Estimation of acoustic macro-models using a genetic full-waveform inversion: applications to the Marmousi model

Angelo Sajeva[1], Mattia Aleardi[1], Eusebio Stucchi[2], Nicola Bienati[3], Alfredo Mazzotti[1]

Right Running Head: Genetic FWI for acoustic macro-models

[1]University of Pisa, Department of Earth Science, Pisa, Italy. E-mail: angelo.sajeva@for.unipi.it; mattia.aleardi@for.unipi.it; alfredo.mazzotti@unipi.it.

[2]University of Milan, Department of Earth Science, Milano, Italy. E-mail: eusebio.stucchi@unimi.it.

[3]ENI, Upstream and Technical Services Division, San Donato Milanese, Milano, Italy. E-mail: nicola.bienati@eni.com.

ABSTRACT

We present a stochastic full-waveform inversion that uses genetic algorithms (GA FWI) to estimate acoustic macro-models of the P-wave velocity field. Stochastic methods such as GA severely suffer the curse of dimensionality, meaning that they require unaffordable computer resources for inverse problems with many unknowns and expensive forward modeling. To mitigate this issue, we propose a two-grid technique, that is, a coarse grid to represent the subsurface for the GA inversion and a finer grid for the forward modeling. We applied this procedure to invert synthetic acoustic data of the Marmousi model. We show three different tests. The first two tests use as prior information a velocity model derived from standard stacking velocity analysis and differ only for the parameterization of the coarse grid. Their comparison shows that a smart parameterization of the coarse grid may significantly improve the final result. The third test uses a linearly increasing 1D velocity model as prior information, a layer-stripping procedure, and a large number of model evaluations. All the three tests return velocity models that fairly reproduce the long-wavelength structures of the Marmousi. First-break cycle skipping related to the seismograms of the final GA-FWI models is significantly reduced compared to the one computed on the models used as prior information. Descent-based FWIs starting from final GA-FWI models yield velocity models with low and comparable model misfits and with an improved reconstruction of the structural details. The quality of the models obtained by GA FWI + descent-based FWI is benchmarked against the models obtained by descent-based FWI started from a smoothed version of the Marmousi and started directly from the prior information models. The results are promising and demonstrate the ability of the two-grid GA FWI to yield velocity models suitable as input to descent-based FWI.

INTRODUCTION

Full-waveform inversion is usually based on iterative local optimization methods and exploits the entire information of the seismogram to determine a high-resolution image of the subsurface. Accordingly, FWI is a very promising method, and, over the last decades, it has received growing expectations by industry and academy (e.g., Mora, 1988; Pratt, 1990; Virieux and Operto, 2009; Vigh, 2010; Morgan et al., 2013). Nevertheless, this method is limited by its local nature, i.e., it terminates in the nearest minimum of the misfit function, which may not coincide with the global minimum. To mitigate this issue, a smoother misfit function, which is obtained by progressively incorporating the higher frequencies of the data (multiscale technique, Bunks, 1995), is commonly used in the inversion, and the initial model is carefully chosen. Theoretically, a good starting model for FWI is a model which lies in the basin of attraction of the global minimum of the data misfit surface. In practice, good starting models are usually required to be smooth (Asnaashari et al., 2013) and able to match the main events of the observed seismogram with an error smaller than half of the wavelet wavelength to avoid the so-called cycle-skipping artifact (Beydoun and Tarantola, 1988). Several methods can be used to produce a starting model for FWI. In the oil and gas industry, the most common procedure is to apply reflection tomography followed by migration velocity analysis (see Woodward et al., 2008 for a review of the tomographic workflow). Other popular methods are first-arrival traveltime tomography (Nolet, 1987), stereotomography (Billette and Lambaré, 1998; Lambaré, 2008), Laplace domain and Laplace-Fourier domain inversions (Shin and Cha, 2008; Shin and Ha, 2008). Note that building a good enough initial model with these or other methods requires a significant amount of qualified human resources and a not negligible computing time. In fact, many processing iterations and adjustments, which include also traveltime picking, are usually needed to obtain the desired starting model.

A different approach to full-waveform inversion may consist in using a stochastic global method in place of a local iterative method. The advantage of using stochastic global methods is that they are able to jump out from local minima. Consequently, stochastic methods are less affected by the

choice of the starting point for the inversion (some stochastic methods, such as genetic algorithms, do not even require to define a starting model). The drawback of employing stochastic methods is that their computational cost increases exponentially with the number of unknowns. This is due to the search area of the stochastic inversion which grows exponentially with the number of unknowns. Such problem for stochastic methods is usually referred to as the curse of dimensionality (Bellman, 1957). Because FWI commonly inverts for more than hundreds of thousands of unknowns, the resulting computational time for a stochastic inversion appears to be unaffordable.

To attenuate this problem, we propose to represent the subsurface by means of a two-grid technique, i.e., we discretize the P-wave velocity model using a coarse grid in the stochastic inversion and a fine grid in the modeling of the wavefield propagation (Sajeva et al., 2014). The coarse grid permits to reduce the number of unknowns, thus mitigating the curse of dimensionality in the stochastic inversion, whereas the fine grid is used to perform the finite-difference forward modeling on an interpolated coarse grid model. Tuning the fine-grid step size allows us to propagate a certain frequency range independently of the coarse-grid parameterization. Because of the coarse grid employed in the inversion, the final model after GA FWI is at low resolution and thus it can not be directly compared with the result of FWI that uses local-iterative methods (descent-based FWI) which may solve for the finer details of the subsurface. Instead, using stochastic FWI, we aim to build a macro-model that contains the correct long-wavelength structure of the subsurface. Recently, other authors (Gao et al. 2014, Datta 2015) proposed to determine starting models for acoustic FWI using stochastic inversion methods combined with different strategies to reduce the number of unknowns.

Several stochastic methods exist and among them genetic algorithms (GA, Holland, 1975), simulated annealing (SA, Kirkpatrick et al., 1983) and neighborhood algorithm (NA, Sambridge, 1999a) are the most popular methods applied to geophysical problems (Sen and Stoffa, 2013). In this paper, we perform the stochastic inversion using a real-valued genetic algorithm. In real-valued

GAs the model parameters are represented as real numbers (Wright, 1991; Eshelman and Schaffer, 1993). Real-valued GAs are an evolution of the first implementations of GAs, which were binary-coded, i.e., whose model parameters were encoded with a binary scheme (De Jong, 1975; Goldberg, 1989). We have chosen a real-coded GA because it proved to perform better than NA and adaptive SA (Ingber, 1989) for high dimensional spaces, using a simple 1D elastic model and analytic functionals (Sajeva et al., 2014). Another advantage of GA over SA is that it can be easily parallelized, because GA inverts collectively a population of models.

In the first part of the paper, we describe the salient methodological aspects of our method and in particular: the workflow of the GA FWI, the two-grid technique, the numerical method used to solve the forward modeling, the misfit functions, and the layer-stripping approach. In the second part, we show three examples of GA FWI applied to the Marmousi model. The first two examples are identical apart from the parameterization of the coarse grid and are discussed together. The third example is more complex, it uses different inversion parameters and inverts for a higher-number of model parameters. This set of three examples verifies the feasibility of using the stochastic approach to FWI to invert for the large structures of the acoustic Marmousi model. As an application of the method, we tested the GA-FWI models as starting models for descent-based FWI. Finally, we quantify the cycle-skipping artifacts and we compare the results of the sequence GA FWI + descent-based FWI with those obtained using descent-based FWI only.

THE METHOD

The genetic-algorithm inversion

Genetic algorithms are stochastic heuristic search algorithms that mimic the natural selection and evolution processes of biological species to search for optimal solutions in a model space (Holland, 1975). Candidate solutions to the optimization problem are interpreted as individuals of a population which undergo the evolution process. Evolution is simulated via the operations of natural selection (performed according to a fitness function), recombination or mating, and mutation. First implementations of GA were binary coded and thereafter real-coded genetic algorithms were proposed to solve a number of problems (Deb and Kumar, 1995; Bessaou and Siarry, 2001; Wu et al, 2007). In 1991, Janikow and Michalewicz demonstrated that real-coded GA outperform binary-coded GA in many optimization problems, especially in high-dimensional nonlinear problems. Consequently, we used a real-coded GA in this work.

When setting up a genetic algorithm, choices related to genetic operators and tuning parameters must be made. Such choices are often problem dependent and they may be crucial to the success of the algorithm. We employed the GA operators described in Chipperfield et al. (1994) and here below we give a brief summary of the GA workflow we used

1. Randomly generate a set of models (individuals) within a defined search area. This set of models is called initial population.

2. Solve the forward model for each individual of the current population and evaluate the misfit between the observed and the simulated data.

3. Sort the individuals according to their data misfit and assign to each individual a rank-based fitness value (Baker, 1987); hence, select the subset of individuals apt for mating using stochastic universal sampling (Baker, 1987) as the selection method.

4. Pair the selected individuals and recombine their values to generate new individuals. The set of new individuals is called offspring.

5. Randomly mutate a fraction of the parameters of the offspring.

6. Allow for some parents to survive in the next generation by adding the most fit parents to the offspring. To this end and to maintain constant the desired number of models in the new generation, a subset of the offspring may be rejected. The resulting set of models constitutes the new generation.

7. Repeat steps from 2 to 6 until a stopping criterion is attained.

8. Extract the best-fitting model.

The search area is determined by the available prior information and should contain any valuable information of the subsurface. Several stopping criteria may be chosen, such as the achievement of a predetermined data misfit value, the consumption of the allocated resources (computer time, number of model evaluations), or if the data misfit does not decrease for a certain number of generations. At the end of the inversion, we simply extract the best-fitting model instead of evaluating the posterior probability distribution (PPD) of the models. This is because we mainly aim to estimate a single model, which can be used, for example, as starting model for descent-based FWI. In addition, it is demonstrated that genetic algorithms are not Monte Carlo methods, and, then, they produce a biased PPD (Sen and Stoffa, 1996). Thus, if desired, the PPD must be reconstructed from the set of computed models using different methods (Sambridge, 1999b; Hong and Sen, 2009; Mallick, 1999; Aleardi and Mazzotti, 2014). Such methods add a computational cost to the GA FWI. For these reasons, in the following we neglect to evaluate the PPD associated to the GA results.

The forward modeling

We perform the forward modeling by numerically solving the 2D acoustic wave equation:

$$\left(\frac{1}{v^2}\frac{\partial^2}{\partial t^2} - \nabla^2\right)P(x,z,t) = -\boldsymbol{\nabla}\cdot\boldsymbol{f}, \tag{1}$$

where P is the scalar pressure field, $\boldsymbol{f} = \boldsymbol{f}(x, z, t)$ is the external force field, $v = v(x, z)$ is the acoustic wave speed, and ∇^2 is the 2D Laplacian. Despite the fact that equation 1 is strictly valid only in fluid and gaseous media, it is frequently used in active-source full waveform inversion because its numerical solution is computationally inexpensive compared to the solution of the elastic wave equation (Fichtner, 2011).

For each source term $\boldsymbol{f}$, the predicted data consists in the solution of equation 1 at the receiver positions. To numerically solve the acoustic wave equation, we employ a simple explicit time-stepping scheme starting from a local perturbation of the wavefield at the source position. The scheme is implemented using a time-domain finite-difference method having an accuracy of second order in time and fourth order in space. We position absorbing cells at the boundaries of the model to avoid artifacts and to suppress surface-related multiples. In the following, we compactly denote with g the forward operator, and we explicit the dependence on the source position for the pressure field

$$P(x_r, x_s, t) = g(v(x, z), \boldsymbol{f}(x_s, t)), \qquad (2)$$

where x_r are the inline receiver positions ($z_r=0$) and x_s are the inline source positions ($z_s=0$).

The two-grid technique

Several methods have been proposed in the literature to reduce the number of model parameters that describe the subsurface by using geological information. For instance, Ma et al. (2012) used the depth-migrated image to sparsely sample the subsurface and derive an image-guided interpolator that permits to go back and forth between the sparse representation and the finely and uniformly sampled representation. In this section, we propose to use a simple low-resolution grid with large sized rectangular cells that is resampled to a finer grid prior to the forward modeling.

To achieve this goal, we introduce two parameterizations of the subsurface, represented by the matrices U and V. The matrix $U \in \mathbb{R}^{n \times m}$ contains the acoustic velocity values of the subsurface arranged in a coarse grid, at positions $x_1, x_2 = x_1 + dx, \ldots, x_n = x_1 + (n-1)dx$ and depths $z_1, z_2 = z_1 + dz, \ldots, z_n = z_1 + (m-1)dz$. We employ this coarse grid to parameterize the subsurface for the GA inversion. Note that this parameterization is completely defined by the horizontal and vertical step sizes (dx and dz). The values of dx and dz depend on prior information that we have on the subsurface. For instance, if, given a P-wave velocity prior model, we observe a vertical variability higher than the horizontal variability, then we will use $dz < dx$. Prior information may be provided by a number of different means, including standard stacking velocity analysis as done in two examples of the present work. The choice of dx and dz has a component of subjectivity, in fact the user may choose smaller step-sizes to increase resolution, or larger step-sizes to reduce the number of unknowns. Since the matrix U contains a limited number of unknowns, the exploration of the model space can be more thorough (in fact, the number of unknowns equals the dimensionality of the model space) and the statistical relevance of the inversion is higher.

Differently, we use a finer grid to represent the subsurface for the forward modeling. To obtain the fine-grid model $V \in \mathbb{R}^{N \times M}$ associated with U, where $N>n$ and $M>m$, we apply a bilinear interpolator (Proakis and Manolakis, 1996), I, to the coarse-grid model U, that is $V=I(U)$. Using V in the forward modeling permits to introduce constraints in the model, such as the position of the water bottom, and it allows us to simulate higher frequencies without numerical dispersion. We denote with $DX (=DZ)$ the side of the square cells of the fine-grid.

A variation of the simple coarse-grid implementation makes use of grid step sizes (dx,dz) that vary with depth, i.e., $(dx(z),dz(z))$. In this case, the grid is more properly represented by a vector $v \in \mathbb{R}^d$, where d is the number of grid points and the position of the grid points depends on the step-size functions $dx(z)$ and $dz(z)$. We use this coarse-grid implementation to enlarge the step sizes in depth

according to the loss of resolution with depth that occurs in seismic exploration in which sources and receivers are placed on the surface.

The misfit functions

Given a candidate solution U of the inverse problem, a simple data misfit function can be written as

$$\|d_{obs} - g(I(U))\|_i, \quad i=1,\ 2, \tag{3}$$

where $\|\cdot\|_{i=1,2}$ is the L1 or L2 norm, and d_{obs} is the observed acoustic field (in the case of synthetic tests $d_{obs} = g(v_{true}, f)$ with v_{true} indicating the true velocity model). Some preprocessing operations, such as low-pass/band-pass filtering, trace-by-trace normalization, muting of specific events, may be applied to the observed and/or to the predicted seismograms prior to the data misfit computation.

We also use a more sophisticated data misfit function composed by the weighted sum of the misfit components referred to the reflected and to the transmitted (or diving) wavefields:

$$\|(d_{obs})_{diving} - (g(V))_{diving}\|_i + \alpha\|(d_{obs})_{reflected} - (g(V))_{reflected}\|_i, \quad i=1,\ 2, \tag{4}$$

where $\alpha \in \mathbb{R}$ is the weighting parameter. The choice of α is dictated by the error component we would like to prevail and then be reduced by the inversion. For instance, when $\alpha=0$, the misfit function is given by the difference between the modeled and observed diving-waves, so that the optimization focuses on the reduction of this mismatch only. Otherwise if the value of α is chosen such as the error components due to the diving and to the reflected waves balance, that is

$$\alpha \approx \frac{\|(d_{obs})_{diving} - (g(V))_{diving}\|_i}{\|(d_{obs})_{reflected} - (g(V))_{reflected}\|_i}, \quad i=1,\ 2, \tag{5}$$

then the optimization operates to equally reduce both mismatches simultaneously.

The data misfit computed by means of equation 4 facilitates a layer-stripping approach to the inversion. It consists in solving the inverse problem starting from the shallower layers and progressively including the deeper layers (Yagle and Levy, 1983; Koster, 1991, among others). We use a layer-stripping procedure that consists of three steps:

1. First, estimate the shallow part of the model using the transmitted waves only to guide the inversion within a predetermined model search range (equation 4 with $\alpha=0$).

2. Next, keep fixed the upper part of the model retrieved from step 1 (or narrow the search range of the GA around the solution found in step 1) and estimate the deeper part of the model using a weighted sum of reflected and transmitted waves (equation 4 with $\alpha\neq0$).

3. Finally, as a possible step, readjust both the shallow and the deep parts of the model, using again a weighted sum of reflected and transmitted waves as misfit function.

The depth of the first layer may be inferred by a few try and error tests or by ray tracing over an approximate image of the subsurface. The first layer should embrace the area in which the majority of the turning rays travel into.

For what concerns the model misfit we use a simple average absolute error formula that is, given a fine-grid model V, its model misfit χ is

$$\chi = \frac{\|V_{ref}-V\|_1}{N\times M}, \qquad (6)$$

where V_{ref} is a reference model, and $N\times M$ is the number of nodes in the fine grid.

TESTS ON THE MARMOUSI MODEL

We discuss three tests in which the two-grid GA-FWI method is applied to the acoustic 2D Marmousi model (Versteeg, 1994, Figure 1a). Figure 1b shows the leftmost shot of the observed

data that were generated by means of the 2D finite-difference (FD) engine introduced in the theory section, using the derivative of the Ricker wavelet with 10 Hz central frequency as seismic wavelet. The three tests differ in terms of (1) the search ranges of the GA FWI, (2) the misfit function, (3) the parameterization of the model grids and (4) the simulated acquisition geometry. We can summarize the three tests as follows

1. **Velocity-analysis test with uniform *dx* and *dz* (test-V1):**

   o   search ranges: 1200 m/s wide and centered on a simple model determined via standard stacking velocity analysis and Dix formula.

   o   misfit function: equation 3 with $i=2$, i.e., L2 norm;

   o   the two grids:

      ▪   coarse grid: $n=17$, $m=15$, corresponding to $dx \approx 570$ m and $dz \approx 150$ m, and number of grid points $d=255$;

      ▪   fine grid: $N=384$, $M=92$, corresponding to $DX=DZ=24$ m and a maximum frequency of 12.5 Hz.

2. **Velocity-analysis test with *dx* and *dz* increasing with depth (test-V2):**

   o   search ranges: same as in test-V1;

   o   misfit function: L2 norm (same as in test-V1);

   o   the two grids:

      ▪   coarse grid: number of grid points $d=140$, *dx* and *dz* increase with depth according to the loss of both vertical and horizontal seismic resolution with depth;

      ▪   fine grid: same as test-V1.

3. **Linear-velocity-increase test with a layer-stripping approach (test-L):**

   o   search ranges: 2000 m/s wide in the first two steps of the layer stripping and centered on a 1D model with linearly increasing velocity with depth from 1500 m/s to 4000 m/s; 1000 m/s wide in the last step and centered on the resulting best model of the previous two steps.

   o   misfit function of equation 4 with $i=1$, i.e., L1 norm, and variable α in a three-step layer stripping procedure;

   o   the two grids:

       ▪   coarse grid: $n=96$, $m=23$ (i.e, $dx=dz=96$ m), and number of grid points $d=2208$;

       ▪   fine grid: $N=192$, $M=46$, corresponding to $DX=DZ=48$ m and maximum frequency of 6.25 Hz.

Note that the third example uses a quarter of the number of points of the first two examples for the fine grid. Accordingly, the fine-grid step size is doubled, the maximum modeled frequency is halved (to guarantee that no numerical dispersion occurs), and, above all, the FD computational cost is reduced. At the opposite, the number of points of the coarse grid in the third example is notably larger than that used for the first two examples. Because the number of points of the coarse grid equals the number of unknowns of the inversion, this test requires an increased number of model evaluations.

Finally, note that we have developed quite different model parameterizations for the coarse grid. In fact, the first example uses an asymmetric parameterization, i.e. $dx \neq dz$, the second example uses an asymmetric parameterization with step-sizes increasing with depth, and in the third example the cells are square ($dx=dz$, symmetric parameterization) and small.

The velocity-analysis tests (test-V1 and test-V2)

The prior information for these tests is derived by standard stacking velocity analysis and Dix formula that enabled us to compute the velocity model shown in Figure 2a. This model is resampled to the coarse grids of test-V1 and test-V2 (Figure 2b and 2c, respectively) and the resulting resampled models define the central values of the search areas for the GA inversions. In both tests, the search area is 1200 m/s wide for each model parameter.

The relevant GA parameters used in both test-V1 and test-V2 are shown in Table 1. The simulated acquisition geometry consists of 31 sources with spacing of 288 m. Each source hits 127 receivers, uniformly spaced at 72 m, and placed at fixed locations on the surface. For the computation of the inversions, we used five compute nodes of a linux cluster (each compute node is a 2 octa-core Intel Xeon CPU E5-2630 v3 at 2.40 GHz). We stopped both inversions after 100 generations, which correspond to 40500 evaluated models, and the runtime was approximately 10 days.

Figure 3 shows the evolution of the data misfit during the GA inversions for test-V1 (red) and test-V2 (green); dashed lines refer to the mean values of the data misfit per generation and continuous lines show the minimum values of the data misfit per generation. Note that for both inversions, the mean value curve approaches the minimum value curve. This indicates a loss of genetic diversity during the inversion (Reeves and Rowe, 2002). When the genetic diversity is low, the genetic optimization is less efficient (Reeves and Rowe, 2002), and further continuing the inversion may bring to minimal improvements at the expense of significant additional CPU time.

The best-fitting model for test-V1 is shown in Figure 4a. This model is interpolated to the fine grid (Figure 4b) and, hence, the reference Marmousi model of Figure 1a is subtracted from the interpolated model for comparison (Figure 4c). Analogously, for test-V2, the best-fitting model is shown in Figure 5a in the coarse grid, in Figure 5b in the fine grid, and Figure 5c shows the

14

difference between the estimated model of Figure 5b and the reference model. Note that the quality of both inversions worsens with depth (see Figures 4c and 5c).

To evaluate the quality of the final models, we calculated the average absolute error χ using equation 6 with respect to the reference Marmousi model of Figure 1a. For both test-V1 and test-V2 χ=254 m/s. Nevertheless, if we limit to the shallow half of the model, the average absolute errors are χ=188 m/s and χ=115 m/s for testV1 and testV2, respectively. This means that the shallower half of the model is better resolved in test-V2. In addition, test-V2 produces a smoother model with less anomalies (see Figure 5b) than the final model of test-V1 (Figure 4b). We believe that the smoothness of the final model and the more correct reconstruction of the shallow part are indications that the result of test-V2 is an improvement with respect to test-V1.

The inspection of the seismic data offers some insights into the seismic events that guide the inversion. Figure 6 shows a zoom of the leftmost shot of Figure 3 and displays the observed data (Figure 6a), the predicted data for test-V1 and test-V2 (Figures 6b and 6d, respectively), and the differences between predicted and observed data for test-V1 and test-V2 (Figures 6c and 6e, respectively).

Note that the predicted data do not show strong reflectors (Figures 6b and 6d), whereas the observed data contains some evident reflections (Figures 6a). This is because the true model presents sharp contrasts and a fine layering (see Figure 1a), which produce strong reflections, whereas the fine-grid final models for test-V1 and test-V2, which are interpolated models, have a smooth structure and thus they can not produce strong reflections. The red circles in Figure 6 highlight an area in which the improvement from test-V1 to test-V2 is evident (compare Figure 6c and Figure 6e). The final data misfit for test-V1 is 2.83 and for test-V2 is 2.80. This further proves the better performance of test-V2 with respect to test-V1.

The linear-velocity-increase test with a layer-stripping approach (test-L)

In this test, we applied the layer-stripping procedure described in the theory section, and we used the misfit function of equation 4. The data portion which includes the diving waves had been low-pass filtered (3 Hz) and subjected to a trace-by-trace normalization. The reflected waves portion had been low-pass filtered (6.25 Hz) and trace-by-trace normalized.

The simulated acquisition geometry employs 11 sources and 96 uniformly spaced receivers. Note that the number of sources is smaller than those used in the previous tests (11 instead of 31), and this reduces the computational cost of the forward modeling. Figures 7a and 7b show the diving and the reflected wavefield components for one of the observed shots which is approximately located at the center of the model (the 5th shot out of 11).

Some relevant GA parameters used in this test are shown in Table 2. Due to the increased number of unknowns, we set a larger number of models per generation (4800 instead of 300 as in the first two tests) resulting in approximately 3×10^6 evaluated models. Nine compute nodes (i7-3770 CPUs at 3.4 GHz) were used to concurrently compute the forward modelings. The overall run time of test-L was about 41 days, corresponding to approximately 1.2 s per forward model. Note that the GA inversion is highly parallelizable, and the runtime would have dropped drastically if a higher number of compute nodes were used as well as if the code had been fully optimized for parallel computing.

In the first step of the layer-stripping procedure, the shallower model parameters (down to 1 km of depth) are allowed to vary inside a search area 2000 m/s wide, centered on a simple 1D model where the acoustic velocity increases linearly with depth from 1500 m/s to 4000 m/s (see Figure 8a). Note that this model contains inaccurate prior information, and we will show in the following sections that the descent-based FWI started from this model fails to converge (see Figure 16a). Hence it is of interest to investigate for an inversion procedure which is different from descent-based FWI and more robust with respect to inaccurate starting models. In this step the α parameter

16

in the data misfit function of equation 5 is 0. We stop the first inversion after approximately 265000

evaluations, when the trend of the minimum value of the misfit per generation begins to flatten (see

Figure 8b). Figure 8c shows the best-fitting model after this first inversion. Note the improvement

in the shallow part of the model, especially in the center.

To preserve genetic continuity, we used the last generation of the first inversion to start the second

step of the layer stripping procedure where we focused on the deeper part of the model. We used a

search area 2000 m/s wide, and we stopped the inversion after approximately 1.4×10^6 model

evaluations. The parameter α was set to 1 such that the contributions of the error components

relative to the diving waves and to the reflected events, were comparable. However, since the

previous inversion step was focused on the optimization of the diving waves only, in the present

step the data misfit of the reflected waves was some two-three times higher than the data misfit of

the diving waves. Thus, it is the reflected events that mainly drive the inversion in the first

generations. The cyan curve in Figure 9a indicates the data misfit evolution of the second inversion

step. We allowed a third inversion step, performing approximately 1.4×10^6 model evaluations and

setting again α=1, in which the model parameters at all depths may vary in a search area ±500 m/s

wide centered on the best-fitting model of the second inversion. The blue curve of Figure 9a

indicates the data misfit evolution, Figure 9b shows the final best-fit model, and Figure 9c displays

the difference between the final best-fit model and the reference model.

The average absolute errors between the best-fitting model and the reference model are 369 m/s in

the first inversion, 251 m/s in the second inversion, and 212 m/s in the third inversion. If we take

only the shallower half of the model, the average absolute errors are 167 m/s, 129 m/s, and 99 m/s,

respectively. Comparing these values with those of test-V1 and test-V2 shows that the final model

of test-L is the one with the smallest model misfit.

As in the previous tests, the diving wavefield (Figure 10) is better recovered than the reflections (Figure 11). However, the match between the predicted reflected waves (Figure 11a) and the low-frequency part of the observed wavefield (Figure 11b) is improved with respect to the previous two tests, whereas, the high-frequency components dominate the difference panel (Figure 11c). Note that, in this test, the low-frequency components of the reflected wavefield can be correctly reconstructed thanks to the finer coarse grid employed in the inversion, which allows for a more detailed representation of the subsurface.

GA-FWI macro-models as starting models for descent-based FWI

Waiting for computers that can permit to run GA-based FWI on industrial scale projects at a convenient performance-to-cost ratio, a possible application of GA FWI is to produce a suitable starting model for descent-based FWI. To this end, we used test-V1, test-V2, and test-L final GA models as starting models for descent-based FWI. The descent-based FWI algorithm that we use employs the steepest-descent method and a multiscale approach (we performed 5 iterations at 4, 5, 6, 8, and 10 Hz). Both inversions and forward problems are formulated in time domain. The starting models for descent-based FWI (i.e. the final models retrieved by GA FWI) for test-V1, test-V2, and test-L, are shown in Figures 12a, 12d, and 12g, respectively, while in Figures 12b, 12e, and 12h the final models after descent-based FWI are illustrated. Figures 12c,f, and i display the differences between the final models after descent-based FWI and the reference model of Figure 1a.

Note that in all the three cases the structure features of Marmousi are fairly reconstructed, especially down to depths of 1 km -1.2 km. With respect to the GA FWI results, the average absolute error obtained with the sequence GA FWI + descent-based FWI decreases to 158 m/s, 150 m/s, and 120 m/s for test-V1, test-V2, and test-L, respectively.

As a benchmark test, we performed a descent-based inversion that starts from a smoothed version of the Marmousi (Figure 13a). This model contains the correct long wavelengths of the true model,

18

and, therefore, it is a suitable starting model for FWI (Virieux and Operto, 2009). Figure 13b shows

the final model after descent-based FWI and Figure 13c displays the difference between the

reference model and the model of Figure 13b. Note that this final model (Figure 13b) correctly

reproduces the intermediate/fine scale of the Marmousi model, and, as expected, it has an average

absolute error (80 m/s) smaller than those of the three GA FWI + descent-based FWI tests.

We prove the benefit of performing GA FWI prior to the descent-based FWI also by analyzing the

cycle skips associated to the first breaks of the observed data and of the data modeled on the

starting models of descent based FWI, for each of the three tests. We restrict the cycle-skip analysis

to the first breaks because these transmission events are crucial to reconstruct the large and

intermediate wavelengths of the model in descent-based FWI (Mora, 1989). To compute the amount

of cycle skipping associated with a given starting model we use the method proposed by Shah et al.

(2012). Their method analyzes the first arrivals of the data, and computes the phase differences

between the observed data and the data predicted on the starting model, at a given frequency, for

each source-receiver pair. In this residual phase domain, the cycle skipping occurs when there is a

2π jump. Figure 14 shows the residual phase associated with the three different models representing

the central models of the search range for GA FWI (Figures 14a, b, and c) and the three best fitting

models found by GA FWI (Figures 14d, e, and f), in the three tests V1, V2, and L. Note that, for

each of the three tests, the final models obtained by GA FWI produce significantly lower cycle

skips than the models corresponding to the prior information, that is the models corresponding to

the central values of model search ranges. Therefore, it is evident the advantage to descent-based

FWI of starting from a model estimated by GA FWI.

The benefit of performing GA FWI prior to descent-based FWI is further proved by comparing each

of the three tests with counterpart descent-based FWIs that start directly from the velocity models

corresponding to the centers of the search ranges for the GA FWIs (Figures 2b, 2c, and 8a),

interpolated in the fine grid. The bar diagram of Figure 15 shows the average absolute error of the

final models from all these six inversions and illustrates that the descent-based only approach returns worse results than the inversions that also make use of GA.

Note that test-L is the one that most benefits from the introduction of the GA inversion. Given the highly inaccurate prior information (see Figure 8a), the descent-based FWI falls, as expected, into a local minimum (Figure 16a). Differently, the GA FWI starting from the same prior information returns a more accurate image of the subsurface, even though at low resolution, which is shown in Figure 16b. We suggest therefore to use this GA model as starting model for descent-based FWI. The final result of the combined GA FWI + descent-based FWI is shown in Figure 16c. Finally, the average absolute errors of the three models of Figure 16 are monotonically decreasing from left to right: 416 m/s, 212 m/s, and 120 m/s.

CONCLUSIONS

We present a genetic algorithm full-waveform inversion (GA FWI) that uses a two-grid technique to estimate a low-resolution acoustic model (macro-model) of the subsurface. The two-grid technique allows us to both reduce the number of unknowns in the inversion; thus mitigating the so-called curse of dimensionality, and to increase the frequency range in the acoustic propagation.

We demonstrate the feasibility of the proposed GA FWI on three different examples from the Marmousi model, using noise-free synthetic data and known source wavelet, a quite convenient situation that clears the ground from other issues than stochastic FWI. The first two examples (test-V1 and test-V2) of GA FWI allow the stochastic algorithm to explore a broad search area centered on a velocity model derived from standard stacking velocity analysis. They share the same input data (31 shots), GA parameters setting, and a data misfit functional that takes into account the whole recorded wavefield. They differ only for the parameterization of the coarse grid as test-V1

employs constant-size grid cells, while test-V2 uses a grid with cell dimensions increasing with depth. In the third example (test-L), the GA FWI starts from a search area centered on a 1D velocity model with velocity linearly increasing with depth, it makes use of a layer-stripping procedure and of a data-misfit equation that is composed of two terms, one associated to the diving waves and the other related to the reflected wavefield. It also differs from the previous two examples because a larger number of model evaluations are performed due to a higher number of unknowns (coarse grid points). This in turn requires different GA parameters, a larger fine-grid step size and a reduced number of input data (11 shots) to allow for a sustainable computational effort.

Comparing the results of the first two tests, it turns out that a smart implementation of the coarse grid, with cell size that increases with depth (test-V2), improves the final result (compare the final model shown in Figure 4b with the one in Figure 5b). This is due to the fact that in the coarse grid more cells are placed on the shallow part of the model, where the seismic illumination is higher, while the deeper part of the model, with poorer illumination, is discretized employing much larger cell sizes. Note that the better results of test-V2 are achieved making use of a total number of grid points (the unknowns in the inversion) that is less than in test-V1 (140 vs 255). In spite of the very simple 1D velocity field that defined the initial model parameter search, the GA FWI performed in the third example (test-L) is the one that returns the best result. This highlights the importance of letting the GA perform a large exploration of the model space (3×10^6 models were computed for this test) and of choosing an efficient inversion strategy, such as the layer-stripping approach, for the success of the stochastic inversion.

Note that in all cases, the major differences between the estimated GA FWI models and the reference model (Figures 4c, 5c, and Figure 9c) are concentrated at the edges and at the base of the models, where the seismic illumination is poor and any purely data driven inversion cannot do much. However, all the estimated macro-models fairly reproduce the long-wavelength structures of the Marmousi model.

Starting descent-based FWI from the three best models estimated via GA FWI, we found that the final models improve the reconstruction of the finer details (see Figures 12b, 12e, and 12h). Moreover, the final models resulting from the sequence GA FWI + descent-based FWI turn out to be quite similar to the model obtained starting the descent-based FWI from a smooth version of the true Marmousi (compare Figures 12b, 12e, and 12h with Figure 13b). Considering the average absolute model error for the final models of the sequence GA FWI + descent based FWI, we have 158 m/s, 150 m/s, 120 m/s for test-V1, test-V2, and test-L, respectively, while we register 80 m/s for the final descent based FWI model obtained starting from the smoothed version of the Marmousi. In fact, the models resulting from GA FWI fulfill an important prerequisite for the success of descent-based FWI, that is the seismograms computed on them show minor first-breaks cycle skips with respect to the observed seismic data. In particular, we verified that GA FWI macro-models give rise to seismograms with significant increases of first-breaks matching compared to the seismograms of the prior models (corresponding to the central values of the GA FWI search ranges). Descent-based FWI started straight on the prior models brings, as expected, to considerably worse results.

We interpret the increased likeness among the three final models after the sequence GA FWI + descent-based FWI, their decreased model misfits and their good similarity with the model estimated from the smoothed Marmousi model as positive indications of the ability of the proposed two-grid GA FWI to estimate reliable velocity macro-models of the subsurface, suitable as input for further descent-based inversions.

22

comments. The analysis and visualisation of the seismic data was performed by means of the

Promax 3D Software of Landmark Graphics Corporation.

REFERENCES

Aleardi, M., and A. Mazzotti, 2014, 1D Elastic FWI and Uncertainty Estimation by Means of a Hybrid Genetic Algorithm-Gibbs Sampler Approach: 76th Conference and Exhibition, EAGE, Extended Abstracts, DOI: 10.3997/2214-4609.20140858.

.

Asnaashari, A., R. Brossier, S. Garambois, F. Audebert, P. Thore, and J. Virieux, 2013, Regularized seismic full waveform inversion with prior model information: GEOPHYSICS, **78**, no. 2, R25-R36.

Bäck, T., and F. Hoffmeister, 1991, Extended selection mechanisms in genetic algorithms: Morgan Kaufmann publisher.

Baker, J. E., 1987, Reducing bias and inefficiency in the selection algorithm: 2nd International Conference on Genetic Algorithms, 14-21.

Bellman, R. E., 1957, Dynamic programming: Princeton University Press.

Bessaou, M. and Siarry, P., 2001. A genetic algorithm with real-value coding to optimize multimodal continuous functions: Structural and multidisciplinary optimization, **23**, no. 1, 63-74.

Beydoun, W. B., and A. Tarantola, 1988, First Born and Rytov approximations: Modeling and inversion conditions in a canonical example: The Journal of the Acoustical Society of America, **83**, no. 3, 1045-1055.

Billette, F., and G. Lambaré, 1998, Velocity macro-model estimation from seismic reflection data by stereotomography: Geophysical Journal International, **135**, 671–680.

Blickle, T. and L. Thiele, 1995, A comparison of selection schemes used in genetic algorithms, ftp://129.132.2.212/pub/publications/TIK-Report11.pdf, accessed 23 February 2015.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: GEOPHYSICS, **60**, no. 5, 1457-1473.

Chipperfield, A. J., P. J. Fleming, and H. Pohlheim, 1994, A Genetic Algorithm Toolbox for MATLAB, Proceedings of the International Conference on Systems Engineering, 200-207.

De Jong, K.A., 1975, An analysis of the behavior of a class of genetic adaptive systems: Ph.D. thesis, University Microfilms No. 76 9381, University of Michigan, Dissertation Abstracts International, **36**, no. 10, 5140.

Deb, K., and Kumar, A., 1995, Real-coded genetic algorithms with simulated binary crossover: studies on multimodal and multiobjective problems: Complex Systems, **9**, 431-454.

Datta, D., 2015, Estimating Starting Models for Full Waveform Inversion Using a Global Optimization Method: 77th Conference and Exhibition, EAGE, Extended Abstracts, DOI: 10.3997/2214-4609.201412590.

Eshelman, L.J., and Schaffer, J.D., 1993. Real-Coded Genetic Algorithms and Interval-Schemata: Foundations of Genetic Algorithms, **2**, 187-202.

Fichtner, A., 2011, Full seismic waveform modelling and inversion: Springer.

Gao, Z., J. Gao, P. Zhibin, and X. Zhang, 2014, Building an initial model for full waveform inversion using a global optimization scheme: 85th Annual Meeting, SEG, Expanded Abstracts, 1136-1141.

Goldberg, D. E., 1989, Genetic algorithms in search, optimization and machine learning: Kluwer Academic Publishers.

Holland, J. H., 1975, Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence: The University of Michigan Press.

Hong, T., and M. K. Sen, 2009, A new MCMC algorithm for seismic waveform inversion and corresponding uncertainty analysis: Geophysical Journal International, **177**, no. 1, 14-32.

Kirkpatrick, S., C. D. Gelatt, and M. Vecchi , 1983, Optimization by simulated annealing: Science, **220**, no. 4598, 671-680.

Koster, J.K., 1991, A direct layer-stripping approach to the inversion of marine seismic data: Ph.D. thesis, Delft University of Technology.

Ingber, L., 1989, Very fast simulated re-annealing: Mathematical and computer modelling, **12**, no. 8, 967-973.

Janikow, C. Z. and Z. Michalewicz, 1991, An experimental comparison of binary and floating point representations in genetic algorithms: 4th International Conference on Genetic Algorithms, 31-36.

Lambare, G.,2008, Stereotomography: GEOPHYSICS, **73**, no. 5, VE25-VE34.

Ma, Y., D. Hale, B. Gong, and Z. Meng, 2012, Image-guided sparse-model full waveform inversion: GEOPHYSICS, **77**, no. 4, R189-R198.

Mallick, S., 1999, Some practical aspects of prestack waveform inversion using a genetic algorithm: An example from the east Texas Woodbine gas sand: GEOPHYSICS, **64**, no. 2, 326-336.

Mitchell, M., 1996, An introduction to genetic algorithms: Massachusetts Institute of Technology Press.

Mora, P., 1988, Elastic wave-field inversion of reflection and transmission data: GEOPHYSICS, **53**, no. 6, 750–759.

Mora, P., 1989, Inversion= migration + tomography: GEOPHYSICS, **54**, no. 12, 1575-1586.

Morgan, J., M. Warner, R. Bell, J. Ashley, D. Barnes, R. Little, K. Roele, and C. Jones, 2013, Next-generation seismic experiments: wide-angle, multi-azimuth, three-dimensional, full-waveform inversion: Geophysical Journal International, **195**, no. 3, 1657-1678.

Nolet, G., 1987, Seismic tomography with applications in global seismology and exploration geophysics: D. Reidel Publishing Company.

Pratt, R. G., and M. H. Worthington, 1990, Inverse theory applied to multi-source cross-hole tomography. Part 1: acoustic wave-equation method: Geophysical Prospecting, **38**, no. 3, 287-310.

Proakis, J.G., and D. G. Manolakis, 1996, Digital Signal Processing, Principles, Algorithms, and Applications: Pentice Hall.

Reeves, C. R., and J. E. Rowe, 2002, Genetic Algorithms - Principles and Perspectives (A guide to GA Theory): Kluwer Academic Publisher.

Sajeva, A., M. Aleardi, E. Stucchi, A. Mazzotti, and B. Galuzzi, 2014, Comparison of stochastic optimization methods on two analytic objective functions and on a 1D elastic FWI: 76th Conference and Exhibition, EAGE, Extended Abstracts, P0115.

Sajeva, A., N. Bienati. M. Aleardi, E. Stucchi, and A. Mazzotti, 2014, Estimating velocity macro-models using stochastic full-waveform inversion: 85th Annual Meeting, SEG, Expanded Abstracts, 1227-1231.

Sambridge, M., 1999, Geophysical inversion with a neighbourhood algorithm—I. Searching a parameter space: Geophysical Journal International, **138**, no. 2, 479-494.

Sambridge, M., 1999, Geophysical inversion with a neighbourhood algorithm—II. Appraising the ensemble: Geophysical Journal International, **138**, no. 2, 727-746.

Sen, M. K., and P. L. Stoffa, 1996, Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion: Geophysical Prospecting, **44**, no. 2, 313-350.

Sen, M. K., and P. L. Stoffa, 2013, Global optimization methods in geophysical inversion: Cambridge University Press.

Shah, N., M. Warner, T. Nangoo, A. Umpleby, I. Stekl, J. Morgan, and L. Guasch, 2012, Quality assured full-waveform inversion: Ensuring starting model adequacy: 83rd Annual Meeting, SEG, Expanded Abstract, 1-5.

Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, 922–931.

Shin, C. and W. Ha, 2008, A comparison between the behavior of objective functions for waveform inversion in the frequency and Laplace domains: GEOPHYSICS, **73**, no. 5, VE119-VE133.

Versteeg, R., 1994, The Marmousi experience: Velocity model determination on a synthetic complex data set: The Leading Edge, **13**, no. 9, 927-936.

Vigh, D., B. Starr, J. Kapoor, and H. Li, 2010, 3D full waveform inversion on a Gulf of Mexico WAZ data set: 80th Annual Meeting, SEG, Expanded Abstracts, 957-961.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: GEOPHYSICS, **74**, no. 6, WCC1-WCC26.

Woodward, M. J., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: GEOPHYSICS, **73**, no. 5, VE5-VE11.

Wright, A. H., 1991, Genetic Algorithms for Real Parameter Optimization: Foundations of Genetic Algorithms, **1**, 205-218.

Wu, C.H., Tzeng, G.H., Goo, Y.J. and Fang, W.C., 2007, A real-valued genetic algorithm to optimize the parameters of support vector machine for predicting bankruptcy: Expert systems with applications, **32**, no. 2, 397-408.

Yagle, A. E., and B. C. Levy, 1984, Application of the Schur algorithm to the inverse problem for a layered acoustic medium: The Journal of the Acoustical Society of America, **76**, no. 1, 301-308.

LIST OF CAPTIONS

Figure 1: (a) The acoustic Marmousi model. We use it to generate the observed synthetic data and as reference model for the computation of the average absolute model error. (b) The leftmost shot gather of the observed data set. Both reflected waves and transmitted/diving waves can be identified in the seismogram.

Figure 2: (a) The model determined via standard stacking velocity-analysis and Dix equation. This model is resampled to the coarse grid of test-V1 (b) and to the coarse-grid of test-V2 (c). Models of panels (b) and (c) are used as the center of the search areas for test-V1 and test-V2, respectively.

Figure 3: The evolution of the data misfit for test-V1 (red) and test-V2 (green) from generation 20 to 100. Continuous and dashed lines denote the minimum and the mean values of the data misfit per generation, respectively. Note that the minimum value of the data misfit for test-V2 is smaller than that of test-V1 and that the mean value curve approaches the minimum value curve in both tests.

Figure 4: Test-V1: (a) best model in coarse grid at the end of the 100th generation; (b) the same model in fine grid; (c) difference between (b) and the reference model (Figure 1a). The average absolute error between (b) and the reference model is 254 m/s for the entire model, and 188 m/s if we take only the shallower half of the model.

Figure 5: Test-V2: (a) best model in coarse grid at the end of the 100th generation; (b) the same model in fine grid; (c) difference between the best model in fine grid and the reference model (Figure 1a). The average absolute error between (b) and the reference model is 254 m/s and 115 m/s if we take only the shallower half of the model.

Figure 6: (a) Observed data; (b) data predicted in test-V1; (c) difference between (b) and (a); (d) data predicted in test-V2; (e) difference between (d) and (a). All panels are referred to the leftmost shot and share the same gray scale. The red circles delimit an area in which test-V2 clearly matches the observed data better than test-V1.

Figure 7: One of the shots of the observed data in test-L split into two parts: (a) the transmitted/diving wavefield and (b) the reflected waves down to 4.2 s. Both data are low-pass filtered and trace-by-trace normalized.

Figure 8: The first step of the layer-stripping procedure of test-L: (a) the model used to center the search domain; (b) the evolution of the minimum value of the data misfit per generation; (c) the best-fitting model. Note that we updated only the shallow part of the model down to approximately 1 km of depth.

Figure 9: The second and third steps of the layer stripping procedure of test-L: (a) the evolution of the minimum value of the data misfit per generation for the two steps; (b) the best-fitting model after the third step; (c) difference between (b) and the reference model (Figure 1a). The average

absolute error of equation 6 computed with respect to the reference model decreases from 222 m/s

at the end of the second step to 211 m/s at the end of the third step.

Figure 10: Result of the inversion on the diving waves for the 5th shot-gather: (a) final predicted

data after the third inversion step of test-L, (b) observed data, (c) difference between the predicted

and observed data. Note the good fit between the predicted and the observed seismogram attested

by the low residuals.

Figure 11: Result of the inversion on the reflected waves for the 5th shot-gather: (a) final predicted

data after the third inversion step of test-L, (b) observed data, (c) difference between the predicted

and observed data. Note that the low-frequency content of the observed wavefield is correctly

recovered and the mismatch is mainly due to the high-frequency content.

Figure 12: Results of the sequence GA FWI + descent-based FWI. First row shows the outcome

referred to test-V1: (a) starting model for descent based FWI, i.e., the best-fitting model after GA

FWI, (b) final model after descent-based FWI, (c) difference between (b) and the reference model

of Figure 1a. In analogy to the first row, the second and third rows show the outcomes referred to

test-V2 and test-L, respectively.

Figure 13: Benchmark test: (a) smoothed version of the Marmousi used as starting model for

descent-based FWI, (b) final model after descent-based FWI, (c) difference between the descent-

based FWI model and the reference model of Figure 1a. The average absolute error of (b) is 80 m/s.

Figure 14: Phase residual panels (Shah et al., 2012) referred to different starting models: (a) the

first velocity analysis model (the model of Figure 2b after fine interpolation), (b) the second

velocity analysis model (the model of Figure 2c after fine interpolation), (c) the linearly increasing

with depth model (the model of Figure 8a in fine grid). Phase residual panels for (d) the best-fitting

model of test-V1; (e) the best-fitting model of test-V2, (f) the best-fitting model of test-L.

Figure 15: Bar diagram that shows, in blue, the average absolute error after descent-based FWIs started from models that represent the prior information used in the three tests (V1, V2, and L), and, in red, the average absolute error of the final models after GA FWI + descent-based FWI for the three tests (V1, V2, and L). Note that performing a GA FWI prior to descent-based FWI improves the final model misfit for each of the three tests.

Figure 16: Comparison between descent-based FWI, GA FWI, and the sequence GA FWI + descent-based FWI given the same prior information. (a) final model after descent-based FWI using model of Figure 8a as starting model, (b) final model after GA FWI using model of Figure 8a to center the ranges of the stochastic inversion, (c) final model after descent-based FWI using the GA FWI model in (b) as starting model.

a)

b)

Figure 1: (a) The acoustic Marmousi model. We use it to generate the observed synthetic data and as reference model for the computation of the average absolute model error. (b) The leftmost shot gather of the observed data set. Both reflected waves and transmitted/diving waves can be identified in the seismogram.

116x161mm (300 x 300 DPI)
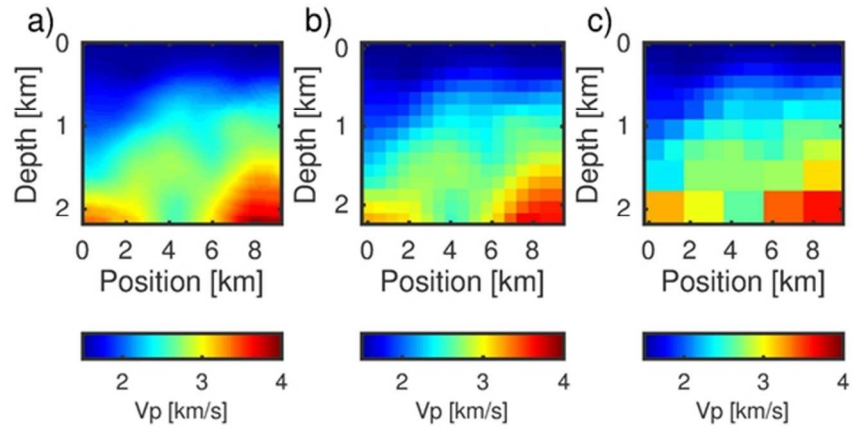
Figure 2: (a) The model determined via standard stacking velocity-analysis and Dix equation. This model is resampled to the coarse grid of test-V1 (b) and to the coarse-grid of test-V2 (c). Models of panels (b) and (c) are used as the center of the search areas for test-V1 and test-V2, respectively.
58x31mm (300 x 300 DPI)
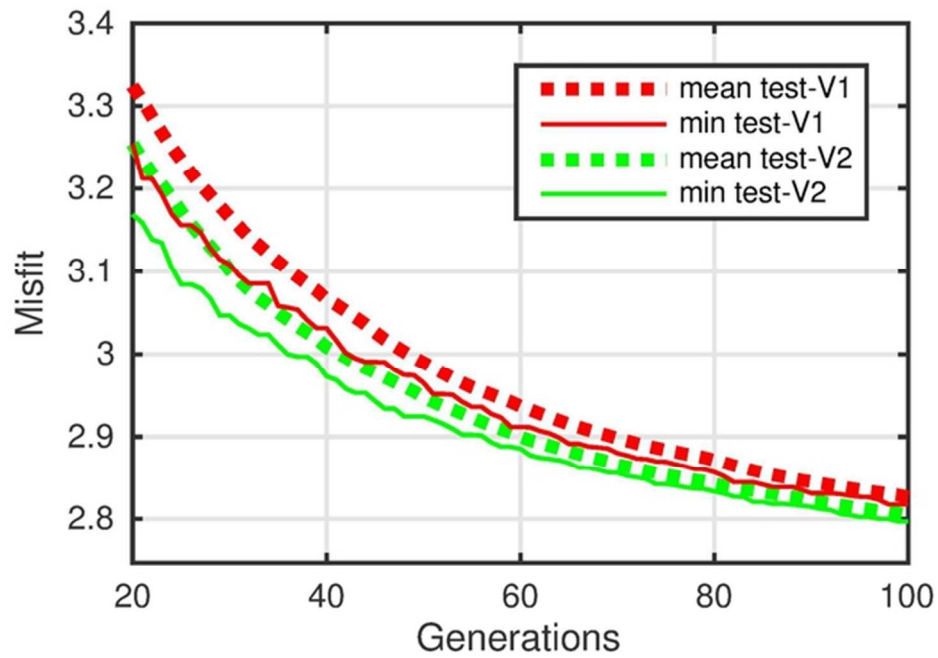
Figure 3: The evolution of the data misfit for test-V1 (red) and test-V2 (green) from generation 20 to 100. Continuous and dashed lines denote the minimum and the mean values of the data misfit per generation, respectively. Note that the minimum value of the data misfit for test-V2 is smaller than that of test-V1 and that the mean value curve approaches the minimum value curve in both tests.

58x40mm (300 x 300 DPI)
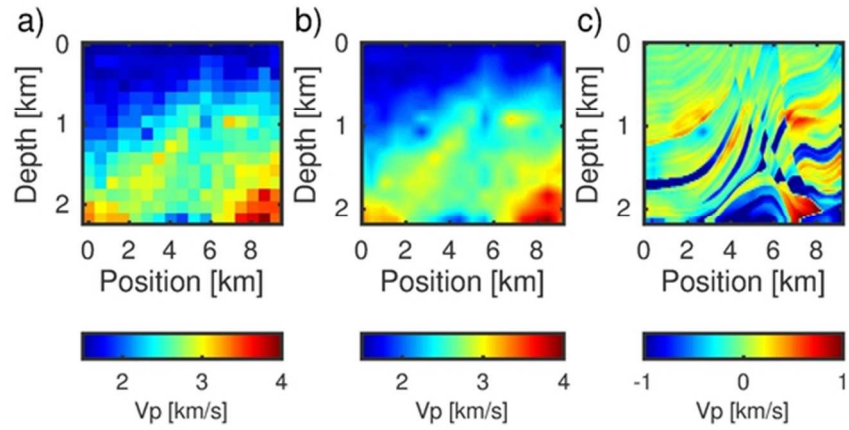
Figure 4: Test-V1: (a) best model in coarse grid at the end of the 100th generation; (b) the same model in fine grid; (c) difference between (b) and the reference model (Figure 1a). The average absolute error between (b) and the reference model is 254 m/s for the entire model, and 188 m/s if we take only the shallower half of the model.

58x31mm (300 x 300 DPI)
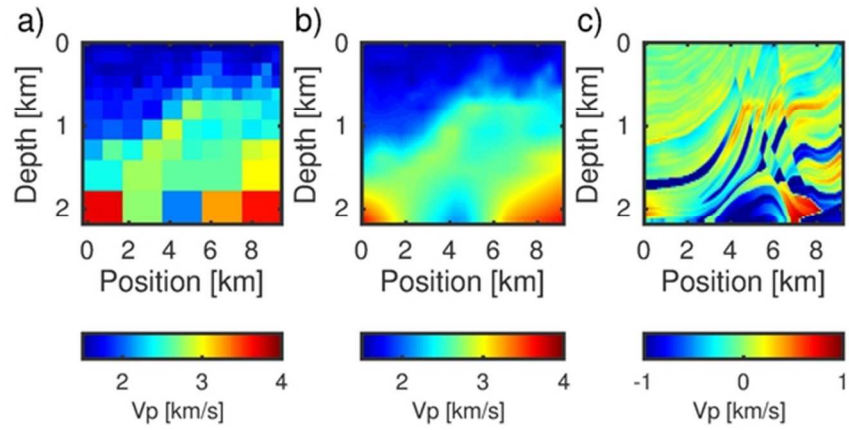
Figure 5: Test-V2: (a) best model in coarse grid at the end of the 100th generation; (b) the same model in fine grid; (c) difference between the best model in fine grid and the reference model (Figure 1a). The average absolute error between (b) and the reference model is 254 m/s and 115 m/s if we take only the shallower half of the model.

58x31mm (300 x 300 DPI)
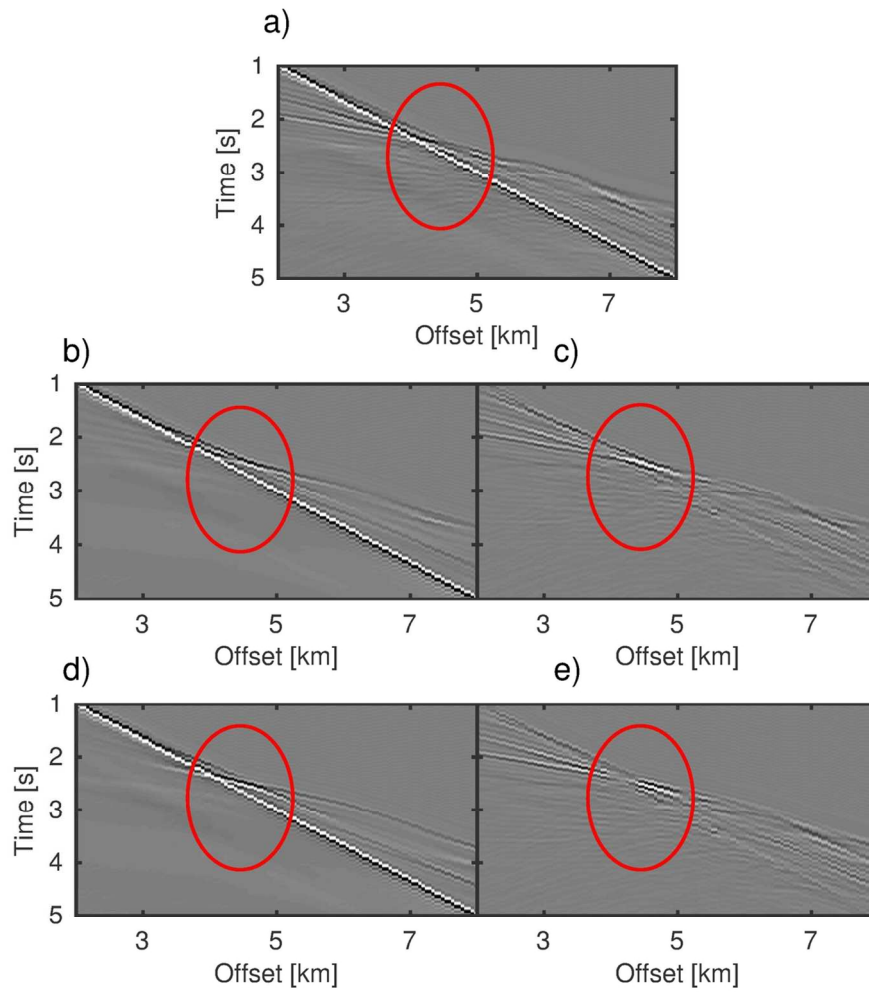
Figure 6: (a) Observed data; (b) data predicted in test-V1; (c) difference between (b) and (a); (d) data predicted in test-V2; (e) difference between (d) and (a). All panels are referred to the leftmost shot and share the same gray scale. The red circles delimit an area in which test-V2 clearly matches the observed data better than test-V1.

116x123mm (300 x 300 DPI)
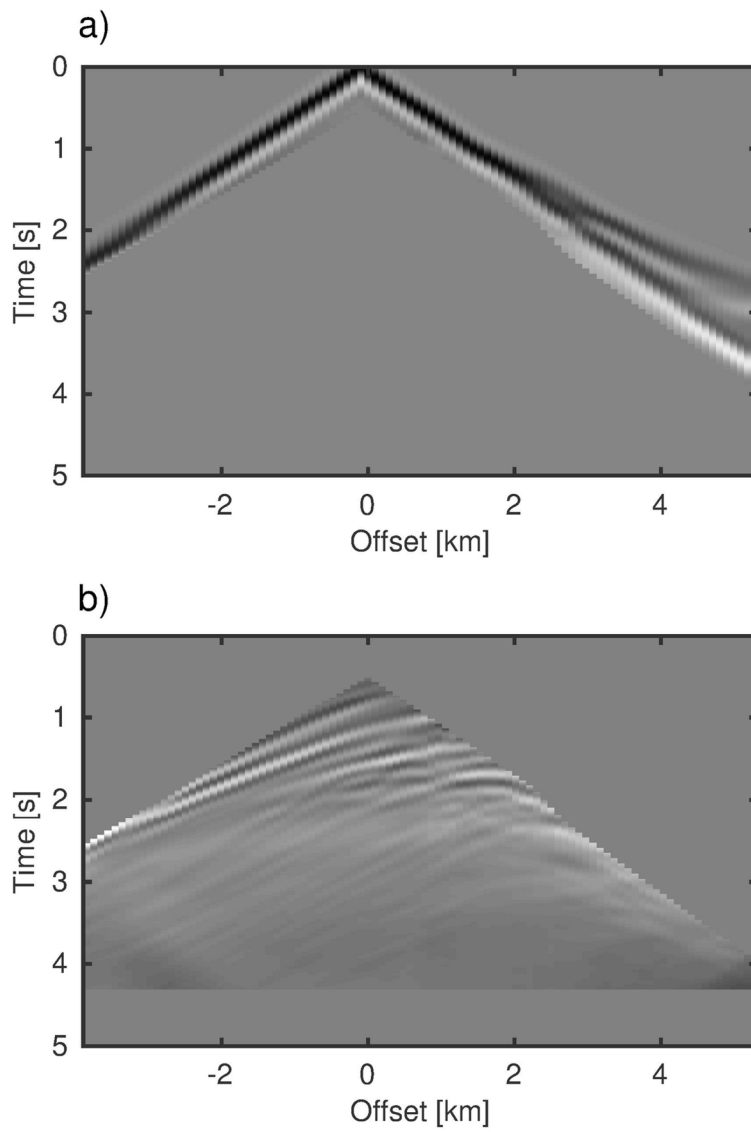
Figure 7: One of the shots of the observed data in test-L split into two parts: (a) the transmitted/diving wavefield and (b) the reflected waves down to 4.2 s. Both data are low-pass filtered and trace-by-trace normalized.

116x161mm (300 x 300 DPI)
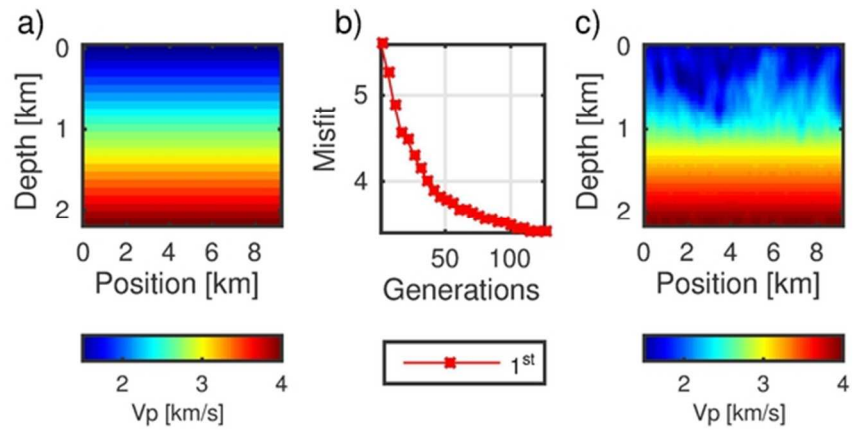
Figure 8: The first step of the layer-stripping procedure of test-L: (a) the model used to center the search domain; (b) the evolution of the minimum value of the data misfit per generation; (c) the best-fitting model. Note that we updated only the shallow part of the model down to approximately 1 km of depth.

58x31mm (300 x 300 DPI)
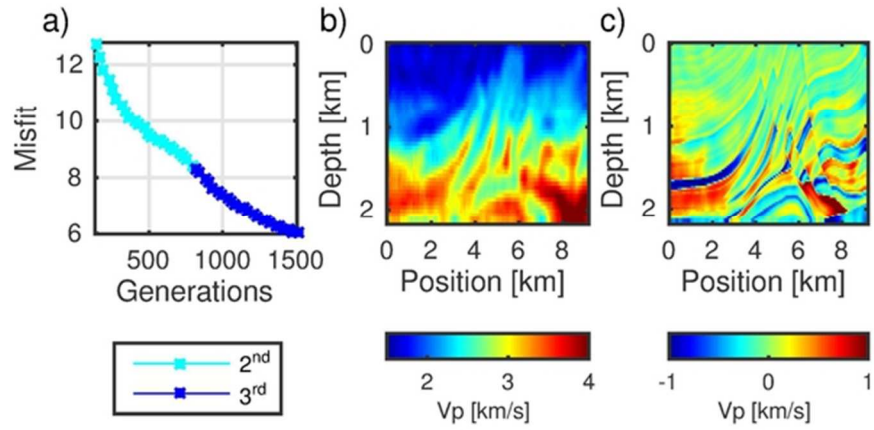
Figure 9: The second and third steps of the layer stripping procedure of test-L: (a) the evolution of the minimum value of the data misfit per generation for the two steps; (b) the best-fitting model after the third step; (c) difference between (b) and the reference model (Figure 1a). The average absolute error of equation 6 computed with respect to the reference model decreases from 222 m/s at the end of the second step to 211 m/s at the end of the third step.

58x31mm (300 x 300 DPI)
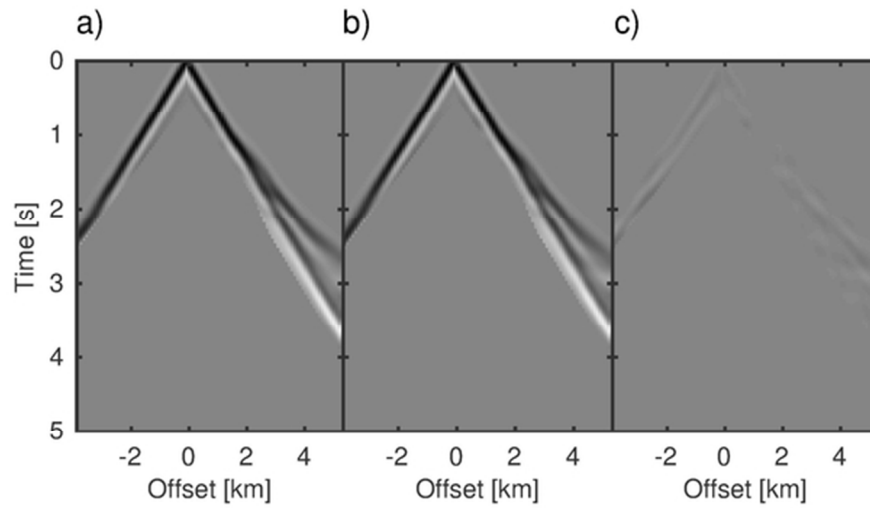
Figure 10: Result of the inversion on the diving waves for the 5th shot-gather: (a) final predicted data after the third inversion step of test-L, (b) observed data, (c) difference between the predicted and observed data. Note the good fit between the predicted and the observed seismogram attested by the low residuals.
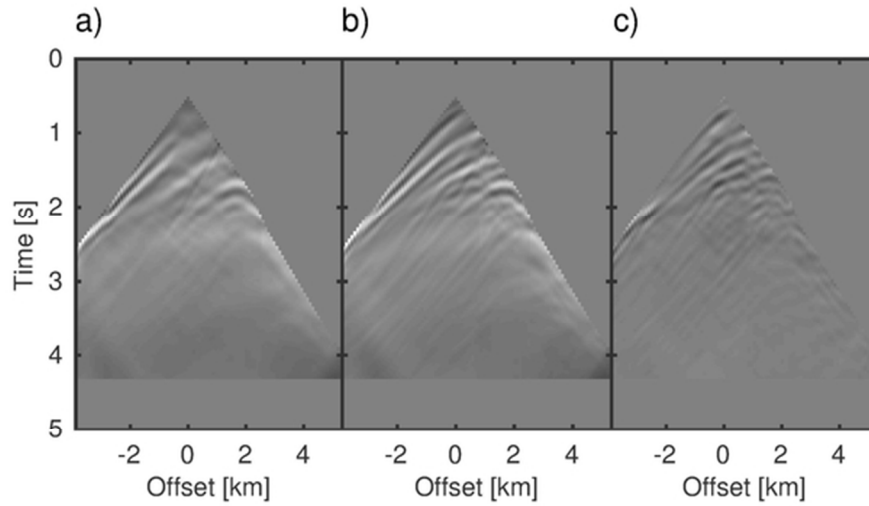
58x31mm (300 x 300 DPI)

Figure 11: Result of the inversion on the reflected waves for the 5th shot-gather: (a) final predicted data after the third inversion step of test-L, (b) observed data, (c) difference between the predicted and observed data. Note that the low-frequency content of the observed wavefield is correctly recovered and the mismatch is mainly due to the high-frequency content.
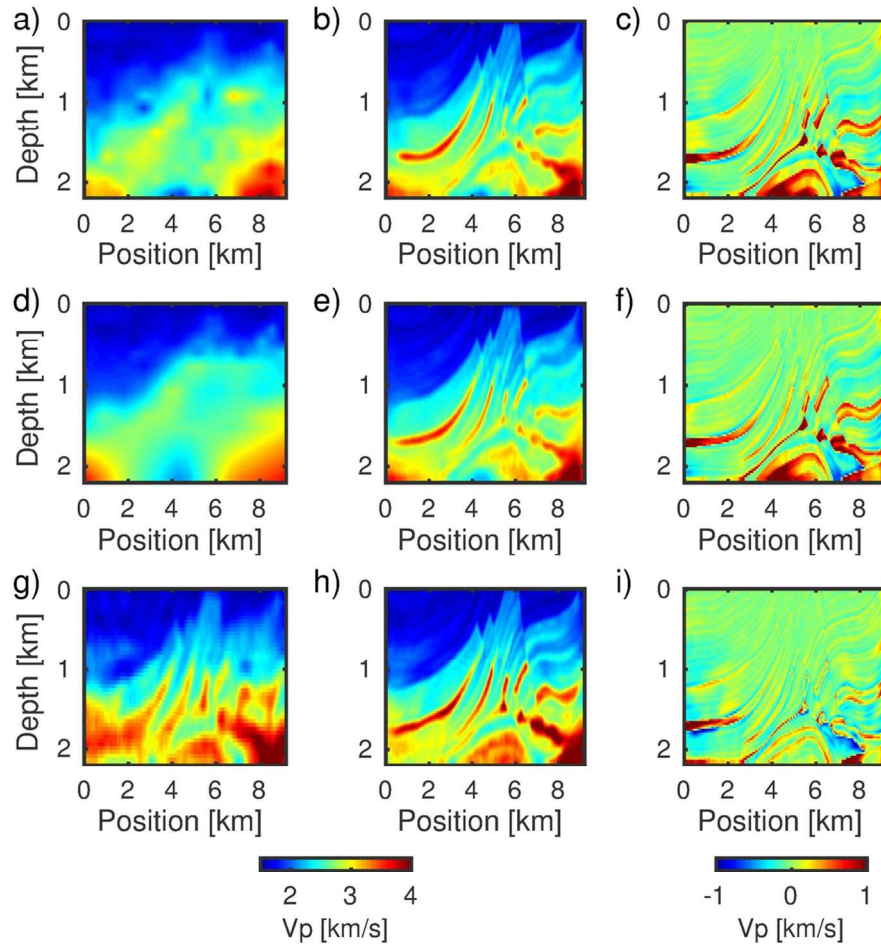
58x31mm (300 x 300 DPI)

Figure 12: Results of the sequence GA FWI + descent-based FWI. First row shows the outcome referred to test-V1: (a) starting model for descent based FWI, i.e., the best-fitting model after GA FWI, (b) final model after descent-based FWI, (c) difference between (b) and the reference model of Figure 1a. In analogy to the first row, the second and third rows show the outcomes referred to test-V2 and test-L, respectively.
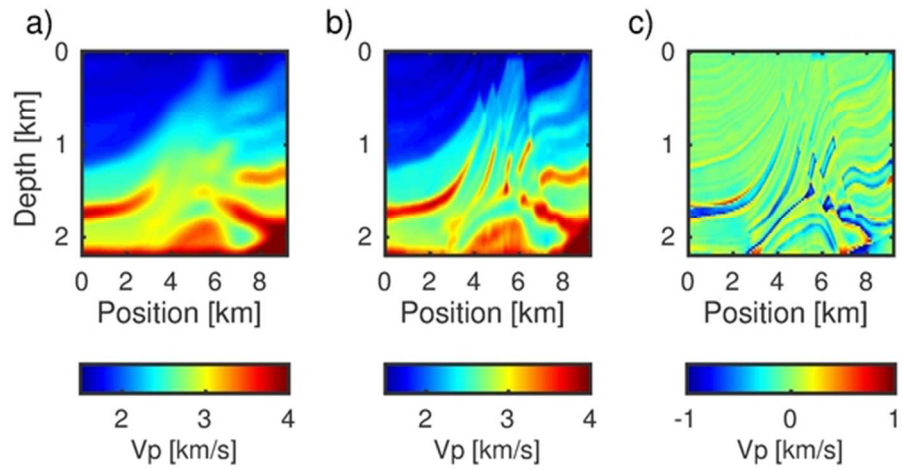
116x123mm (300 x 300 DPI)

Figure 13: Benchmark test: (a) smoothed version of the Marmousi used as starting model for descent-based FWI, (b) final model after descent-based FWI, (c) difference between the descent-based FWI model and the reference model of Figure 1a. The average absolute error of (b) is 80 m/s.\r\n
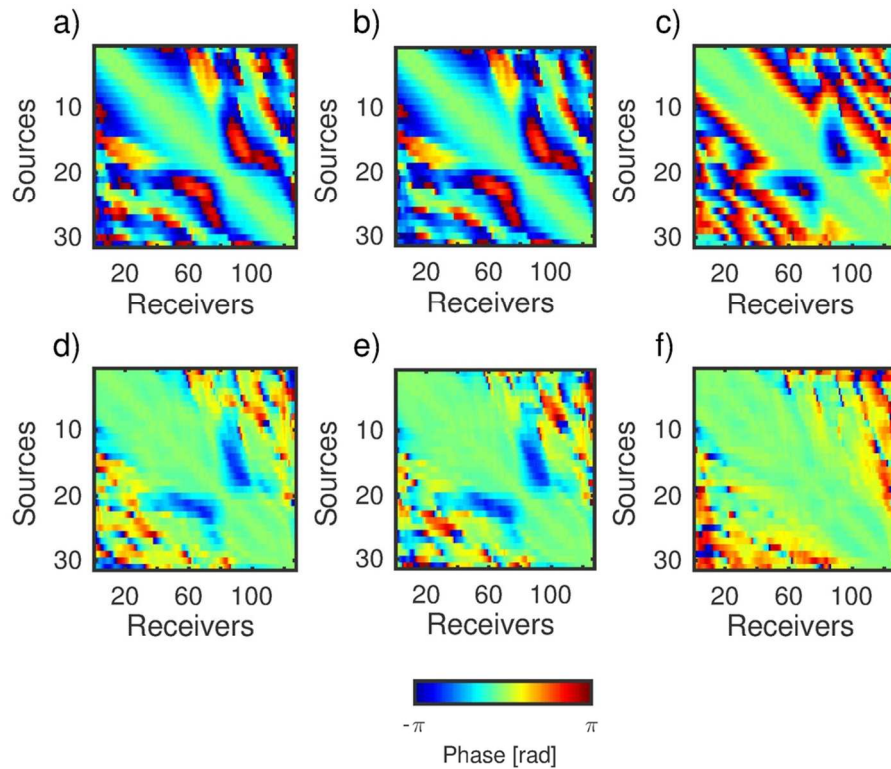58x31mm (300 x 300 DPI)

Figure 14: Phase residual panels (Shah et al., 2012) referred to different starting models: (a) the first velocity analysis model (the model of Figure 2b after fine interpolation), (b) the second velocity analysis model (the model of Figure 2c after fine interpolation), (c) the linearly increasing with depth model (the model of Figure 8a in fine grid). Phase residual panels for (d) the best-fitting model of test-V1; (e) the best-fitting model of test-V2, (f) the best-fitting model of test-L. \r\n
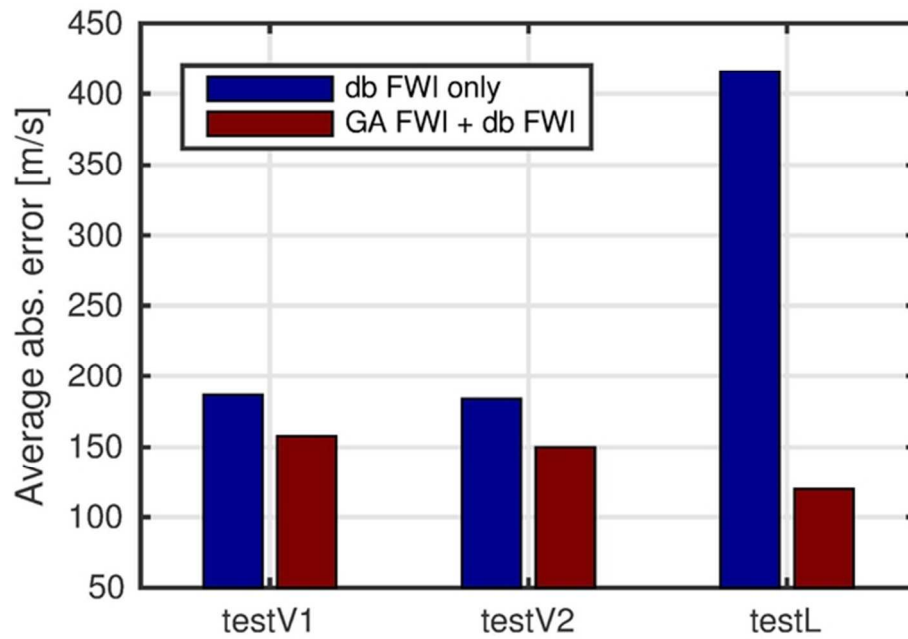101x93mm (300 x 300 DPI)

Figure 15: Bar diagram that shows, in blue, the average absolute error after descent-based FWIs started from models that represent the prior information used in the three tests (V1, V2, and L), and, in red, the average absolute error of the final models after GA FWI + descent-based FWI for the three tests (V1, V2, and L). Note that performing a GA FWI prior to descent-based FWI improves the final model misfit for each of the three tests.
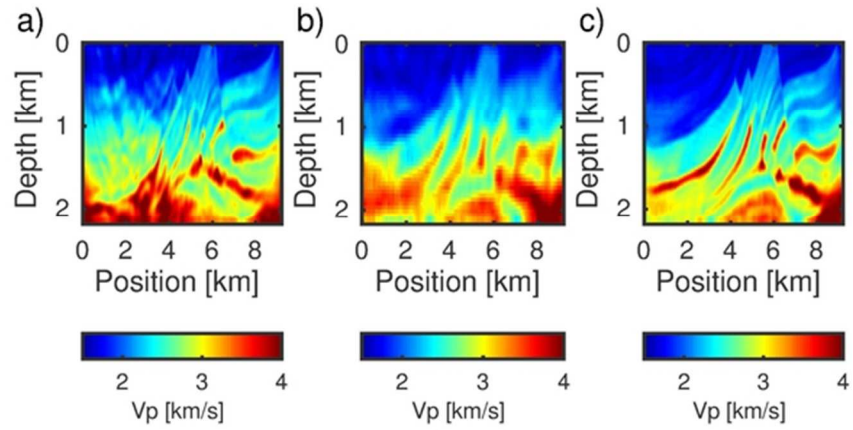
58x40mm (300 x 300 DPI)

Figure 16: Comparison between descent-based FWI, GA FWI, and the sequence GA FWI + descent-based FWI given the same prior information. (a) final model after descent-based FWI using model of Figure 8a as starting model, (b) final model after GA FWI using model of Figure 8a to center the ranges of the stochastic inversion, (c) final model after descent-based FWI using the GA FWI model in (b) as starting model.

58x31mm (300 x 300 DPI)

Table 1 Some important genetic-algorithm parameters used for test-V1 and test-V2.

Parameters	Values
Number of models per generation	500
Ratio of models selected for mating	80%
Ratio of mutated models	10%
Selection method	Stochastic universal sampling

Table 2 Some important genetic-algorithm parameters used for test-L.

Parameters	Values
Number of models per generation	4800
Ratio of models selected for mating	42%
Ratio of mutated models	10%
Selection method	Stochastic universal sampling