Comput Geosci (2017) 21:31–45
DOI 10.1007/s10596-016-9593-0

ORIGINAL PAPER

Acceleration strategies for elastic full waveform inversion workflows in 2D and 3D

Near offset elastic full waveform inversion

Jean Kormann[1,2] · Juan E. Rodríguez[1] · Miguel Ferrer[1] · Albert Farrés[1] · Natalia Gutiérrez[1] · Josep de la Puente[1] · Mauricio Hanzich[1] · José M. Cela[1]

Received: 10 May 2016 / Accepted: 20 September 2016 / Published online: 22 October 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract Full waveform inversion (FWI) is one of the most challenging procedures to obtain quantitative information of the subsurface. For elastic inversions, when both compressional and shear velocities have to be inverted, the algorithmic issue becomes also a computational challenge due to the high cost related to modelling elastic rather than acoustic waves. This shortcoming has been moderately mitigated by using high-performance computing to accelerate 3D elastic FWI kernels. Nevertheless, there is room in the FWI workflows for obtaining large speedups at the cost of proper grid pre-processing and data decimation techniques. In the present work, we show how by making full use of frequency-adapted grids, composite shot lists and a novel dynamic offset control strategy, we can reduce by several orders of magnitude the compute time while improving the convergence of the method in the studied cases, regardless of the forward and adjoint compute kernels used.

Keywords Inversion · Elastic · Near offset

Mathematics Subject Classification (2010) 93.85.Rt · 93.85.Bc

✉ Jean Kormann
jean.kormann@unileoben.ac.at

[1] Barcelona Supercomputing Center (BSC), Jordi Girona 29, 08034 Barcelona, Spain

[2] Montanuniversitaet Leoben, Chair of Applied Geophysics, Peter Tuner Straße, 8700 Leoben, Austria

1 Introduction

The retrieval of physical properties of the Earths interior using seismic data has always been challenging for both academia and industry. As such, it has been subjected to intensive research for the last decades. One of the methods that potentially allows to extract more information from seismic data than travel-time tomography is full waveform inversion (FWI). Theoretically, it can provide models of physical parametres with higher spatial resolution than other classical methods such as travel-time tomography. In parallel, due to the increase in computational power, large-scale complex forward modelling has become affordable.

FWI is commonly formulated as an iterative process that seeks to improve the model by minimizing the differences between the synthetic traces obtained using a reference model (which can be poor in high-frequency content) with the real seismic trace recorded experimentally, hereafter referred to as data, by means of a cost functional [44]. Nevertheless, and given the number of parametres to be inverted, statistical methods such as Monte Carlo are still not applicable for large 3D problems, especially when high-frequency contents are involved. In this sense, the centre piece of FWI is the so-called adjoint method, which allows us to obtain the gradient of the misfit function for the current model by cross-correlating both incident and back-propagated wavefield [15, 16, 31, 32, 44]. With the gradient at hand, an iterative optimization problem can be set in order to find the model that fits best the field data.

Although conceptually simple, FWI application to field data remains challenging due to the numerous inherent problems. Extreme non-linearity, irregular non-convex misfit functions, compensation for geometrical spreading, insufficient knowledge of the physics and inadequate

starting models together with low-frequency limitations in the data are some of the problems that must be overcome when inverting a seismic dataset. This explains the sparking interest for more complex and accurate inversion methods. Additional mathematical insights such as preconditioning and regularization also play an important role in reducing the non-linearity of the problem and are often required in order to ensure convergence [3, 12, 20]. On the other hand, FWI becomes very challenging as at least two parametres must be determined either independently or using constrains, and hence, the parametre space becomes larger and more complex [7, 8, 46].

In addition to the algorithmic challenges of elastic FWI, sometimes referred to as EFWI, there is an inherent barrier which is the cost associated with running the scheme. In order to mitigate the computational burden, previous work has mostly focused on improving the efficiency of the computational kernel [11, 13, 21], which takes most of the FWI execution time. The computational kernel is responsible for simulating seismic wavefields with the purpose of measuring data misfits, generating gradients and obtaining approximations to the Hessian. Some acceleration efforts focusing on the FWI computing kernels include porting these kernels to accelerator-based hardware [5, 11, 39], optimizing the kernels' performance in general purpose processors [23] for off-the-shelf hardware or parallelizing the kernels in many compute units [30, 36, 38]. Nevertheless, we wish to show here that there are workflow strategies that are orthogonal to kernel optimization and which result in notable computational savings, thus making elastic FWI a routinely applicable tool for 3D datasets. Such strategies are mostly independent on the specifics of the computed kernel such as whether it is specified in time or frequency domain, uses finite differences or finite elements or whether it is highly parallel or not. We will exemplify the most important strategies that we have explored for a set of 2D and 3D cases, and show how some of them even result in an improvement in model fit and a reduction in kernel size, both collateral benefits for an FWI application. In summary, we will apply the following strategies: (1) multi-scale grid-adapted gradient generation, (2) shot decimation by means of composite shot lists and (3) dynamic offset control (DOC) to focus on reflection events with minimal user interaction.

This paper is structured as follows: in the first section, we briefly present the adjoint method and the more important features of our algorithm. In the second section, we present our strategy for selecting a dataset and show that it allows for obtaining accurate inversion. We also illustrate the potential of our approach by including density effects in the modelled datasets. In the third section, we will apply our workflow to a 3D elastic model and discuss some computational aspects and advantages of our implementation.

2 Theory

We will briefly outline the general FWI algorithm and also remark which specializations have been made in our implementation, used in the examples hereafter. The mathematical formulation of FWI consists of the minimization of an error or misfit function $E(\mathbf{m})$ [44, 48], where $\mathbf{m}$ is the current model. The computation of E requires modelling accurately the response of the model $\mathbf{m}$ by means of a simulation. Many choices of E functionals exist, L^2 being the most common choice. However, some other norms have been successfully applied to the FWI problem, such as the L^1 norm [7, 8, 45, 48], the Huber norm [19] or separating phase and envelope misfits in the time-frequency domain [16, 24]. In this study, we employ the normalized least square criterion L^2 given by

$$E_k(\mathbf{m}_k) = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{n}\left[\frac{\mathbf{u}_j^i(\mathbf{x}_r,\mathbf{x}_s;\mathbf{m}_k)}{\max_j(|\mathbf{u}^i(\mathbf{x}_r,\mathbf{x}_s;\mathbf{m}_k)|)} - \frac{\mathbf{u}_j^i(\mathbf{x}_r,\mathbf{x}_s)}{\max_j(|\mathbf{u}^i(\mathbf{x}_r,\mathbf{x}_s)|)}\right]^2, \tag{1}$$

where N is the number of receivers, n the number of samples of each trace, $\mathbf{x}_r$ and $\mathbf{x}_s$ are the receiver and source position respectively, $\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s;\mathbf{m}_k)$ the components of the displacement vector obtained from the current model parametres $\mathbf{m}_k$ at the kth iteration and $\mathbf{u}(\mathbf{x}_r,\mathbf{x}_s)$ stands for this recorded in the seismic experiment. A misfit function such as Eq. 1 is typically an irregular surface, and hence, finding a global minimum can be challenging if not impossible. In this context, multi-scale techniques [4, 9, 32, 35, 40, 48], combined with gradient preconditioning and regularization methods, have been developed to sequentially incorporate higher wavenumber information into the inverted model. This results in a reduction of local minima effects by means of either selecting increasing individual frequency samples (in frequency domain) or broadening a low-pass filter by increasing its corner frequency (in time domain).

We represent seismic wavefields with the elastic wave equation and use a time-domain approach to simulate synthetic wavefields. The equation reads

$$\begin{aligned}\rho(\mathbf{x})\dot{\mathbf{v}}(\mathbf{x},t) &= \nabla\cdot\boldsymbol{\sigma}(\mathbf{x},t) + \mathbf{f}_s(\mathbf{x}_s,t),\\ \dot{\boldsymbol{\sigma}}(\mathbf{x},t) &= \mathbf{C}(\mathbf{x}):\nabla\mathbf{v}(\mathbf{x},t)\end{aligned} \tag{2}$$

where f_s is the source function at position $\mathbf{x}_s$, $\mathbf{v}$ the particle velocity, ρ the density and $\boldsymbol{\sigma}$ the stress field, which is related to the strains by $\boldsymbol{\sigma} = \mathbf{C}:\boldsymbol{\varepsilon}$ being $\mathbf{C}$ the stiffness tensor which defines the elastic properties of the model. Notice that for the isotropic case, the model has only three free parametres: λ, μ and ρ. Our seismic modelling scheme is

an eighth order in space and second order in time staggered-grid finite-difference scheme in time domain [28, 47] that uses a mimetic approach to accurately solve the free-surface condition and hence allows for a less restrictive grid spacing criterion in the computations [33]. For the bottom and lateral boundaries, we use sponge zones as reflection-free condition [41].

The model gradient associated with the misfit function E of each source or shot is constructed by cross-correlation of a forward and adjoint wavefield, obtained from an adjoint source as explained by [14]. Using the strains of both forward ($\boldsymbol{\varepsilon}$) and adjoint ($\boldsymbol{\varepsilon}^\dagger$) wavefields, the kernels for ρ, λ and μ are defined by

$$K_\rho = -\int_T \mathbf{v}\cdot\mathbf{v}^\dagger dt$$
$$K_\lambda = \int_T \mathrm{tr}(\boldsymbol{\varepsilon})\mathrm{tr}(\boldsymbol{\varepsilon}^\dagger)dt$$
$$K_\mu = \int_T 2\boldsymbol{\varepsilon} : \boldsymbol{\varepsilon}^\dagger dt\,. \tag{3}$$

where $\mathbf{v}$ and $\mathbf{v}^\dagger$ stand for the forward and adjoint particle velocity fields, respectively, and T the simulation's duration. The search direction $\mathbf{P}_k$ at the kth iteration is obtained by means of a non-linear conjugate gradient method and is given by

$$\mathbf{P}_k = -\nabla_{\mathbf{m}_k} E_k + \beta_k \mathbf{P}_{k-1}\,, \tag{4}$$

where a Polak-Ribière criterion is used to obtain β_k. The last stage consists in a line search algorithm that finds a satisfactory α_k and then updates the models according to

$$\mathbf{m}_k = \mathbf{m}_{k-1} + \alpha_k \mathbf{P}_k. \tag{5}$$

Nevertheless, using gradients obtained directly from the kernel results in a slowly converging algorithm; thus, we apply a two-step preconditioning: At a first stage, we apply a change of variables and choose to work with $\tilde{\mathbf{m}}=\log_e(\mathbf{m})$ instead of $\mathbf{m}$ as explained by [25]. In all the numerical tests, we observed that using $\log_e$ leads to faster convergence, reducing the number of iterations per frequency. The second step compensates for geometrical spreading by means of dividing the gradients by the square of the forward illumination; this combination has shown to be sufficient in all the tests we performed. The complete elastic FWI workflow is shown in Fig. 1 and is rather general for FWI applications. Notice that the scheme typically involves many shots, and hence, it is beneficial to apply parallelization techniques at the shot level, as indicated in the grey boxes of Fig. 1. Particular to our own FWI implementation, we wish to remark that in the following examples, the computational grid does not match the receiver or source grids at any given frequency band. In addition, we have not fixed any part of the model neither have we preprocessed the synthetic data to select particular phases. In order to avoid near-field effects,

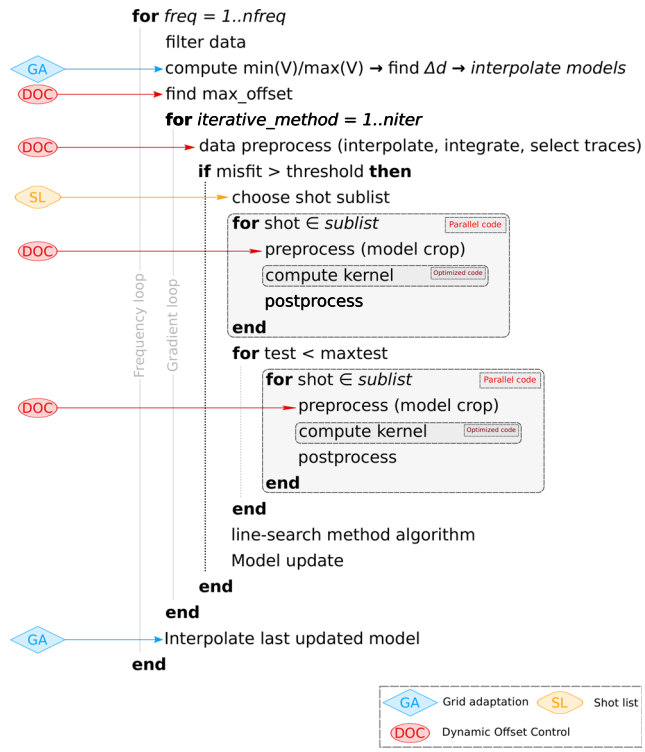

Fig. 1 Algorithm used for full waveform inversion. The optimization and acceleration features are indicated in colour

we typically remove the nearest offset traces for each shot (typically offsets < 150 m). This helps reducing the source footprint in the gradients.

2.1 Workflow optimization overview

Three main optimizations are suggested at the workflow level, which affect the performance and quality of elastic FWI. Before analyzing in detail their behaviour, we proceed to outline their main traits.

– *Frequency-adapted grids*: Full wavefield seismic modelling algorithms are limited by resolution constraints, which are determined by the dispersion properties of the modelling algorithm. As a consequence, the grid spacing Δ needed to simulate wavefields accurately is proportional to the minimum wave velocity in the model and inversely proportional to the maximum frequency modelled. The lower bound to the inverse proportionality constant is called *points per wavelength* or ppw and is constant through all practical frequencies. Notice that ppw is determined by the actual algorithm and the accuracy requirements set by its user. As the ppw is a lower bound, the standard approach is setting the grid spacing Δ to the minimum required by the highest inverted frequency and minimum velocity, thus setting

a constant grid for the complete FWI procedure. Nevertheless, small Δ values result in many grid points and thus expensive simulations. Furthermore, the minimum velocity value might vary due to model updates during the FWI iterations. We propose adapting Δ to the optimum value at each inverted frequency, thus regridding the model to suit the new sampling for modelling [27]. In a typical multi-scale FWI [9, 32], this means large computational savings at the lowest frequencies inverted but no benefit at the highest frequencies, which in turn lead the cost count of the complete FWI procedure. In our case, we use slowness and density trilinear interpolation in a Cartesian grid to move between scales, but different modelling algorithms might favour different regridding options. This reduces the computing time required to obtain a model update, compared to using the maximum frequency grid spacing at each iteration.

– *DOC*: This strategy is a data decimation approach which limits the maximum offset used at each shot. The technique was originally designed so that surface wave effects do not become dominant in the gradient, so that reflection modes keep updating the model in elastic FWI. This does not mean that we neglect surface waves, but their overall impact on the gradient is reduced. As a data decimation approach, DOC requires additional validation of its impact on the resulting models compared to full data approaches. Such validation will be shown in the next section. Incidentally, offset limitation has a strong impact on performance, as the size of each shot resolved is much smaller than the original full-offset shot. Notice that this might end in dissimilarly sized shot domains and, hence, parallel imbalance unless master-worker strategies as those described above can be used.

  This technique is also particularly well suited to master-worker FWI applications [22, 37], where a different amount of computational resources can be allocated to each worker. Furthermore, the parallel efficiency and load balance of such schemes are excellent whenever the number of tasks (e.g. shots) is much larger than the amount of available workers [2].

– *Composite shot lists*: This last strategy [27, 49] simply distributes the complete shot list in a subset of shots which are run sequentially at each gradient iteration. In this way, for a fixed number of gradient iterations, the total amount of shots simulated is smaller, reducing the computational resources needed. Additionally, no cross-talk problems appear such contrary to source encoding techniques [3, 26]. Furthermore, composite shot lists can be combined with the DOC strategy described above and virtually any acquisition geometry can be used. Last but not the least, as in any other multi-shooting strategies, depending on data redundancy, the savings from composite shot lists might be large and apply to all shot loops in the algorithm.

In Fig. 1, we can identify where each of the strategies described above affects the standard elastic FWI workflow. Each strategy is constrained to a particular set of workflow lines and aims at either reducing the size or the amount of the computing kernels that need to be run. In the next sections, we will exemplify the strategies with 2D and 3D models to illustrate both the quality and speedups obtained with respect to standard workflows, using exactly the same computational kernel in all cases.

3 Workflow evaluation in idealized scenarios

3.1 Offset range and inversion quality

In the following, we present the inversion of a 2D model based upon an x-line section of the 3D SEG/EAGE overthrust model [1]. The starting model is presented in Fig. 2 bottom. The acquisition geometry includes sources with a spacing of 100 m starting from $x = 1050$ m, resulting in a total of 151 shots. Receivers have the same spacing but start 50 m from the source, i.e. in a roll-along acquisition ending in roll-on/roll-off near the first and last shot locations. We also have included a free-surface condition at the top and the total simulation time is 6 s. The source used is a vertical force (z-direction) with a Ricker wavelet having a central frequency of 10 Hz. The model dimensions are 16×3.2 km. For the sake of simplicity, the V_s/V_p ratio is constant and set equal to 0.5 and ρ is set to 1000 kg/m^3. Figure 3 shows the multi-component data (P, v_x and v_z, respectively). We observe that for such elastic model, surface waves dominate the amplitudes. Nevertheless, both Lamé parametres are freely and simultaneously inverted.

Such data can be inverted using multi-scale elastic FWI using a low-pass filter with cutoff frequency at 2.5, 3.4, 5.1 and 6.4 Hz sequentially, running 20 iterations per frequency. The starting models consist in a Gaussian smoothing of the true ones (see Fig. 2 bottom) with a correlation length of 500 m as proposed in the literature [7, 46] which is similar in quality to those that could be obtained by travel-time tomography. The comparison between full offset and DOC is presented in Fig. 4 (top and bottom, respectively) and the multi-grid and multi-scale parametres used for inversion are summarized in Table 1. Although the main features are present and hence the inversion for full offset seems successful, the algorithm has been unable to attain a sharper image of the model, because it has become trapped in a local minimum. Our hypothesis is that high-frequency content is strongly related to surface waves having large

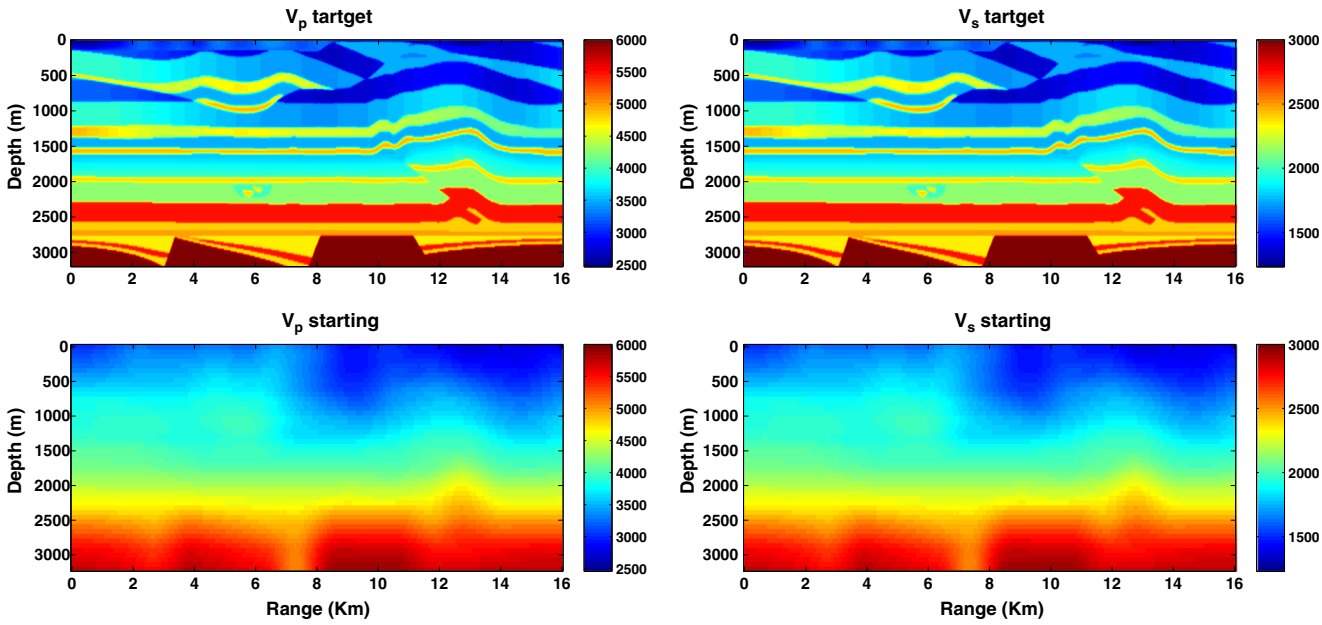

Fig. 2 Target and starting models for elastic inversion. From *top to bottom* and *left to right*: the V_p and V_s target models; V_p and V_s starting models

amplitudes and hence taking a prevalent role in the inversion at the expense of information from deeper structures. Brossier et al. [7] suggested that such kind of data could be better inverted by applying sequential inversion to successively damped traces, in order to reduce the non-linearity of the misfit function. Alternatively, one could remove surface wave from the data and use elastic FWI with absorbing boundaries at the top of the model.

In this work, our approach for elastic FWI consists in employing data as is, without damping, only constraining the maximum offset such that we decimate the data in the spatial domain. In particular, we will limit the maximum

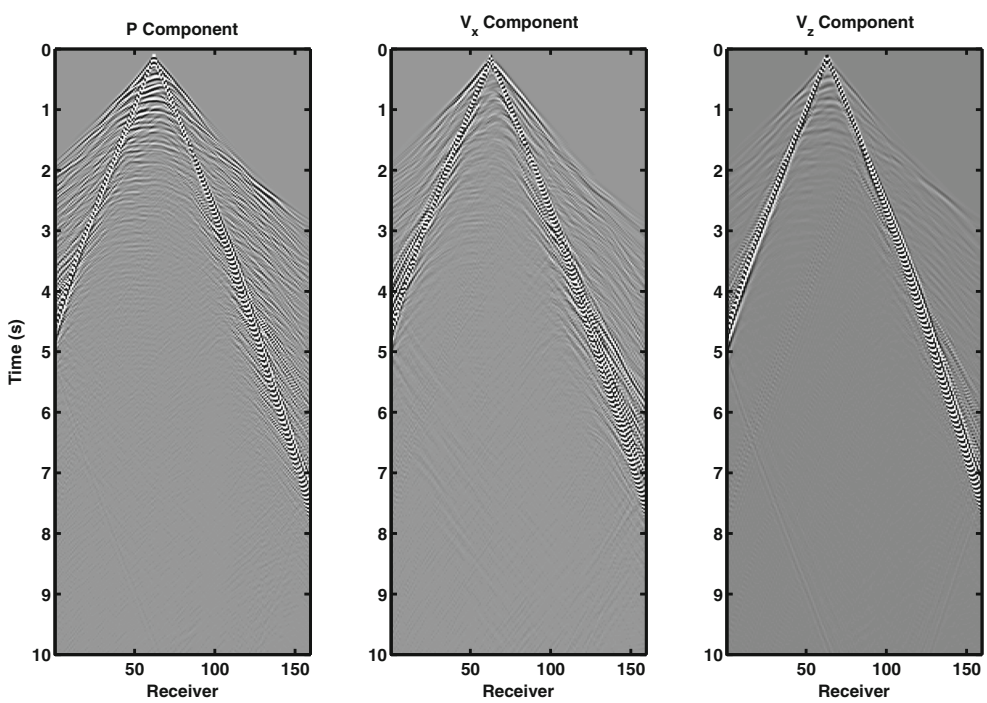

Fig. 3 Three-component shot gather obtained with a Ricker wavelet with central frequency of 10 Hz for a source located at the middle of the domain

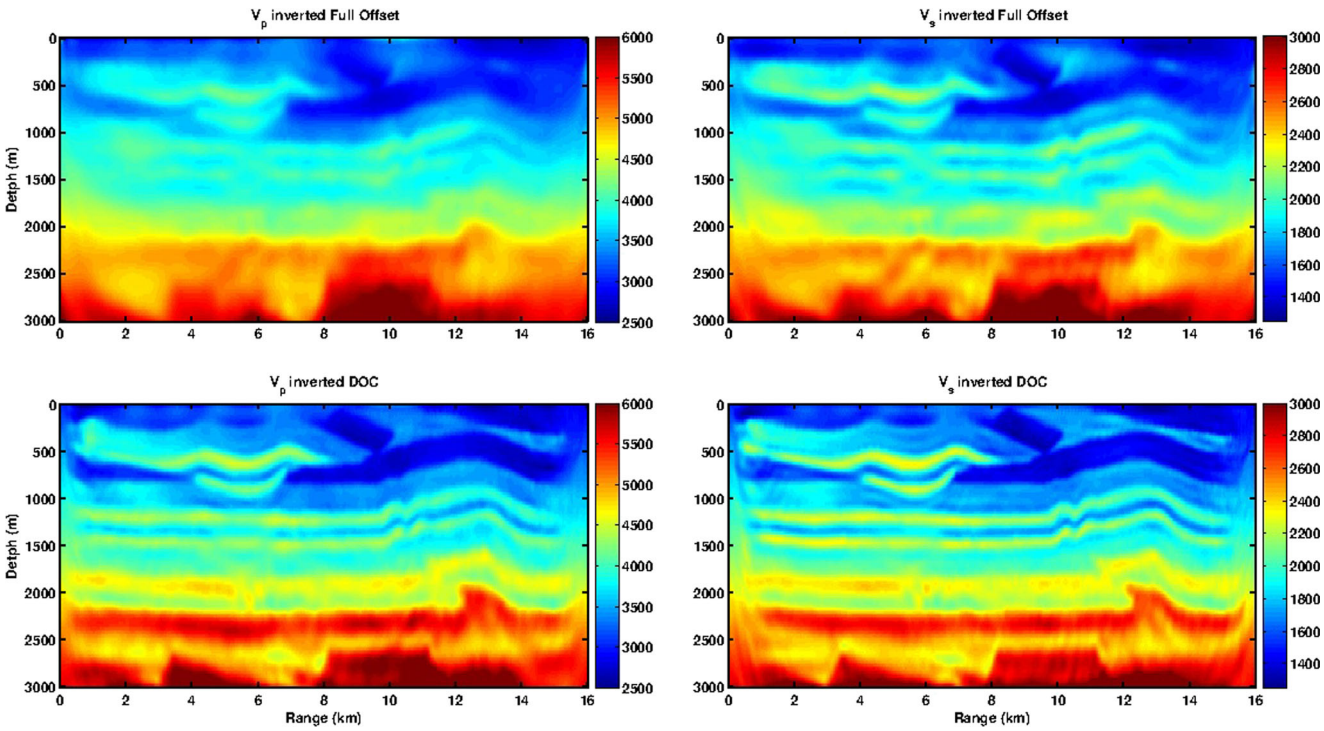

Fig. 4 From *top to bottom* and *left to right*: the V_p and V_s inverted models with full offset; V_p and V_s inverted models with DOC

offsets inverted by a distance proportional to the characteristic wavelength of the model. This means that the inverted offsets are smaller, the higher the corner frequency applied in our multi-scale method. Such strategy allows for a strong reduction of the surface wave footprint with respect to reflections and refractions. Furthermore, the method leads to a drastic reduction of the overall computational effort, depending on the offset selection strategy chosen. The maximum offset for the DOC method is given by

$$\max \|\mathbf{x}_r - \mathbf{x}_s\| = D/f_0 \tag{6}$$

where f_0 is the cutoff frequency of the low-pass filter applied to the data, D a parameter to be defined by the user, and x_s and x_r stand for the source and receiver positions, respectively.

Table 1 Multi-grid and multi-scale parametres for 2D FWI (from top to bottom: cutoff frequency, mesh size, spatial discretization and number of iterations)

FWI strategy				
Frequency	2.5 Hz	3.4 Hz	5.1 Hz	6.4 Hz
Mesh size	72 × 362	92 × 459	128 × 627	164 × 803
Δ	33.4 m	25.8 m	20.48 m	17.0
Offset	4800 m	3530 m	2352 m	1875 m
Number of it.	20	20	20	20

The example above uses $D = 2V_P$. In Fig. 5, we present the velocity maps of the resulting models obtained with DOC after a total of only 80 iterations. We observe very detailed structures at shallow depths; the two dipped structures at ranges 5 to 10 km at the bottom are well recovered as well. The small channel at range 6 km and depth 2 km is also correctly retrieved by our algorithm. Curiously, the V_p model is not as sharply recovered as the one presented in the work of Brossier et al. [7], likely due to using directional force sources instead of explosions and a penalization of direct waves inherent to offset limitation associated with DOC. Figure 6 compares the results of our inversion with vertical logs. We observe a very good match between target and inverted velocities, especially for the shear velocity. It seems clear that the upper part is almost perfectly recovered, with very fine details due to the inclusion of free-surface effects [29]. We acknowledge, nevertheless, that our inversion lost some accuracy at 12 km, seemingly because of cycle-skipping effects although the reflectors are correctly located.

3.2 The choice of parameter *D*

The objective of this subsection is to discuss the influence on our elastic FWI scheme of the parameter D in Eq. 6. To that goal, we perform a series of inversions of a well-known model, so that we can evaluate the model fit for different choices of D. The inversion parametres, starting models and

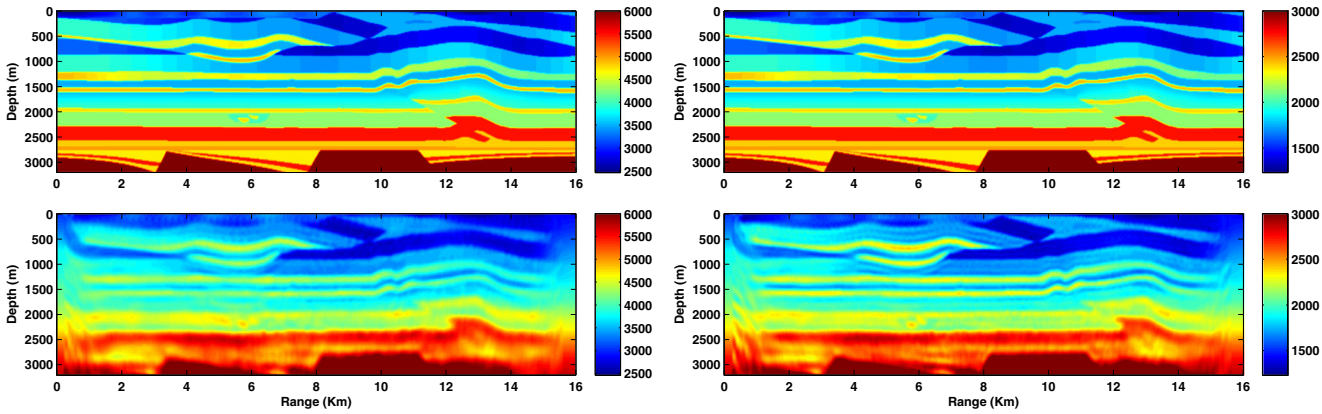

Fig. 5 From *left to right* and *top to bottom*: V_p target model, V_s target model, V_p inverted model, and V_s inverted model obtained with the DOC strategy of Table 1

strategy are those described in the previous section. We use a set of D values and evaluate the quality of the resulting inversions by means of the L^1 norm of the V_P and V_S obtained models; data misfit is harder to compare because inversions at each D value involve different amount of data. The range of D used is $V_s/2$, V_s, V_p, $2V_p$ and ∞.

Figure 7 displays errors at the end of each frequency inversion for each D value. It is clear that full offset is the worst option, already from the first frequency band. We observe that reducing D leads to more accurate inverted models, although there appears to be a limit beyond which reducing D results in worse inversion results. The best overall result is obtained with $D = V_P$, although the range $D \in [0.5V_P, 2V_P]$ seems satisfactory. In fact, there appears to be a different behaviour at low and high frequencies, with smaller D values preferred at low frequencies and slightly

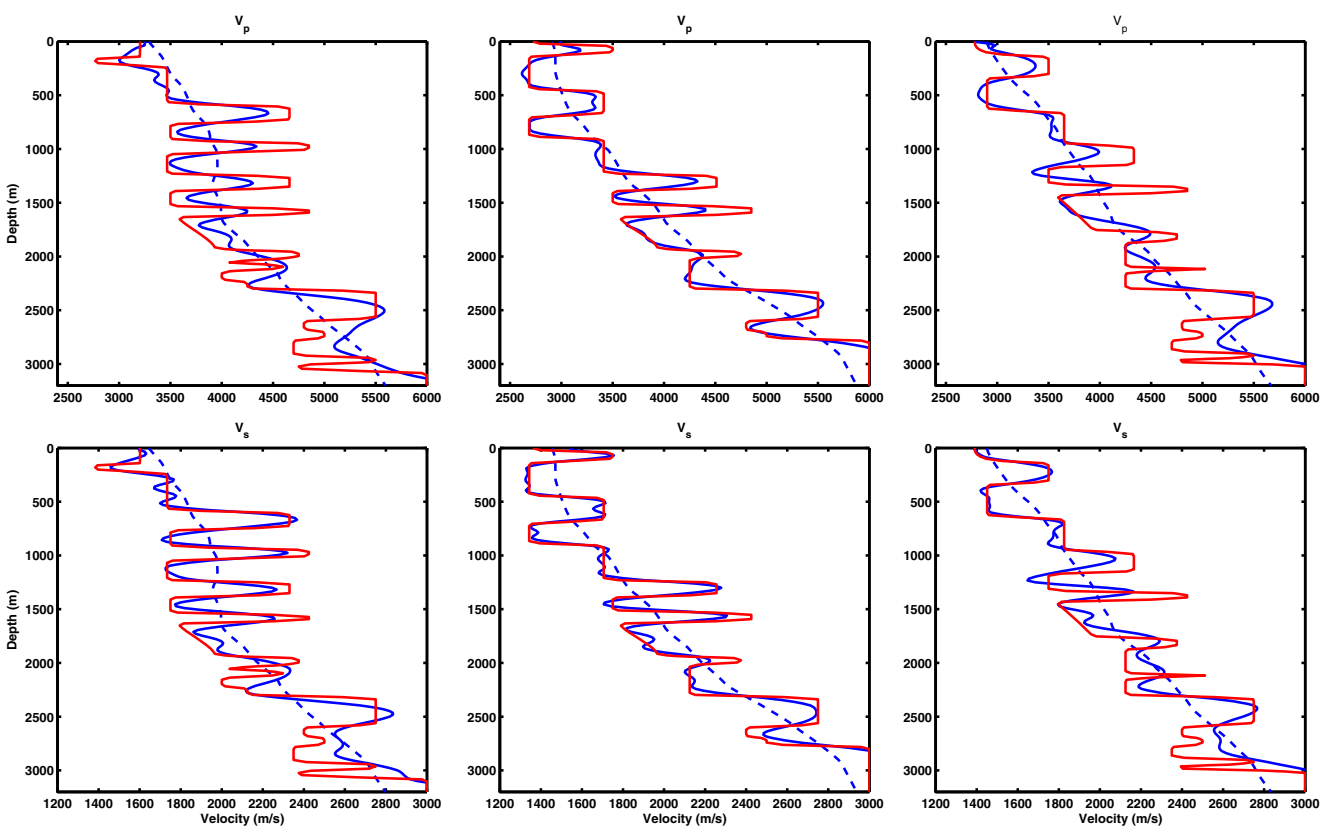

Fig. 6 V_p and V_s (*top to bottom*) profiles obtained with the DOC strategy of Table 1 at range 6, 9 and 12 km (*left to right*). The inverted, starting and target models are the *blue*, the *dash* and the *red line,* respectively

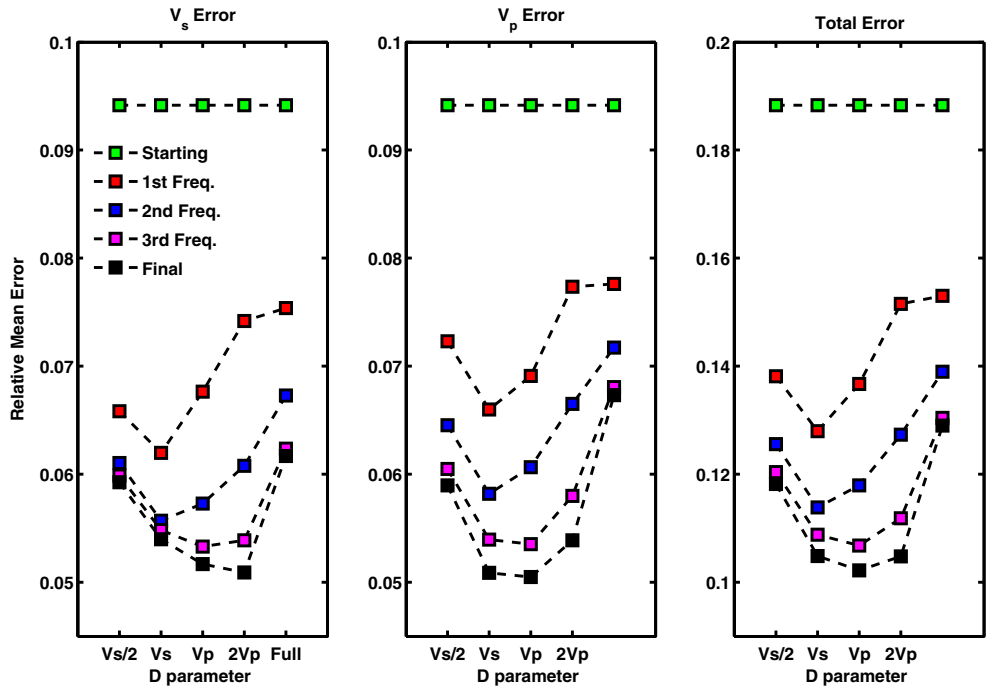

Fig. 7 Inverted V_p and V_s mean relative error with respect to the true velocity models: starting model, first, second, third and fourth frequency in *green*, *red*, *blue*, *magenta* and *black*, respectively

higher values for higher frequencies. This suggests that the optimal D value might depend on frequency. Nevertheless, the best value of D in real applications might be obtained by QC for a range of D at the lowest frequencies and keeping D constant throughout the frequency range.

We remark that DOC results in drastic reduction of the computational needs. If combined with frequency-adapted grid spacing and model slicing, we could obtain gradients of each shot with the least possible computational resources. In our case, shots cover the whole receiver span plus a padding region of 1500 m at both sides and absorbing boundaries. In Fig. 8, we plot the computational savings at each frequency for the 2D case previously depicted and a 3D scenario shown later in this article. In this figure, we use $D = V_P$ and savings are expressed as the fraction of the compute time required for a full offset (i.e. $D = \infty$). Clearly, the higher

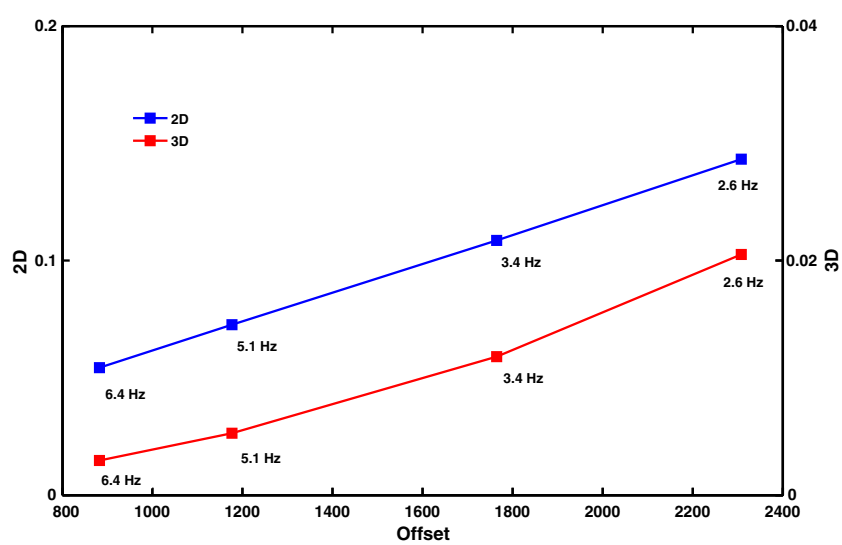

Fig. 8 Ratio of computational cost with and without DOC as a function of the offset for four frequencies in the 2D case (in *blue*, *left axis*) and in the 3D case (in *red*, *right axis*)

Table 2 Multi-grid and multi-scale parametres for 2D FWI when there is no linear relationship between shear and compressional velocities (from top to bottom: cutoff frequency, mesh size, spatial discretization and number of iterations)

FWI strategy						
Frequency	1.7 Hz	1.7 Hz	2.6 Hz	3.4 Hz	5.2 Hz	6.4 Hz
Mesh size	48 × 230	48 × 230	72 × 362	92 × 459	128 × 627	164 × 803
Δ	45.0 m	45.0 m	33.0 m	25.8 m	20.0 m	16.3 m
Offset	7058 m	7058 m	4800 m	3530 m	2352 m	1875 m
Number of it.	20	20	20	20	20	20
Time window	3.4 s	✗	✗	✗	✗	✗

the frequency, the higher the savings in CPU time, reaching up to a ×20 faster runs at 6.4 Hz. In 3D, the case is even stronger towards using DOC.

3.3 Test cases with independent V_s and V_p

We conclude this section with two test cases (namely cases 1 and 2), using uncorrelated V_s and V_p for modelling. For this purpose, we add a low-frequency perturbation invariant in the x-direction to the shear wave velocities of Fig. 2 (target) as illustrated in Fig. 10. Density is 0.415 times the V_p values. The starting models are those of the previous subsection. In case 1, we use sources in the z-direction and also include free-surface into the modelling. For case 2, we use explosive sources and model an infinite medium. Because of the missing low-frequency in the model, we can not make use of all the dataset as we do in the previous section, such that the inversion strategy differs somewhat from that presented in Table 1 for the starting frequency. First, we fail in starting at 2.5 Hz, regardless of the selected strategy or windowing. So we start to invert at 1.7 Hz. We also perform a first set of 20 iterations with only the first 3.4 s, in order to constrain our inversion to the first arrivals, which seems reasonable. In a second step, we relax the windowing and include all the arrivals. Table 2 summarizes the inversion strategy.

Figure 9 presents the inverted velocity maps for cases 1 and 2. The results confirm those of the previous subsection. We observe that the best images are obtained when using explosion sources and no free surface. Reflectors are mostly flat, with fine details in the geological structures. Even the low-velocity reservoir at range 6 km and depth 2000 m is correctly picked by our algorithm. We also observe a quite strong overshooting on the fracture zones, although the reflectors are correctly positioned. When using vertical force sources and free surface, reflectors are no longer flat as seen in the previous section. Figure 10 presents three velocity log profiles at ranges 4, 8 and 12 km, respectively.

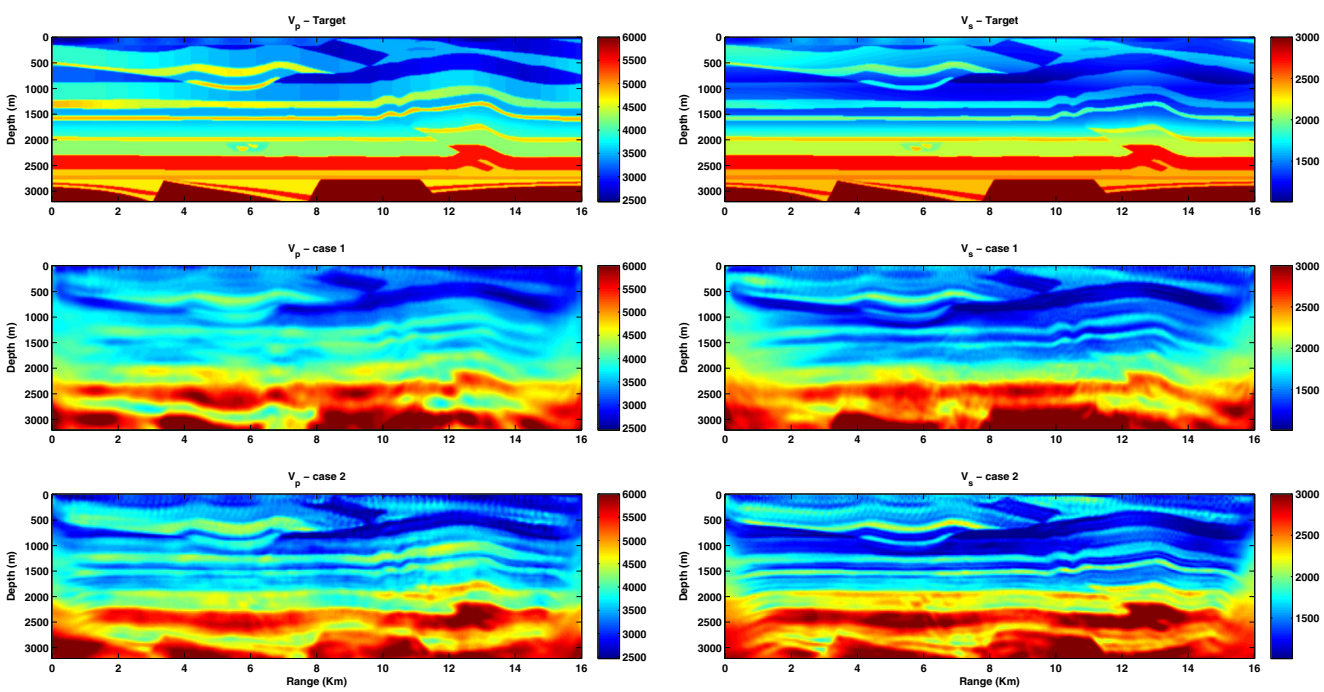

Fig. 9 Target (*top*) and final velocity models for V_p (*left*) and V_s (*right*) obtained with the DOC strategy of Table 2 for test cases 1 (*middle*) and 2 (*bottom*)

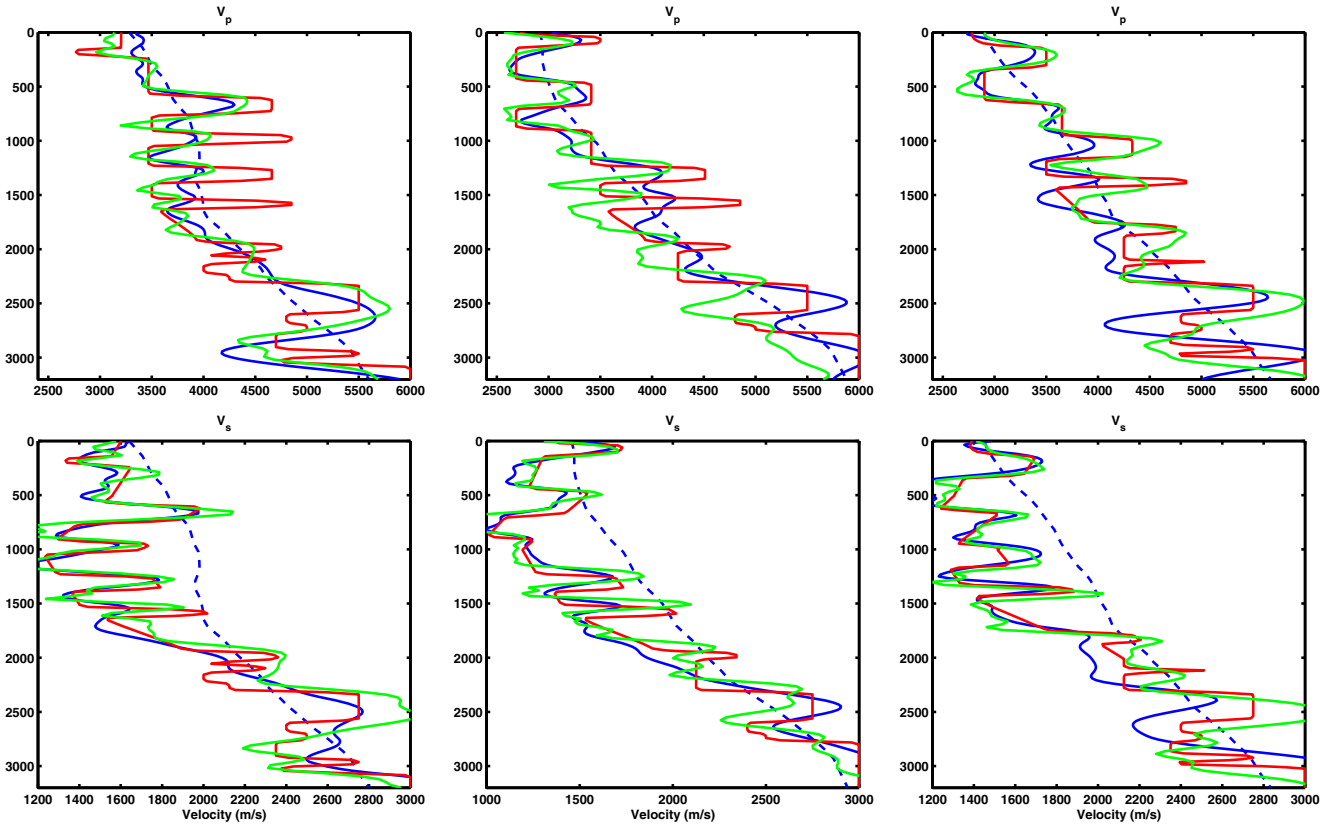

Fig. 10 Final velocity profiles for V_p and V_s (*top to bottom*) profiles obtained with the DOC strategy of Table 2 at range 4, 8 and 12 km (from *left to right*) for test cases 1 (*blue line*) and 2 (*green line*). The starting and target models are the *red* and *dash lines*, respectively

Detailed observations show that the V_p wave velocity is better recovered for test case 2. On the other hand, shear wave velocities appear sharper when the free surface is not included.

Nevertheless, in both cases, we remark that the algorithm has been able to converge efficiently towards an acceptable solution. On the other hand, we observe that the convergence is slightly smoother for case 2 than for case 1 although in both cases, data misfits are diminished to below 1 %.

4 Application to 3D elastic full waveform inversion

Previous studies have already presented results of 3D elastic full waveform inversion: [6, 10, 17, 18, 34, 42, 43, 46]. In order to validate our approach in 3D, the SEG/EAGE Overhrust model is used. The total model size is 16 × 16 × 3.2 km^3. The shear velocity is taken as half of the compressional velocity and density is constant and set to 1000 kg/m^3. The geometry acquisition consists in 79 fixed lines of 79 receivers each, resulting in 6241 four-component receivers equally spaced in both the x- and y-directions. This acquisition differs somewhat from that proposed in [46], in particular we use more sources but less receivers. Sources are located at the centre of each cell of four receivers (see Fig. 11) and only have a component in the z-direction (vertical forces). We used the same Ricker wavelet as described previously. The modelling algorithm is an eighth-order in space and second-order in time, explicit time-domain, staggered-grid, finite-differences method, which includes the free-surface reflections at the top of the model by means of mimetic operators

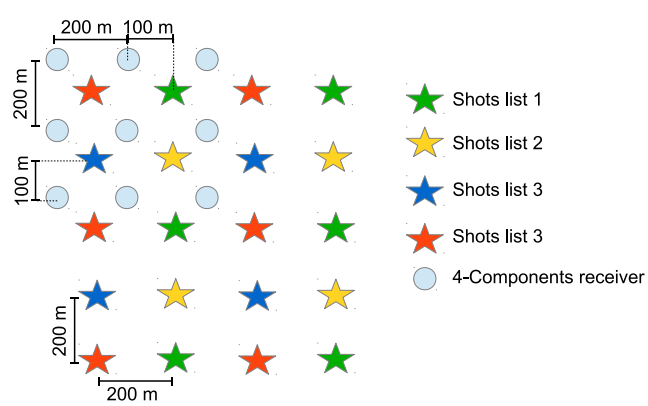

Fig. 11 Acquisition geometry for modelling together with the composite shot list distribution used for the 3D elastic full waveform inversion

Table 3 Multi-grid and multi-scale parametres for 3D FWI with constant density

FWI strategy				
Frequency	2.6 Hz	3.4 Hz	5.1 Hz	6.8 Hz
Mesh size	72 × 362 × 362	92 × 459 × 459	128 × 627 × 627	164 × 803 × 803
Δ	42.7 m	35.2 m	25.2 m	19.63 m
Offset	2310 m	1770 m	1170 m	885 m
Number of it.	10	10	10	10

[33]. The grid size, equal to the amount of staggered grid cells used for modelling, is 440 × 2201 × 2201 with $\Delta x = \Delta y = \Delta z = 7.25$ m. We use MPI-based domain decomposition, using 20 nodes for each shot computation. Inside each node, OpenMP takes care of a multi-thread approach to maximize the use of all available cores. We run the whole simulation on MareNostrum3[1] supercomputer resulting in 2.8 TB of synthetic traces. For inversion, we use only the first 55 lines, resulting in 3621 shots. At this point, inverting directly for all the shots is too expensive; hence, we use the composite shot lists strategy proposed by Warner et al. [49] and split the complete shot list in four sublists, in a staggered manner as shown in Fig. 11. This mostly leads to a similar illumination for each sublist, thus justifying the use of a non-linear conjugate gradient method for optimization, such that we can use the last search direction to update the new one. As a further acceleration strategy, we apply the DOC technique in order to reduce the domain size for forward, backward and test simulation, respectively, thus significantly reducing the computational cost, especially for the highest frequencies. As a side effect, we are no longer forced to use domain decomposition, because for all the frequencies inverted, all shots fit into a single node's memory. As mentioned before, this also reduces drastically the total amount of data used in the inversion, and consequently, the associated data processing operations such as low-pass filtering, time interpolation, adjoint source calculation or intermediate storage. In this numerical experiment, we choose to use the maximum of the shear velocity as D parametre, thus allowing for very small model sizes for each shot during inversion. The whole strategy for inversion is summarized in Table 3. The complete 3D elastic FWI run was executed in 88 nodes resulting in 3 % of the computational ressources used to simulate the dataset. Figure 12 shows the 3D volumes obtained after inversion. The overall results show a good match between target and inverted models. For both shear and compressional velocities, the fracture zone is clearly delineated, as well as the thinner layers. We also examine three random velocity profiles shown in Fig. 13, with $[x, y]$ positions at [6020 m 10010 m], [8016 m 10010 m] and [2034 m 4027 m]. We observe a good agreement between target and inverted velocities, especially for V_s. On the other hand, V_p lacks some resolution, allegedly due to the missing shorter wavelengths compared to V_s at the same frequency, but also because we did not carry out more iterations for computational reasons. At this point, we wish to recall the strategies which have allowed to speed up the computation of elastic 3D FWI in this case and what has been their relative contribution. This is summarized in Table 4. Due to the cost of running the FWI system without using the strategies presented in this work, in the following section, we will depict how to extrapolate such cost without actually running the whole inversion.

[1] http://www.bsc.es/marenostrum-support-services/mn3

5 Discussion

In the last example, we have shown how our elastic 3D FWI-accelerated workflow is able to sustain roughly a ×100 theoretical speedup compared to a conventional 3D elastic FWI workflow. By conventional workflow, we mean no grid coarsening neither DOC nor shot lists. In Table 4, it is clear how DOC compensates for the diminishing effect of grid adaptation to frequency, thus obtaining a significant speedup for the complete frequency range of our inversion. This, however, is a theoretical number that does not account for a number of effects and consequences of our strategy, namely overheads due to model and data interpolations, gradient tapering, load unbalancing or increased boundary/main-grid ratios. We can, nevertheless, use the run time of our forward modelling for this particular test in order to see how far we are from that theoretical speedup. We can do so because both the forward modelling (FM) and FWI kernels share the same code for wave propagation, boundary conditions, source insertion and receiver data retrieval. Taking t^{FM}_{10Hz} the total simulation time of the modelling, we can compute the cost $t^{convFWI}_{6.8Hz}$ of a conventional 3D FWI using the complete offset range but only the shot number and frequency content associated with the highest inverted band in our accelerated algorithm. The count is

$$t^{convFWI}_{6.8Hz} = \mathtt{it}_{\mathtt{FWI}} \cdot \mathtt{spg} \cdot \mathtt{scale}_{\mathtt{freq}} \cdot \mathtt{scale}_{\mathtt{shot}} \cdot t^{FM}_{10Hz}, \tag{7}$$

in which we use the total iteration count $\mathtt{it}_{\mathtt{FWI}} = 40$ as in Table 2, the amount of simulations per gradient computation

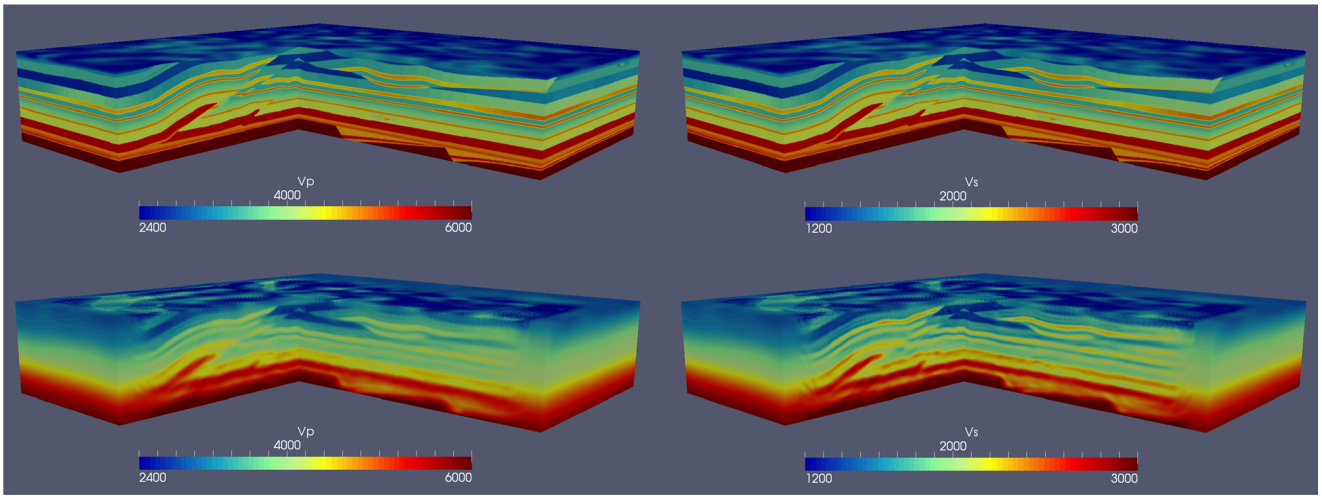

Fig. 12 From *left to right* and *top to bottom*: V_p target model, V_s target model, V_p inverted model, and V_s inverted model obtained with the DOC strategy

$\texttt{spg} = 4$ (two simulations to obtain a gradient and two tests to optimize the step length, ideally), the cost of modelling a lower frequency $\texttt{scale}_{\texttt{freq}} = 1/53$ (the amount of cell updates related to using $\Delta = 19.63$ m instead of $\Delta = 7.25$ m in explicit finite-differences) and that of using fewer shots $\texttt{scale}_{\texttt{shot}} = 0.718$ (from using 3612 shots instad of the original 5041). The aggregate results in

$$t_{6.8Hz}^{convFWI} = 2.17 \cdot t_{10Hz}^{FM} \,. \tag{8}$$

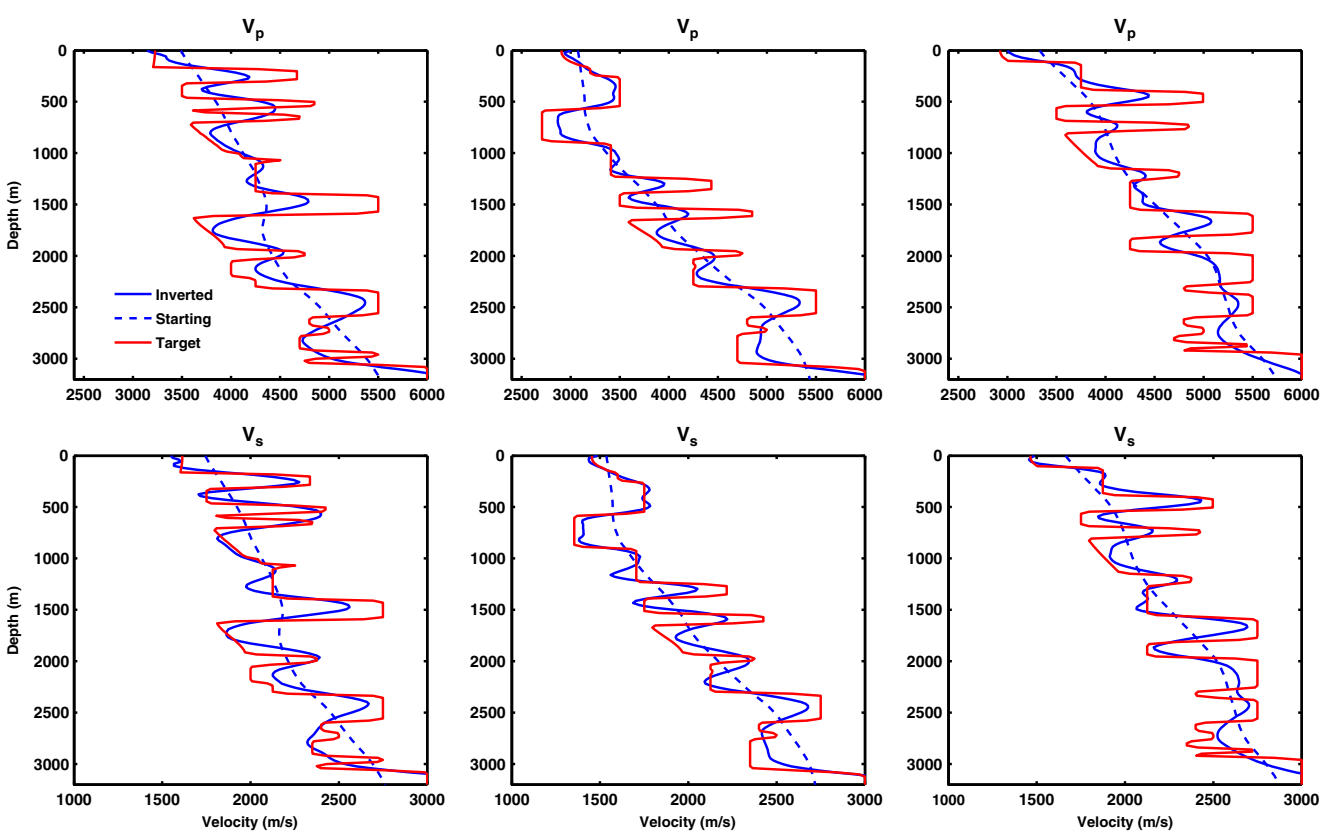

Fig. 13 *Top*: V_p profiles. *Bottom*: V_s for three particular locations. Inverted, starting and target models are the *blue*, the *dashed* and the *red lines*, respectively

Table 4 Theoretical speedups obtained by each acceleration strategy at each frequency

Accelerated vs conventional 3D EFWI speedup, theoretical				
Frequency	Grid adaptation	DOC	Shot lists	Total
2.6 Hz	×22.4	×4.4	×4	×394
3.4 Hz	×10.3	×6.0	×4	×247
5.1 Hz	×2.7	×9.0	×4	×97
6.8 Hz	×1.0	×11.3	×4	×45
All	×2.6	×6.7	×4	×103

Every speedup is computed with respect to running a conventional elastic 3D FWI workflow at the same frequencies and with the same shots and iterations. We have taken into account the additional 1500 m padding per side used in our DOC implementation

Now, normalizing the compute count of our inversion vs our modelling, we find

$$t_{6.8Hz}^{accFWI} = 53 \cdot t_{6.8Hz}^{convFWI} . \quad (9)$$

Hence, we can conclude that we are able to obtain an aggregate ×53 speedup compared to the cost of running the same inversion with a conventional 3D elastic FWI workflow, independent of the kernel architecture or its particular optimizations. This means that we are obtaining about a 51 % of the expected (theoretical) performance gain, for the reasons briefly described at the beginning of this section.

As for the result quality, arguably, using less data also results in a smaller inversion effort, although the data subtracted from the inversion dataset will not be able to improve the model. Nevertheless, in the previous sections, we have focused on explaining how, in our test cases, and without data pre-processing, such long-offset data often results in worse inverted models, most likely due to the algorithm getting stuck in a local minimum if no damping has been previously applied to traces. Similarly, the continuous use of interpolating schemes in our accelerated algorithm can account for extra errors in the inversion, as does narrowing the distance to the artificial absorbing boundary conditions resulting from DOC which can introduce extra artefacts into the gradients. Finally, the composite shot list strategy somehow should result in slower convergence. We do not account for the individual effects of all of these limiting issues in the present study. Nevertheless, the final inverted models show a good enough quality to justify the vast computational improvement allowed by the workflow presented here.

It is worth paying special attention to the novel dynamic offset control technique here, based upon reducing the maximum apperture of each shot in FWI, hence reducing the computational cost when moving into higher frequencies. The technique allows for fast convergence with no data picking even in the presence of ground roll. We have exemplified the method by analyzing the effects of larger offsets on inversion results, reaching the conclusion that a suitable value for D should be $\max(V_s) \leq D \leq 2\max(V_p)$. Nevertheless, we remark that we lack prior ways of determining the optimal D value, which can be considered a drawback of this technique. It is important to notice that the strategies shown here are not limited to a single acquisition type, but could be used for any land or marine acquisition without fundamental differences. This is in contrast to other strategies such as encoded multi-shooting [3, 26] which behaves poorly whenever no fix-spread acquisition is present.

We also remark that, in order to be able to take full advantage of the strategies presented here, it is recommended to use a workflow which can afford severe parallelism imbalances at the shot level. In our case, the choice has been a master-worker workflow where each frequency is an independent workflow itself, and pre-processing, post-processing and kernel are the main worker roles.

6 Conclusions

We have presented a set of general strategies towards reducing the cost associated with 3D elastic FWI. We have validated the most novel of the strategies, called DOC, in complex 2D scenarios and concluded that it does not only result in important computational savings but also allows for obtaining better model fits. This is due to reducing the footprint of the ground roll in the data and using gradients which are very local to the shot locations. The other acceleration techniques include grid size adaptation to frequency and composite shot lists. We remark that all of these strategies are independent of each other but can also be applied simultaneously for maximum gain. It is also important to notice that the improvements are orthogonal to any optimization carried out at the level of the numerical kernel used to solve the wave equation for forward and adjoint fields, but works better when the workflow is flexible enough to accommodate differently sized shots, such as in a master-worker pattern that separates individual shots in different

worker tasks. Our example shows that, rather than the theoretical FWI vs modelling 1000-fold increase in cost, our scheme achieves a mere tenfold increase in cost compared to that of modelling synthetic data when both modelling and inversion cover the same frequency range. In particular, we show positive 3D elastic FWI results with a $\times 53$ speedup compared to the time required for the inversion using a convential FWI workflow.

Acknowledgments Open access funding provided by Montanuniversität Leoben. The authors thank REPSOL for the permission to publish the present research and for funding through the AURORA project. J. Kormann also thankfully acknowledges the computer resources, technical expertise and assistance provided by the Barcelona Supercomputing Center - Centro Nacional de Supercomputactión together with the Spanish Supercomputing Network (RES) through grant FI-2014-2-0009. This project has received funding from the European Union's Horizon 2020, research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 644202. The research leading to these results has received funding from the European Union's Horizon 2020 Programme (2014–2020) and from the Brazilian Ministry of Science, Technology and Innovation through Rede Nacional de Pesquisa (RNP) under the HPC4E Project (www.hpc4e.eu), grant agreement no. 689772. We further want to thank the Editor Clint N. Dawson for his help, and Andreas Fichtner and an anonymous reviewer for their comments and suggestions to improve the manuscript.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Aminzadeh, F., Brac, J., Kunz, T.: 3-D salt and overthrust models: SEG/EAGE 3-D modeling series 1. SEG (1997)
2. Bekakos, M.P., Michail, H., Bowring, N., Michailidis, P.D., Cao, J., Milovanović, E.I., et al.: Supercomputing Research Advances. Nova Science Publishers, Inc (2008)
3. Ben-Hadj-Ali, H., Operto, S., Virieux, J.: An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. Geophysics **76**, R109–R124 (2011)
4. Boonyasiriwat, C., Schuster, G.T., Valasek, P., Cao, W.: Applications of multiscale waveform inversion to marine data using a flooding technique and dynamic early-arrival windows. Geophysics **75**, R129–R136 (2010a)
5. Boonyasiriwat, C., Zhan, G., Hadwiger, M., Srinivasan, M., Schuster, G.: Multisource reverse-time migration and full-waveform inversion on a gpgpu. In: 72nd EAGE Conference & Exhibition (2010b)
6. Borisov, D., Singh, S.: An efficiennt 3d elastic full waveform inversion of time-lapse seismic data using grid injection method, pp. 954–958. SEG Technical Program Expanded Abstracts (2013)
7. Brossier, R., Operto, S., Virieux, J.: Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. Geophysics **74**, WCC105–WCC118 (2009)
8. Brossier, R., Operto, S., Virieux, J.: Which data residual norm for robust elastic frequency-domain fullwaveform inversion. Geophysics **75**, R37–R46 (2010)
9. Bunks, C., Saleck, F.M., Zaleski, S., Chavent, G.: Multiscale seismic waveform inversion. Geophysics **60**, 1457–1473 (1995)
10. Butzer, S., Kurzmann, A., Bohlen, T.: 3D elastic full-waveform inversion of small-scale heterogeneities in transmission geometry. Geophys. Prospect. **61**, 1238–1251 (2013)
11. Caballero, D., Farres, A., Duran, A., Hanzich, M., Fernandez, S., Martorell, X.: Optimizing fully anisotropic elastic propagation on intel xeon phi coprocessors. In: Second EAGE Workshop on High Performance Computing for Upstream, EAGE (2015)
12. Causse, E., Mittet, R., Ursin, B.: Preconditioning of full-waveform inversion in viscoacoustic media. Geophysics **64**, 130–145 (1999)
13. Etienne, V., Tonellot, T., Thierry, P., Berthoumieux, V., Andreolli, C.: Speeding-up FWI by one order of magnitude. In: EAGE Workshop on High Performance Computing for Upstream, EAGE (2014)
14. Fichtner, A.: Full Seismic Waveform Modelling and Inversion. Springer Science & Business Media (2010)
15. Fichtner, A., Bunge, H.P., Igel, H.: The adjoint method in seismology 2. Applications : traveltimes and sensitivity functionals. Theory. Phys. Earth. Planet. Int. **157**, 86–104 (2006)
16. Fichtner, A., Kennet, B.N.L., Bunge, H.P., Igel, H.: Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency domain. Geophys. J. Int. **175**, 665–685 (2008)
17. Fichtner, A., Kennet, B.N.L., Igel, H., Bunge, H.P.: Full waveform tomography for radially anisotropic structure: new insights into present and past states of the Australasian upper mantle. Earth. Planet. Sci. Lett. **290**, 270–280 (2010)
18. Guasch, L., Warner, M., Nangoo, T., Morgan, J., Umpleby, A., Stekl, I., Shah, N.: Elastic 3d full-waveform inversion, 1–5 (2012). doi:10.1190/segam2012-1239.1
19. Guitton, A., Symes, W.: Robust inversion of seismic data using the Huber norm. Geophysics **68**, 1310–1319 (2003)
20. Guitton, A., Gboyega, A., Diaz, E.: Constrained full-waveform inversion by model reparameterization. Geophysics **77**, R117–R127 (2012)
21. Haberdar, H., Siddiqui, S., Feki, S.: Automatic performance tuning of parallel and accelerated seismic imaging kernels. In: EAGE Workshop on High Performance Computing for Upstream, EAGE (2014)
22. Hanzich, M., Rodriguez, J., Gutierrez, N., de la Puente, J., Cela, J.: Using HPC software frameworks for developing BSIT: a geophysical imaging tool. In: Proceedings of WCCM XI, ECCM V, ECFD VI III, pp. 2019–2030 (2014)
23. Hanzich, M., Rubio, F., Farrès, A.: Optimizing an elastic wave propagation by means of a roofline-based strategy on xeon processors. In: Enabling Software and Hardware Techologies Towards Exascale, 13th US National Congress on Computational Mechanics, San Diego (USA) (2015)
24. Jiménez Tejero, C., Dagnino, D., Sallarés, V., Ranero, C.: Comparative study of objective functions to overcome noise and bandwidth limitations in full waveform inversion. Geophys. J. Int. **203**, 632–645 (2015)
25. Kormann, J., Rodriguez, J.E., Gutierrez, N., de la Puente, J., Hanzich, M., Cela, J.: Using power-model based preconditioners for 3D acoustic full waveform inversion. SEG Technical Program Expanded Abstracts, 1105–1109 (2013)
26. Krebs, J.R., Anderson, J.E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., Lacasse, M.D.: Fast full-wavefield seismic inversion using encoded sources. Geophysics **74**(6), WCC177–WCC188 (2009)

27. van Leeuwen, T., Herrmann, F.: 3D frequency-domain inversion with controlled sloppiness. J. Sci. Comput. **36**(5), S192–S217 (2014). doi:10.1137/130918629
28. Levander, A.R.: Fourth-order finite difference P-SV seismograms. Geophysics **53**, 1425–1436 (1988)
29. Mora, P.: Nonlinear two dimensional elastic inversion of multioffset seismic data. Geophysics **52**, 1211–1228 (1987). doi:10.1190/1.1442384
30. Operto, S., Virieux, J., Amestoy, P., L Excellent, J.Y., Giraud, L., Ali, H.B.H.: 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: a feasibility study. Geophysics **72**(5), SM195–SM211 (2007)
31. Plessix, R.E.: A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. Geophys. J. Int. **167**, 495–503 (2006)
32. Pratt, R.G.: Seismic waveform inversion in the frequency domain, part 1: theory and verification in a physical scale model. Geophysics **64**, 888–901 (1999). doi:10.1190/1.1444597
33. de la Puente, J., Ferrer, M., Hanzich, M., Castillo, J.E., Cela, J.M.: Mimetic seismic wave modeling including topography on deformed staggered grids. Geophysics **79**(3), T125–T141 (2014)
34. Raknes, E.B., Arntsen, B., Weibull, W.: Three-dimensional elastic full waveform inversion using seismic data from the Sleipner area. Geophys. J. Int. **202**, 1877–1894 (2015)
35. Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., Dell'Aversana, P.: Multi-scale imaging of complex structures from multi-fold wide- aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt. Geophys. J. Int. **159**, 1032–1056 (2004)
36. Riyanti, C., Kononov, A., Erlangga, Y.A., Vuik, C., Oosterlee, C.W., Plessix, R.E., Mulder, W.A.: A parallel multigrid-based preconditioner for the 3d heterogeneous high-frequency Helmholtz equation. J. Comput. Phys. **224**(1), 431–448 (2007)
37. Rodriguez, J.E., Hanzich, M., de la Puente, J., Kormann, J., Gutierrez, N., Puzyrev, V., Cela, J.M.: Towards an efficient and reliable HPC software platform for 3D geophysical inversion. In: ICE Barcelona - AAPG/SEG International Conference & Exhibition, AAPG/SEG, Barcelona (Spain) (2015)
38. Roy, I.G., Sen, M.K., Torres-Verdin, C.: Full waveform seismic inversion using a distributed system of computers. Concurrency and Computation: Practice and Experience **17**(11), 1365–1385 (2005). doi:10.1002/cpe.897
39. Rubio, F., Hanzich, M., Farrés, A., de la Puente, J., María Cela, J.: Finite-difference staggered grids in GPUs for anisotropic elastic wave propagation simulation. Comput. Geosci. **70**, 181–189 (2014). doi:10.1016/j.cageo.2014.06.003
40. Sirgue, L., Pratt, R.G.: Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies. Geophysics **69**, 231–248 (2004)
41. Sochacki, J., Kubichek, R., George, J., Fletcher, W., Smithson, S.: Absorbing boundary conditions and surface waves. Geophysics **52**(1), 17–48 (1987). doi:10.1190/1.1442241
42. Son, W., Pyun, S., Shin, C., Kim, H.J.: Laplace-domain wave-equation modeling and full waveform inversion in 3d isotropic elastic media. J. Appl. Geophys. **105**, 120–132 (2014). doi:10.1016/j.jappgeo.2014.03.013
43. Tape, C., Liu, Q., Maggi, A., Tromp, J.: Seismic tomography of the southern California crust based on spectral-element and adjoint methods. Geophys. J. Int. **180**, 433–462 (2010). doi:10.1111/j.1365-246X.2009.04429.x
44. Tarantola, A.: Inversion of seismic reflection data in the acoustic approximation. Geophysics **8**, 1259–1266 (1984)
45. Tarantola, A.: Inversion problem theory and methods for model parameters estimation, sIAM edn (1986)
46. Vigh, D., Jiao, K., Watts, D., Sun, D.: Elastic full-waveform inversion application using multicomponent measurements of seismic data collection. Geophysics **79**(2), R63–R77 (2014)
47. Virieux, J.: P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method. Geophysics **51**, 889–901 (1986)
48. Virieux, J., Operto, S.: An overview of full-waveform inversion in exploration geophysics. Geophysics **74**, WCC127–WCC152 (2009)
49. Warner, M., Ratcliffe, A., Nangoo, T., Morgan, J., Umpleby, A., Shah, N., Vinje, V., tekl, I., Guasch, L., Win, C., Conroy, G., Bertrand, A.: Anisotropic 3D full-waveform inversion. Geophysics **78**, R59–R80 (2013)