = 3.3$ km. (b) Horizontal slice at $z = 1.5$ km.

The FWI velocity models after inversion of the 3.5-, 5-, and 7-Hz frequencies are shown in Figures 19–21. One can note a square pattern superimposed on the horizontal and vertical slices of the FWI velocity models obtained after inverting frequencies of 5 and 7 Hz (Figures 20 and 21). The size of the square matches the shot and receiver spacing, suggesting that it corresponds to the footprint of the coarse acquisition geometry. This footprint increases from 5 to 7 Hz as model resolution increases. The acquisition footprint has no preferential orientation because both shots and receivers are deployed uniformly all over the surface, with a constant spacing in the dip and cross directions.

Another illustration of the footprint of acquisition coarseness on 3D frequency-domain FWI is illustrated by Sirgue et al. (2007a). For this application, we ran 40 MPI processes distributed over 10 dual-core biprocessor nodes (four MPI processes/node) for the 150-m grid (frequency = 3.5 Hz), 60 processes distributed over 20 dual-core biprocessor nodes (three MPI processes/node) for the 100-m grid (frequency = 5 Hz), and 90 processes distributed over 30 dual-core biprocessor nodes (three MPI processes/node) for the 75-m grid (frequency = 7 Hz). The number of processes per dual-core biprocessor node decreased from four to three as the size of the problem increased in order to increase the amount of shared memory assigned to each processor for large problems. This process optimized memory use at a partial expense of run time because memory overhead decreased with number of processes. The 10 iterations took about 24, 72, and 120 hours for the 3.5-, 5-, and 7-Hz frequencies, respectively. More detailed information is shown in Table 2.
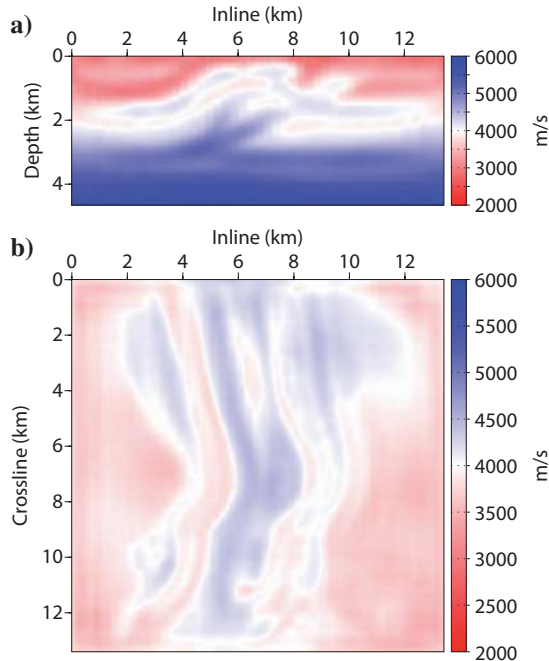

Figure 20. Imaging the thrust system in the overthrust model, 5-Hz FWI velocity model. (a) Cross section at $x = 3.3$ km. (b) Horizontal slice at $z = 1.5$ km.
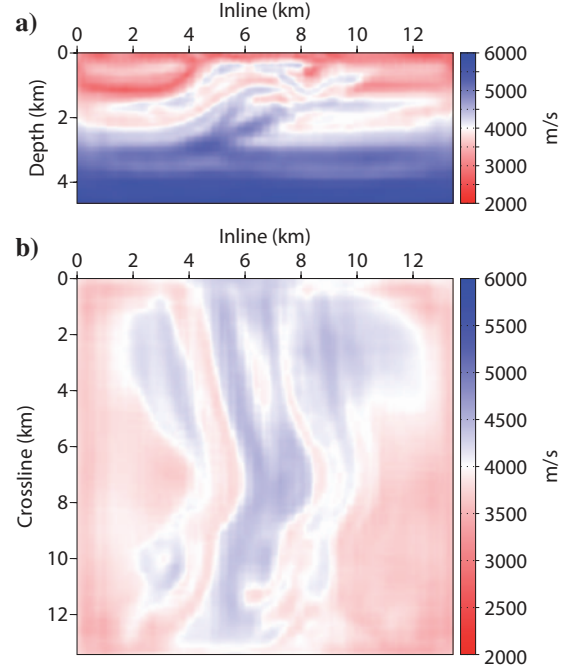

Figure 21. Imaging the thrust system in the overthrust model, 7-Hz FWI velocity model. (a) Cross section at $x = 3.3$ km. (b) Horizontal slice at $z = 1.5$ km.

We also performed the simulations using multithreading. We used 20, 40, and 60 processes distributed over 10, 20, and 30 dual-core bi-processor nodes (i.e., one process per processor) for the 3.5-, 5-, and 7-Hz inversions, respectively, and two threads per processor in each case (i.e., one thread per core). Fewer processes were used in the multithread configuration, so less distributed memory was allocated during LU factorization as a result of reduced memory overhead (Table 3). Moreover, CPU time was reduced significantly for the LU-factorization and multiple-shot solution phases (Table 3). These results suggest that nodes with a large amount of shared memory and multiple cores provide the optimal architecture to perform FD wave modeling based on a direct solver to take advantage of multithreading while mitigating the memory requirement.

Figure 22 shows the convergence rate for each processed frequency. This convergence rate can be compared with that shown in Figure 7a of Sirgue et al. (2007a) for 3.5-Hz frequency, keeping in mind that the whole overthrust model is imaged in Sirgue et al. (2007a).

In Figures 23 and 24, the data fit is illustrated in the frequency domain for the 3.5-, 5-, and 7-Hz frequencies. We compare the spectral amplitude and phase of the monochromatic wavefields at the receiv-er positions computed in the true velocity model and in the FWI models at the first and last iterations of the three monofrequency inversions. One shot is located at the upper-left corner of the receiver plane (Figure 23); the second shot is in the middle of the receiver array (Figure 24). The misfit reduction between the first and last iterations is more obvious at 3.5 Hz, illustrating slower convergence as frequency increases (Figure 22). We also note this misfit reduction is more effective for the shot located in the middle of the receiver array, again illustrating the difficulty of matching the lower-amplitude arrivals recorded at larger offsets. As for the channel case study, the objective function is less sensitive to the residuals of these low-amplitude arrivals; more iterations would have been required to cancel them.

Table 2. Computational cost of imaging the overthrust model (thrust system). The memory available per MPI process is 2, 2.7, and 2.7 GB for frequencies of 3.5, 5, and 7 Hz, respectively.

Requirement term	At 3.5 Hz frequency	At 5.0 Hz frequency	At 7.0 Hz frequency
N_P	40	60	90
GRID	$90 \times 90 \times 32$	$135 \times 135 \times 47$	$180 \times 180 \times 63$
$MEM_{FACTO\ ALL}$ (GB)	11.5	45.0	124.0
$MEM_{FACTO\ PROC}$ (GB)	0.3	0.8	1.8
$TIME_{FACTO}$ (s)	72.0	340.0	1850.0
$TIME_{SOLVE\ ALL}$ (s)	310.0	990.0	3450.0
$TIME_{SOLVE\ SOURCE}$ (s)	0.2	0.6	1.8
$TIME_{GRADIENT}$ (s)	0.7	1.5	35.0
$TIME_{diag\ HESSIAN\ a}$ (s)	1999.0	3432.0	4000.0
$TIME_{ITERATION}$ (s)	2940.0	13,650.0	44,870.0

N_P — Number of MPI processes
GRID — Dimension of the 3D FD grid
$MEM_{FACTO\ ALL}$ — Total memory allocated during factorization
$MEM_{FACTO\ PROC}$ — Average allocated memory per working processor during factorization
$TIME_{FACTO}$ — Elapsed time for factorization
$TIME_{SOLVE\ ALL}$ — Total elapsed time for multi-shot resolution
$TIME_{SOLVE\ SOURCE}$ — Elapsed time for resolution of one source
$TIME_{GRADIENT}$ — Elapsed time to compute gradient
$TIME_{diag\ HESSIAN\ a}$ — Elapsed time to compute diagonal Hessian
$TIME_{ITERATION}$ — Average elapsed time to MPI process one iteration

Table 3. Computational cost of imaging the overthrust model (thrust system) using a multithread configuration (two threads per MPI process). The memory available per MPI process is 4 GB for the three frequencies. Compare these values with those of Table 2.

Requirement term	At 3.5 Hz frequency	At 5.0 Hz frequency	At 7.0 Hz frequency
N_P	20	40	60
$MEM_{FACTO\ ALL}$ (GB)	9.9	40.0	108.0
$MEM_{FACTO\ PROC}$ (GB)	0.5	1.0	1.4
$TIME_{FACTO}$ (s)	56.0	222.0	650.0
$TIME_{SOLVE\ ALL}$ (s)	290.0	650.0	1375.0

N_P — Number of MPI processes
$MEM_{FACTO\ ALL}$ — Total memory allocated during factorization
$MEM_{FACTO\ PROC}$ — Average allocated memory per working processor during factorization
$TIME_{FACTO}$ — Elapsed time for factorization
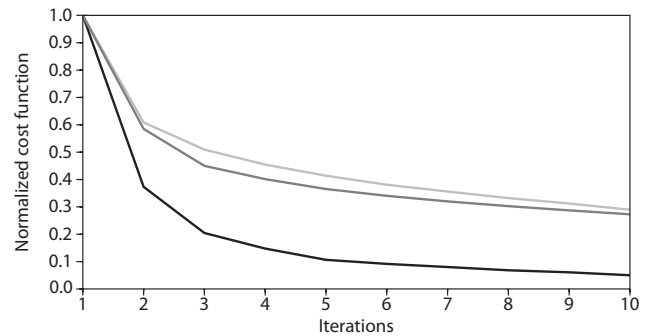$TIME_{SOLVE\ ALL}$ — Total elapsed time for multishot resolution

Figure 22. Imaging the thrust system in the overthrust model: objective function versus iteration number for the 3.5-, 5-, and 7-Hz frequencies (dark to light gray for increasing frequency).
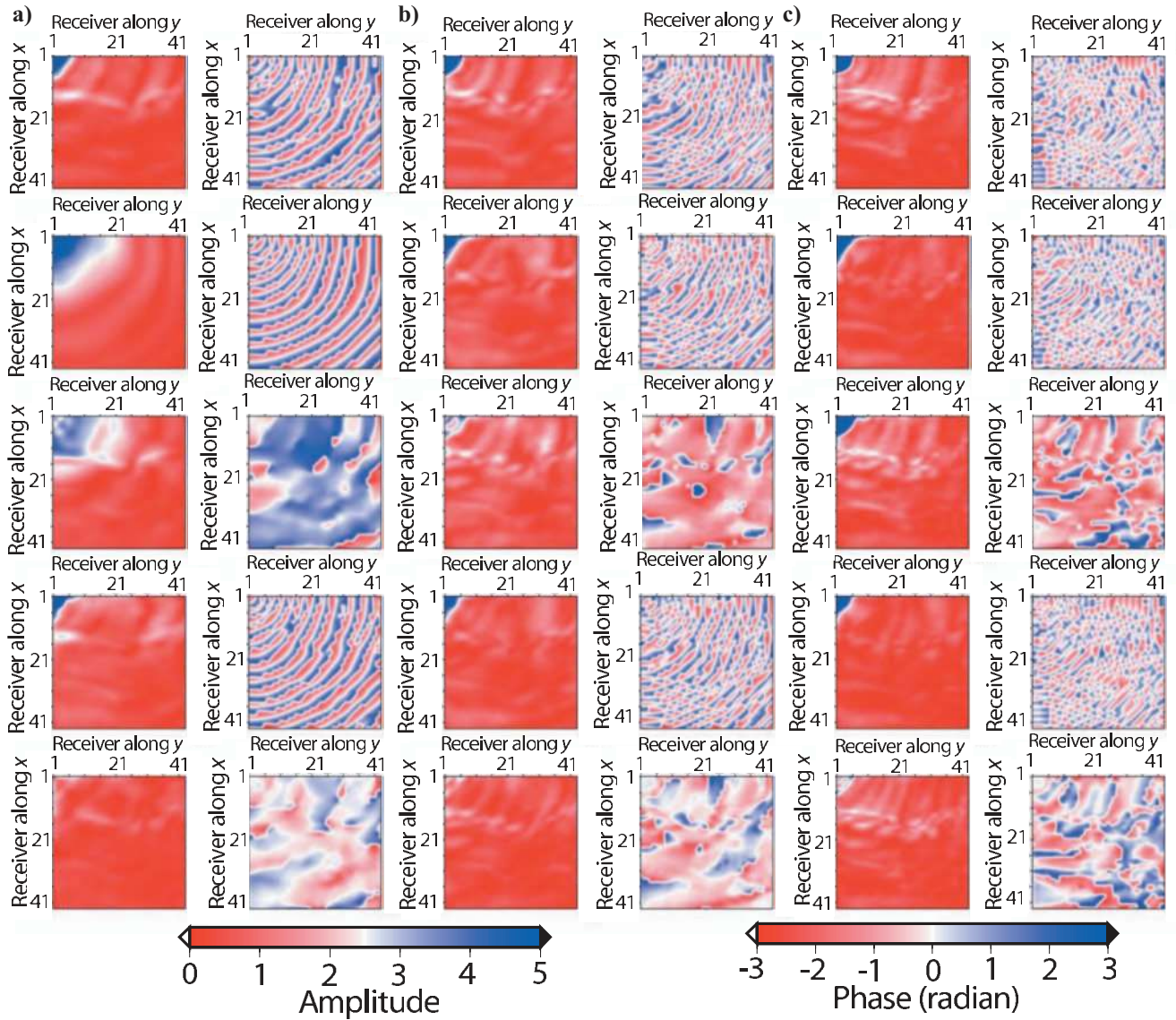
Figure 23. Imaging the thrust system in the overthrust model. Amplitude (left) and phase (right) of the (a) 3.5-Hz, (b) 5-Hz, and (c) 7-Hz wavefields at the receiver positions. The horizontal and vertical axes label the receiver number in the dip-line and crossline directions, respectively. The source is located in the upper-left corner. From top to bottom, the first and second panels show the wavefield computed in the true and the starting models, respectively; the third is the difference between the two wavefields. The fourth panel shows the wavefield computed in the final model of the frequency inversion, and the fifth is the residual between the wavefields computed in the true model and in the final model of the frequency inversion.
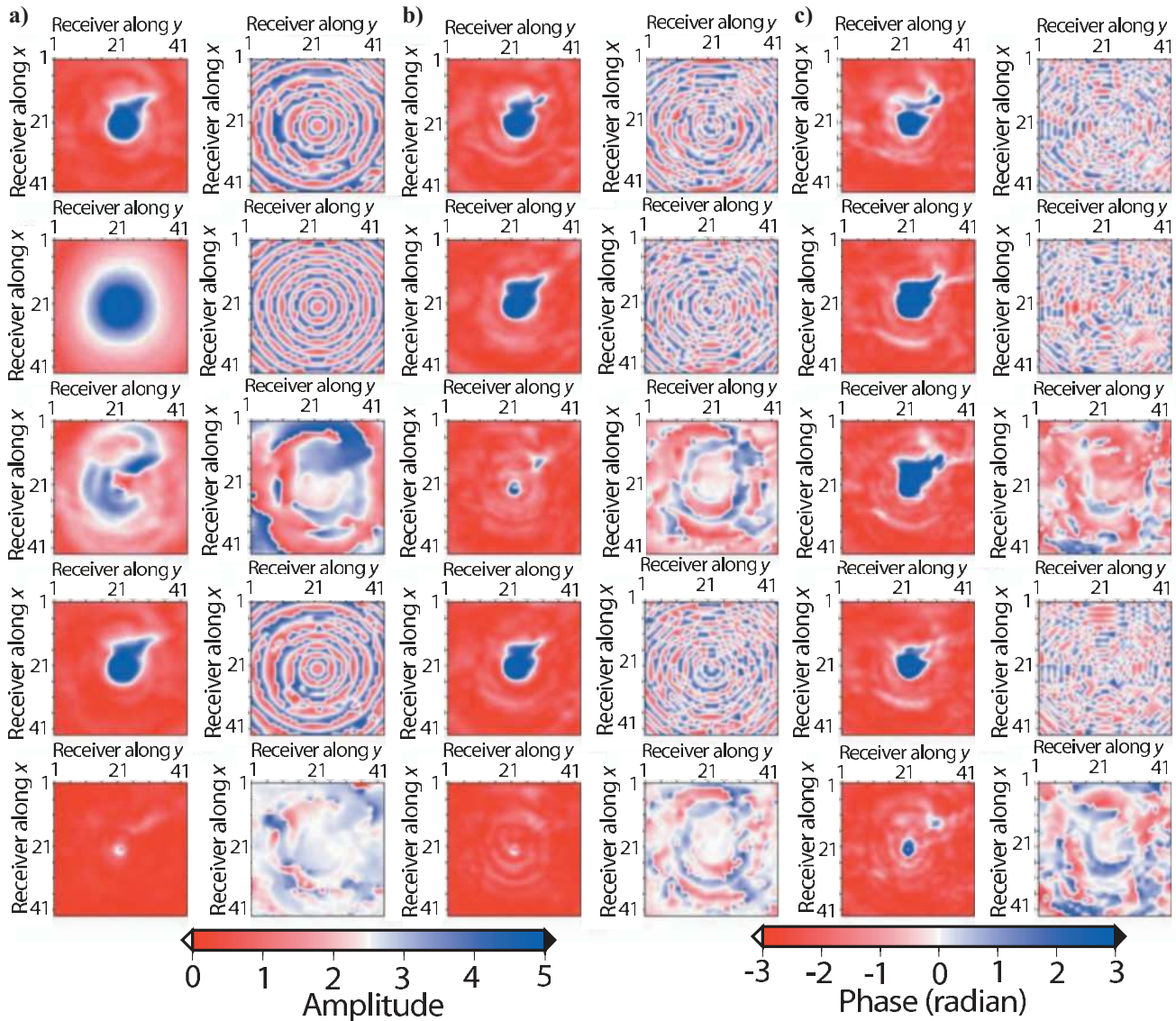
Figure 24. Imaging the thrust system in the overthrust model. The same as for Figure 23, but for a source located in the middle of the receiver array.

CONCLUSION

We have presented a 3D massively parallel frequency-domain FWI algorithm based on a direct solver. Advantages of our approach include the robustness of the forward problem provided by a high-performance direct solver, its efficiency to perform multishot simulations in relatively small 3D finite-difference grids, and a straight-forward parallelization of the inverse problem resulting from a domain decomposition of the monochromatic wavefields performed by the direct solver. By robustness, we mean that the time required to perform a simulation with a direct solver depends less on the complexity of the velocity model than the time required to perform this simulation with an iterative solver. Its main drawbacks are the memory and CPU time complexity of the LU factorization phase and its limited scalability, which restricts the size of the models and the frequency bandwidth that can be addressed on realistic distributed-memory platforms.

We have presented several applications on synthetic examples of increasing complexity to validate the algorithm and to give insight into the feasibility of our approach. Some preliminary applications to the overthrust model suggest that frequency-domain FWI can be applied successfully at low frequencies (< 7 Hz) on limited-size PC clusters to develop 3D velocity models with a maximum resolution on the order of half the wavelength (that is, 285 m for a velocity of 4000 m/s). This resolution scale should be compared with that provided by reflection tomography and migration-based velocity analysis to assess whether velocity models developed by 3D FWI can provide accurate background models for wave-equation PSDM.

Assessment of velocity models developed by 3D FWI as macro-models for PSDM will be one aim of future work. A second aim will be to assess whether accurate starting models for 3D FWI can be developed by first-arrival traveltime tomography, provided that global-offset recording is available. A third objective will be to mitigate the

memory limitations imposed by using a direct solver only.

In addition to time-domain and iterative approaches, a possible direction is to evolve toward hybrid direct-iterative solvers implemented in a domain-decomposition method based on the Schur complement approach, for which the direct solver is applied to subdomains of limited dimension while the iterative solver is applied to the grid points located at the boundaries between the subdomains. Compared to a purely iterative approach, the expected benefit is that the iterative solver in the hybrid approach is applied to a system of smaller dimension (the Schur complement system) that is better preconditioned theoretically. The drawback is that the hybrid approach remains more memory demanding because of the direct solver. The relevance of this approach will have to be demonstrated when many sources are considered.

REFERENCES

Amestoy, P. R., I. S. Duff, J. Y. L'Excellent, and J. Koster, 2007, Multifrontal massively parallel solver (MUMPS version 4.7.3) users'guide, http://enseeiht.fr/apo/MUMPS/, accessed June 2008.
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/EAGE 3-D modeling series 1, SEG.
Ashcraft, C., and J. W. H. Liu, 1998, Robust ordering of sparse matrices using multisection: SIAM Journal on Matrix Analysis and Applications, **19**, 816–832.
Berenger, J.-P., 1994, A perfectly matched layer for absorption of electromagnetic waves: Journal of Computational Physics, **114**, 185–200.
Bickel, S. H., 1990, Velocity-depth ambiguity of reflection traveltimes: Geophysics, **55**, 266–276.
Biondi, B., and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.
Bishop, T. N., K. P. Bube, R. T. Cutler, R. T. Langan, P. L. Love, J. R. Resnick, R. T. Shuey, and D. A. Spinder, 1985, Tomographic determination of velocity and depth in laterally varying media: Geophysics, **50**, 903–923.
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.
Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, 1457–1473.
Capdeville, Y., 2005, Towards global earth tomography using the spectral element method: A technique based on source stacking: Geophysical Journal International, **162**, 541–554.
Claerbout, J. F., 1971, Toward a unified theory of reflector mapping: Geophysics, **36**, 467–481.
Clarke, R., G. Xia, N. Kabir, L. Sirgue, and S. Michell, 2007, Processing of a novel deepwater, wide-azimuth node seismic survey: The Leading Edge,

26, 504–509.
Duff, I. S., and J. K. Reid, 1983, The multifrontal solution of indefinite sparse symmetric linear systems: ACM Transactions on Mathematical Software: ACM Transactions on Mathematical Software, **9**, 302–325.
Forgues, E., and G. Lambaré, 1997, Parameterization study for acoustic and elastic ray + Born inversion: Journal of Seismic Exploration, **6**, 253–278.
Hicks, G., 2002, Arbitrary source and receiver positions in finite-difference schemes using Kaiser windowed sinc functions: Geophysics, **67**, 156–166.
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.
Hustedt, B., S. Operto, and J. Virieux, 2004, Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling: Geophysical Journal International, **157**, 1269–1296.
Jo, C. H., C. Shin, and J. H. Suh, 1996, An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator: Geophysics, **61**, 529–537.
Lailly, P., 1984, The seismic inverse problem as a sequence of before stack migrations: SIAM Conference on Inverse Scattering, Theory and Application, 206–220.
Marfurt, K., 1984, Accuracy of finite-difference and finite-elements modeling of the scalar and elastic wave equation: Geophysics, **49**, 533–549.
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.
Operto, S., G. Lambaré, P. Podvin, and P. Thierry, 2003, 3D ray-Born migration/inversion, part 2: Application to the SEG/EAGE overthrust experiment: Geophysics, **68**, 1357–1370.
Operto, S., C. Ravaut, L. Improta, J. Virieux, A. Herrero, and P. Dell' Aversana, 2004, Quantitative imaging of complex structures from multifold wide aperture seismic data: Geophysical Prospecting, **52**, 625–651.
Operto, S., J. Virieux, P. Amestoy, J.-Y. L'Éxcellent, L. Giraud, and H. Ben-Hadj-Ali, 2007, 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver, A feasibility study: Geophysics, **72**, no. 5, SM195-SM211.
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**, B09306 doi: , doi: 10.1029/2005JB003835.
Operto, S., J. Virieux, and A. Ribodetti, 2005, Multiple arrival migration/inversion of real global aperture seismic data: 67th Annual Conference and Exhibition, EAGE, Extended Abstracts, P272.
Operto, S., J. Virieux, and F. Sourbier, 2008, Documentation for FWM2D/FWT2D programs: 2D acoustic frequency-domain waveform modelling/tomography, http://seiscope.unice.fr/FWT/doc.html, accessed December 2007.
Operto, S., S. Xu, and G. Lambaré, 2000, Can we image quantitatively complex models with rays?: Geophysics, **65**, 1223–1238.
Plessix, R. E., 2007, A Helmholtz iterative solver for 3D seismic-imaging problems: Geophysics, **72**, no. 5, SM185-SM194.
Pon, S., and L. Lines, 2005, Sensitivity analysis of seismic depth migrations: Geophysics, **70**, no. 2, S39-S42.
Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, part I: Theory and verification in a physic scale model: Geophysics, **64**, 888–901.
———, 2004, Velocity models from frequency-domain waveform tomography: Past, present and future, Presented at the 66th Annual Conference and Exhibition, EAGE, workshop W8.
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 M. H. Worthington, 1990, Inverse theory applied to multisource cross-hole tomography. Part 1: Acoustic wave-equation method: Geophysical Prospecting, **38**, 287–310.
Ravaut, C., S. Operto, L. Improta, J. Virieux, A. Herrero, and P. dell' Aversana, 2004, Multi-scale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-wavefield inversions: Application to a thrust belt: Geophysical Journal International, **159**, 1032–1056.
Riyanti, C. D., A. Kononov, Y. A. Erlangga, C. Vuik, C. Oosterlee, R. E. Plessix, and W. A. Mulder, 2007, A parallel multigrid-based preconditioner for the 3D heterogeneous high-frequency Helmholtz equation: Journal of Computational Physics, **224**, 431–448.
Sava, P. C., and B. Biondi, 2005, Wave-equation migration velocity analysis — I, Theory: Geophysical Prospecting, **52**, 593–606.
Shin, C., K. Yoon, K. J. Marfurt, K. Park, D. Yang, H. Y. Lim, S. Chung, and S. Shin, 2001, Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion: Geophysics, **66**, 1856–1875.
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, 2007a, 3D full waveform inversion: Wide versus narrow azimuth acquisitions: 77th Annual International Meeting, SEG, Expanded Abstracts, 1760–1764.

Sirgue, L., T. J. Etgen, U. Albertin, and S. Brandsberg-Dahl, 2007b, System and method for 3D frequency-domain waveform inversion based on 3D time-domain forward modeling: U.S. Patent US2007/0282535 A1.

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

Song, Z., P. Williamson, and R. Pratt, 1995, Frequency-domain acoustic-wave modeling and inversion of crosshole data, Part 2: Inversion method, synthetic experiments and real-data results: Geophysics, **60**, 784–795.

Sourbier, F., S. Operto, J. Virieux, P. Amestoy, and J. Y. L'Excellent, 2007, FWT2D: A massively parallel frequency-domain full-waveform algorithm for imaging acoustic media: 77th Annual International Meeting, SEG, Expanded Abstracts, 1893–1897.

Stekl, I., and R. G. Pratt, 1998, Accurate viscoelastic modeling by frequency-domain finite difference using rotated operators: Geophysics, **63**, 1779–1794.

Stekl, I., M. R. Warner, and A. P. Umpleby, 2007, 3D frequency domain waveform inversion — Synthetic shallow channel example: 69th Annual Conference and Exhibition, EAGE, Extended Abstracts, C026.

Stork, C., 1992, Reflection tomography in the post migrated domain: Geo-physics, **57**, 680–692.

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

——, 1987, Inverse problem theory: Methods for data fitting and model parameter estimation: Elsevier Scientific Publishing Co., Inc.

Toksöz, M. N., and D. H. Johnston, 1981, Seismic wave attenuation: SEG

Warner, M., I. Stekl, and A. Umpleby, 2007, Full wavefield seismic tomography — Iterative forward modelling in 3D: 69th Annual Conference and Exhibition, EAGE, Extended Abstracts, C025.

Williamson, P., 1991, A guide to the limits of resolution imposed by scattering in ray tomography: Geophysics, **56**, 202–207.

Williamson, P. R., and M. H. Worthington, 1993, Resolution limits in ray tomography due to wave behavior, Numerical experiments: Geophysics, **58**, 727–735.

Wu, R.-S., and M. N. Töksoz, 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11–25.

Zhang, Y., S. Xu, and G. Zhang, 2006, Imaging complex salt bodies with turning-wave one-way wave equation: 76th Annual International Meeting, SEG, Expanded Abstracts, 2323–2327.

Zhang, Y., G. Zhang, D. Yingst, and J. Sun, 2007, Explicit marching method for reverse-time migration: 77th annual meeting, SEG, Expanded Abstracts, 2300–2304.