Two-grid genetic algorithm full-waveform inversion

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

Abstract

Full-waveform inversion (FWI) tries to estimate velocity models of the subsurface with improved accuracy and resolution compared to conventional methods. To be successful, it needs input data that is rich in low frequencies and possibly characterized by long source-to-receiver offsets. The correct solution of the inverse problem by means of local methods is facilitated if the starting model lies in the "valley" of the cost-function global minimum. We explore the possibility of relaxing this requirement by using genetic algorithms, a stochastic optimization method, as the driver of the FWI (GA FWI). However, stochastic methods are affected by the "curse of dimensionality," meaning that they require huge and sometimes even unaffordable computer resources for inverse problems with many unknowns and costly forward modeling. Therefore, we need to adopt proper stratagems in the inversion and limit our goal to the estimation of a velocity macromodel that is of a model with only the long-wavelength velocity structures, which could eventually act as the starting model for a local, higher-resolution gradient-based inversion. To this end, in the GA FWI we parametrize the subsurface with two grids: (1) a coarse grid with widely spaced nodes, that is unknowns, for the inversion, and (2) a fine grid with shorter spacing for the modeling. As a side result, we can also have an estimate of the uncertainty at the solution nodes of the grid. The approach we discuss is 2D acoustic in the time domain, with finite difference forward modeling. The examples we show refer to the Marmousi model and to a marine field data set.

Introduction

Full-waveform inversion simultaneously considers the traveltimes, amplitudes, and shapes of the recorded wavelets — that is, the entire information of the seismogram — to estimate an optimal and high-resolution velocity field of the subsurface. Successfully applying FWI to a wide variety of geologic models is a very ambitious and much desired goal, and thus FWI has received growing attention from the geophysical community since the earliest works of Tarantola (1984), Mora (1988), and Pratt and Worthington (1990). Successful FWI applications that deal with acoustic, elastic, or anisotropic cases are well described in scientific literature (Brossier et al., 2009; Plessix and Perkins, 2010; Sirgue et al., 2010; Prieux et al., 2011; Morgan et al., 2013 — among others). In general, FWI is carried out by means of iterative local optimization methods that require the computation of the gradient of a misfit function, which is rather complicated due to the nonlinearity and the ill-posedness of this challenging inverse problem (Virieux and Operto, 2009). Therefore, noise contamination, lack of low frequencies, and inaccuracies of the starting model make it difficult for gradient-based optimization algorithms to find the global minimum of a misfit surface affected by many local minima; unless the starting model is in the basin of attraction of the global minimum, there is a significant risk to get trapped in a local minimum.

To reduce this risk, the inversion can be performed starting from the low frequencies only of the data and progressively including the higher frequencies, which is a method known as the multiscale technique (Bunks et al., 1995). Good starting models are generally required to be smooth (Asnaashari et al., 2013) and with the associated synthetic seismograms matching the events (particularly the refracted and diving waves) of the observed seismogram with errors smaller than half of the wavelet period to avoid cycle-skipping artifacts (Beydoun and Tarantola, 1988). Suitable starting models for gradient-based FWI usually are derived by employing reflection and refraction tomography, PSDM velocity analysis, first-arrival traveltime tomography (Nolet, 1987), stereotomography (Billette and Lambaré, 1998), or ensembles of different methods trying also to integrate the available geophysical and geologic data. Laplace domain and Laplace-Fourier domain (Shin and Cha, 2008; Shin and Ha, 2008) have also been proposed. All of these techniques have demonstrated their effectiveness, but they usually require a significant amount of qualified human resources and computing time.

Stochastic methods are less affected than gradient-based methods by the presence of local minima in the error surface and, consequently, are less dependent on the starting model for the inversion. For instance, genetic algorithms do not even require the definition of a starting model because they actually start from a population of models within a range that includes the candidate solutions. Applications of genetic algorithms and simulated annealing in geophysics are numerous; among others, we mention the works of Sen and Stoffa (1991), Stoffa and Sen (1991), Tran and Hiltunen (2012), and Datta and Sen (2016). Unfortunately, within the constraint of a reasonable time budget, stochastic optimization methods still may require unaffordable computing resources if applied to inverse problems with large dimensions of the model space, because the search area of the stochastic inversion grows exponentially with the number of unknowns.

However, the use of genetic algorithms, which allow for a parallel implementation of the code, combined with a tailored parametrization of the subsurface, significantly attenuates the computing-time problem. We chose to adopt a specific implementation of real-valued genetic algorithms following a comparative test between this method and the neighborhood algorithm (Sambridge, 1999a) and the adaptive simulated annealing (Ingber, 1989), in which this genetic-algorithm implementation displayed a better performance in case of high-dimensional spaces (Sajeva et al., 2014a). In addition, genetic algorithms provide a wide exploration of the model space that allows for the estimation of uncertainties in the final result (Aleardi and Mazzotti, 2016). Concerning the subsurface parametrization, we propose to employ a two-grid technique in which the subsurface is described by

[1]University of Pisa, Department of Earth Sciences.
[2]Eni, Upstream & Technical Services.
[3]University of Milan, Department of Earth Sciences.

means of a fine grid for the finite-difference forward modeling and by a coarse grid for the stochastic inversion (Sajeva et al., 2014b; Sajeva et al., 2016).

In what follows, we first outline the proposed method, which we name two-grid genetic algorithm full-waveform inversion (GA FWI), and then we move on to illustrate two examples: one pertinent to 2D synthetic data of the Marmousi model and the other to a 2D marine actual data case. We limit to 2D acoustic cases. Finally, we discuss the pros and cons of our method, the high-performance computing implications, and the future perspectives.

Two-grid GA FWI method

Subsurface parametrization. We discretize the subsurface with two grids: a "coarse" inversion grid and a "fine" modeling grid. Each node of the inversion grid corresponds to an unknown of the GA optimization, which, in the acoustic approximation we use, is the P-wave velocity at that node. The horizontal and vertical step sizes of the inversion grid cells determine the ultimate resolution of the estimated velocity model. The grid cells can be of fixed dimensions or can vary laterally and/or with depth according to the presumed illumination given by the source-receiver layout and to Fresnel zone concepts. Ideally, they can also be adapted during the inversion following the evolution of the velocity model being estimated by the GA. The dimensions of the grid cells determine the number of grid nodes that is the number of unknowns in the inversion, and thus they must be chosen also to limit the GA optimization to a reasonable amount of time, given the available computing resources. For this reason, the inversion grid is "coarse." The higher the computing resources, the less "coarse" can become the inversion grid and the higher can be the resolution of the estimated final model, within the limits of the considered frequency band. Figure 1 shows the inversion grid used for the Marmousi example that we will discuss in the next section. It consists of an irregular grid with variable cell size that increases with depth in accordance to the loss of resolution of the seismic data. The total number of grid nodes that is of unknowns is 143.

A bilinear interpolation converts the inversion grid to the modeling grid that is characterized by shorter and fixed dimensions of the grid cell, enabling a more accurate description of the subsurface morphology (such as the geometry of the seafloor) and the computation of higher frequencies without numerical dispersion. We compute the predicted data employing an acoustic FD modeling with an accuracy of the fourth order in space and of the second order in time. In the modeling grid used for the Marmousi example, the grid nodes have a constant spacing in the horizontal and vertical
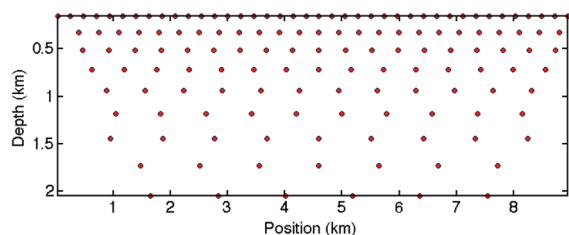
directions equal to 24 m, allowing us to compute synthetic seismograms up to 12 Hz with negligible numerical dispersion.

Data misfit and genetic-algorithm optimization. We compute the L1 or L2 norm misfit between observed and predicted data considering either the waveforms, the energy, the envelopes, or whatever other attribute we may deem as appropriate for the specific case. For instance, we try envelopes when the observed data quality is poor and when it is difficult to estimate a reliable source wavelet. As in many previous works (Chironi et al., 2006; Vigh et al., 2010; Bi and Lin, 2014), the misfit function can be devised as to implement a layer stripping, an offset stripping procedure, or to refer to different wave portions of the data, such as in equation 1 where one term includes the refracted and diving waves, and the second term considers the reflected wavefield balanced by a weight parameter α.

Data Error = (observed diving − predicted diving) +
α (observed reflected − predicted reflected)   (1)

The data misfit, that is the L1 or L2 norm of the data error, drives the real valued GA optimization that we employ for FWI. Genetic algorithms (Holland, 1975) are a class of stochastic optimization methods that search for the global minimum of the data misfit function within a given search range and do not require any calculation of derivatives of the data error surface. Therefore, they are less prone to get trapped into local minima. Genetic algorithms carry out the search of the model space by mimicking the natural evolution processes and evolving a population of velocity models (individuals) toward the attainment of a higher fitness that is of a lower data misfit between the corresponding synthetic seismograms and the observed data. Three main operations drive the GA optimization (Figure 2): selection, recombination, and mutation. An initial population of velocity models is randomly generated within predefined bounds, and the fitness of each candidate model is evaluated. Then, different models are stochastically selected on the basis of their associated
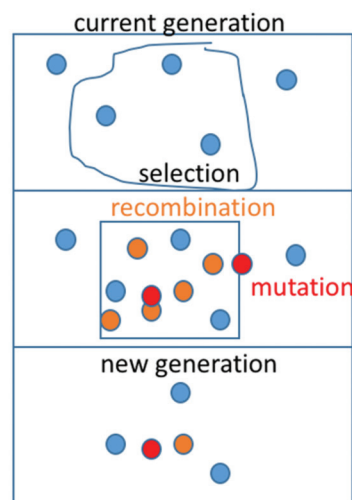

Figure 1. The coarse and irregularly spaced inversion grid used for the GA FWI example on the Marmousi data. The number of grid nodes is 143 while the number of nodes in the fine and regularly spaced modeling grid (not shown) is approximately 35,000 (more than 200 times the inversion grid points).


Figure 2. From a population model (five blue dots), three models are selected because of their associated low data misfit. Recombination and mutation change the "genes" (velocity values) of the selected models creating several offspring (red and orange dots), which are again selected to form a new population. Note that the number of individuals for each generation remains constant.

data misfit; models giving rise to lower data misfit are more likely to be selected. Recombination and mutation operators intervene to modify the selected models to form new models (offspring individuals) that are a new population, which is then the input for the next iteration (generation).

The evolution of a GA optimization is controlled by the setting of several parameters that define the number of individuals of the population, the type of selection mechanism adopted, the intensity of the selection pressure, the possibility to explore the model space by different subpopulations, the possibility to reinsert the best parents in the next population (elitist strategy), and so on (Mitchell, 1996). We have carried out extensive tests on both analytical and experimental data misfit surfaces to gain an understanding of the impact of each parameter on the ability of the algorithm to find the global minimum and on the speed with which convergence is attained. The most important parameter is the number of individuals that must always be greater (possibly much greater) than the number of unknowns to facilitate a thorough exploration of the model space, while the use of subpopulations generally increases the speed of convergence. The initial population distribution and the width of the search ranges within the model space are chosen on the basis of the a priori knowledge we have on the velocity field: uniform distribution and large search ranges are appropriate when a priori info is missing or unreliable. In the examples that follow, we specify the parameters we adopted; detailed descriptions can be found in Sajeva et al. (2016) and Tognarelli et al. (2015).

Uncertainty estimation via Gibbs sampling. Genetic algorithms, like other global optimization methods, explore the model space and, besides the best model, gather many different models that we generally discard. Instead, the ensemble of models sampled by the GA exploration is of great use if we wish to measure the uncertainty of the final result, that is if we wish to
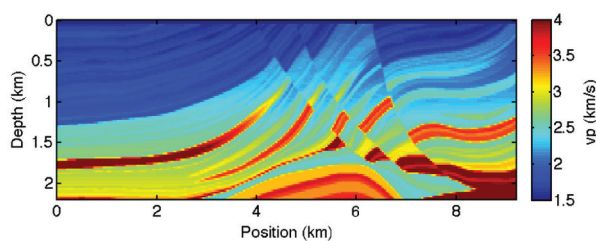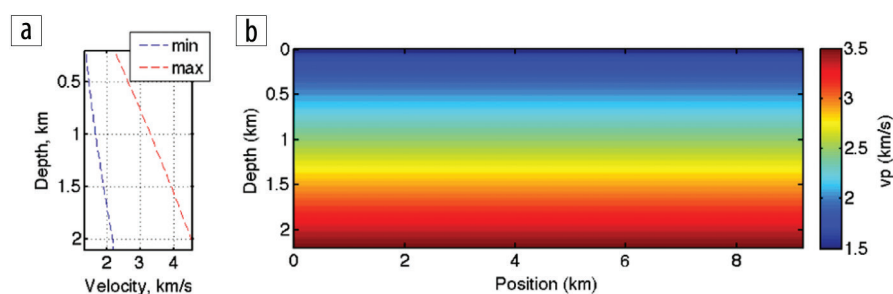
represent the final solution by the posterior probability distributions (PPDs) in model space. However, it is not possible to derive an unbiased estimate of the PPDs directly from the ensemble of GA models because GAs are not a Markov chain Monte Carlo (MCMC) method (Rubinstein and Kroese, 2011). In practice, GAs underestimate the true uncertainties (variances) because they tend to oversample the model space zones corresponding to the lowest data misfit.

Among the methods presented in the literature to derive an unbiased estimation of the PPDs, we adopt a MCMC method known as Gibbs sampler (Geman and Geman, 1984) that performs a resampling of the model space making use of all the models and their respective data misfits found by the GA inversion. In particular, the model space explored by GA is first partitioned into Voronoi cells, each one associated with a single GA model and its misfit value, building a multidimensional interpolant that is successively sampled by the Gibbs sampler (Sambridge, 1999b). Note that with this method, there is no need to perform additional forward modeling computations than those performed during the GA optimization. Specific details on the hybrid genetic algorithms + Gibbs sampler approach in a framework of full-waveform inversion can be found in Aleardi and Mazzotti (2016).

Marmousi: 2D acoustic example

We have extensively tested the GA FWI on the Marmousi model (Figure 3), trying different subsurface parametrization, different search areas, and a priori information for the GA optimization and different GA parameter settings (Sajeva et al., 2016). It is on this reference model that we compute the synthetic "observed" seismograms, simulating an acquisition with 31 equally spaced sources and 127 equally spaced receivers at the surface, with a 6 Hz Ricker wavelet as the source signature.

In the example we discuss here, the unknowns for the GA optimization are the P-wave velocities at the 143 nodes of the irregular coarse grid shown in Figure 1, where the size of the grid cells increases with depth. The search area for the inversion (Figure 4a) is centered on a simple 1D velocity model (Figure 4b) and the initial population with uniform distribution is randomly selected within the search ranges.

The coarse inversion grid of Figure 1 is bilinearly interpolated to the fine modeling grid for computing the predicted data. We employ the same finite-difference code, with an accuracy of second order in time and fourth order in space, for both computing the "observed" data on the true Marmousi model of Figure 3 and the predicted data during the GA optimization. Thanks to the peculiarly favorable situation, with no noise and with known source wavelet, the data misfit is computed including simultaneously all the recorded events with no layer stripping and no weighting, that is with $\alpha = 1$ in equation 1. We employ the L2 norm applied to low-pass filtered (0–3 Hz) and trace-by-trace normalized data. Note that we have no significant energy at 0 Hz (it is less than -40 dB), and the majority of the signal is carried by frequencies around 3 Hz. The GA

Figure 3. The Marmousi model that is the reference model used for testing the two-grid GA FWI.

Figure 4. (a) The search range used for the GA optimization is centered on (b) the simple 1D linearly increasing velocity model.

optimization is launched with the setting parameters indicated in Table 1 for a total of 40,500 evaluated models.

Figure 5 shows the evolution of the mean and minimum data misfit (green and red curve, respectively) with advancing generations. After the initial drop, the two curves gradually converge as the algorithm approaches a minimum and at the end of 100 generations, that is after 40,500 models have been evaluated, the mean and minimum values are almost coincident indicating that genetic diversity is lost and that computing further models will not improve the result. The final best-fitting model is shown in Figure 6: note that it fairly reproduces the long wavelength velocity structure of the Marmousi and that it is a smooth and low-resolution velocity model. The smoothness and the low resolution are the consequence of the large spacing of the inversion grid and of the bilinear interpolation, which operates to bring the velocities on the coarse grid to the fine grid. Decreasing the size of the inversion grid cells would allow for a more detailed reconstruction of the subsurface but at the expense of increasing the number of unknowns.

Collecting the models explored by the GA optimization enables us to quantify the uncertainty affecting the best model. In Figure 7, the green bars show the distribution of the models sampled by the GA (which we name GA distributions) for the six grid nodes indicated by the red dots in Figure 6. These GA distributions are the input to the Gibbs sampler to retrieve a reliable estimation of the true posterior probability distributions. The blue curves overlapped to the GA distributions represent the 1D marginal PPD for the same six grid nodes. The red dashed lines indicate the best model velocities. The PPDs become broader and bimodal moving from the center (grid node N.2) to the sides (grid nodes N.1 and N.3) of the model and for increasing depths (grid nodes N.4, 5, and 6; please note the different scales of the horizontal axes), which is where we expect a loss of data information due to a poor seismic illumination. Therefore, for each grid node of interest, we not only can estimate an optimal velocity value but we also can provide its uncertainty — information which may be of use for further applications.

Depending on the degree of resolution reached by the two-grid GA FWI, the best model can be tried as the velocity field for prestack depth migration, or it can become the starting model for a successive gradient-based FWI to gain the fine details of the velocity structure, without risking to start from a "bad" model that is too far from the global minimum. In fact, tests on Marmousi (Sajeva et al., 2016) and on other data have shown that GA FWI models yield a significant decrease of cycle skips compared to the prior models, particularly for refracted/diving-wave events.

Taking the GA FWI model of Figure 6 as the starting model for a time-domain gradient-based FWI (based on the steepest descent method) and running five iterations at 4, 5, 6, 8, and 10 Hz we get the result shown in Figure 8. The gradient-based FWI that started from the two-grid GA FWI best model

Table 1. Relevant parameters of the genetic algorithm FWI for the Marmousi test.

Individuals per population	500
Number of subpopulations	5
Selection probability	80%
Mutation probability	0.7%
Number of generations	100
Number of evaluated models	40,500
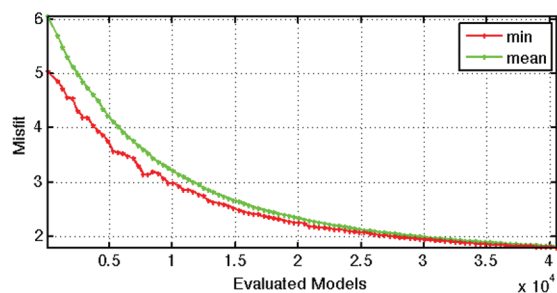
Figure 5. Evolution of the data misfit during the GA optimization. The green curve indicates the evolution of the mean error, while the red curve shows the evolution of the minimum error (that is, the error associated to the best model of each generation). Note how the two curves converge with increasing number of evaluated models.
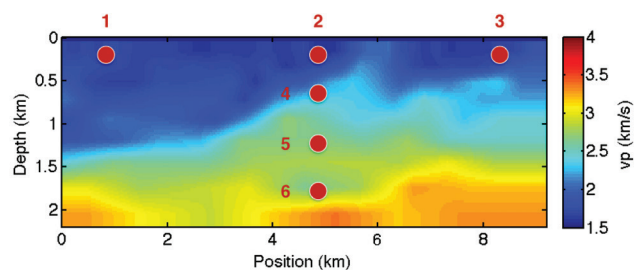

Figure 6. Best model resulting from the two-grid GA FWI. Note its smoothness and low resolution, and that it fairly reproduces the general structure of the Marmousi model. The greatest discrepancies are located at the edges of the model and at depth where the seismic illumination is poor. The numbered dots indicate the position where uncertainties are estimated.
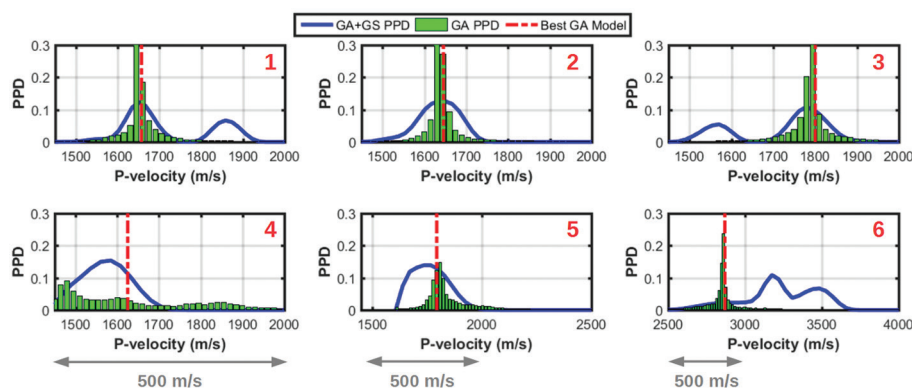
Figure 7. Uncertainty estimates for the six grid nodes indicated in the previous figure. The green bars represent the GA distributions that are the ensemble of models explored by the GA optimization. The blue curves represent the posterior probability distributions estimated by means of the Gibbs sampler. Note that the horizontal axes are plotted with different scales to better evidence the different shapes of the distributions. The vertical red dashed lines indicate the best model velocities.

has successfully retrieved many details, and the final model of Figure 8 nicely matches the true Marmousi of Figure 3. Again, most of the mismatches are located at the edges of the model and at depth where the seismic illumination is scarce and any data-driven inversion cannot do much.
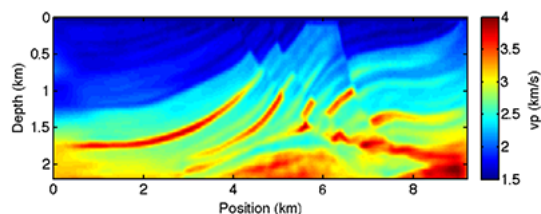

Figure 8. Final model after gradient-based FWI up to 10 Hz started from the two-grid GA FWI model of Figure 6. Note the significant increase of details brought by the gradient-based inversion.

Marine 2D real data example

We illustrate the application of the two-grid GA FWI on an inline from a marine 3D survey. In particular, we test whether we can estimate a preliminary, quick-look velocity field for prestack depth migration, with no a priori information on the velocities and with very limited processing effort, such as would be the case in an early stage of a seismic exploration project. We consider 56 shot gathers with source-to-receiver offset from 180 m to 4000 m evenly distributed along the line for a total of 6517 recorded traces. Figure 9 shows a sample shot gather and its amplitude spectrum. Note the lack of low frequencies. The source wavelet is estimated from the data.

The initial population of velocities for the GA optimization has a uniform probability distribution centered on a 1D linearly increasing velocity with depth (from 1500 m/s at the surface to 3500 m/s at 1300 m depth), with range +/-500 m/s.
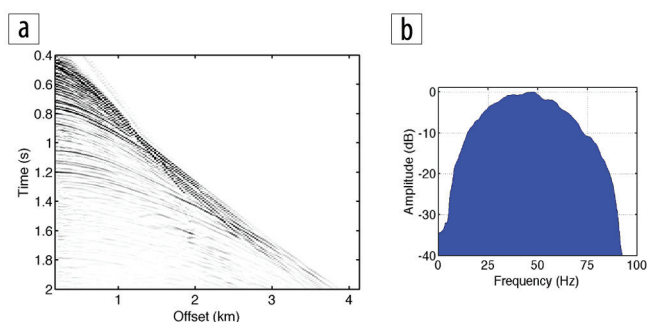
Figure 9. (a) A shot gather taken from the inline data and (b) its amplitude spectrum.

Table 2. Relevant parameters of the genetic algorithm FWI for the marine data test.

Individuals per population	500
Number of subpopulations	2
Selection probability	80%
Mutation probability	1.3%
Number of generations	100
Number of evaluated models	40,500

As compared against the previous synthetic example, the inversion grid is now regular with rectangular cells sized 600 m by 150 m, and the number of nodes (and unknowns) is only 78 (13 in the horizontal and 6 in the depth direction, respectively). Table 2 presents the main GA settings used for this example. Other initial populations and other inversion grids have been presented in Tognarelli et al. (2015). The modeling grid is of 242 × 40 nodes, with an isometric grid size of 30 m. The data misfit is the L1 norm between the envelopes of the observed and predicted data, both low-pass filtered (0–6 Hz) and trace-by-trace normalized. In this example, we show the results for the inversion of the diving waves only, which have been selected using pre-defined mute functions. Convergence is attained after 40,500 models have been evaluated by the GA optimization and the resulting best model is shown in Figure 10a, while the comparison between the predicted and the observed diving waves for two shot gathers along the inline is shown in Figure 10b.

To check whether this preliminary result may be of any practical value, we prestack depth migrate the seismic data making use of the best velocity model of Figure 10a as the migration velocity field: the degree of horizontal alignments of the events in common-image gathers (CIGs) is the quickest way to assess the results. To this end, in Figure 11, we show 15 CIGs evenly spaced along the profile after PSDM. The maximum depth shown (1.5 km) corresponds to the approximate penetration depth of the refracted and diving waves we considered in the inversion. Band-pass filtering, trace-by-trace normalization, and gain were applied for display purposes. Migrating the data with the prior velocity field, that is with the 1D velocity model at the center of the search range for the GA optimization, yields the migrated gathers in Figure 11a. Obviously, severe misalignments are present. Instead, migrating the CIGs with the best velocity model (Figure 10a) resulting from the two-grid GA FWI, we obtain the migrated gathers of Figure 11b where a significant improvement of the horizontal alignment of the events can be observed up to 1 km of depth. Note that below that threshold, the gathers still exhibit complex moveout; this could be due to the fact that FWI used only the refracted/diving waves, which likely explore the upper layers. Consequently, this velocity model can be further improved, for instance by including the reflected events in an additional step of two-grid GA FWI, or it can be used as the starting model for a gradient-based FWI.

Discussion and further work

The main pros and cons of the two-grid GA FWI we propose are:

1) It is less affected than gradient-based FWI methods by the local minima issue, and thus it mitigates the need of very low frequencies (down to few cycles per second) with a good S/N in the observed data or of a "good" starting model.

2) Therefore, it can be performed with virtually no a priori information, as it can be started from an initial random population of velocity models, such as centered on a very simple 1D velocity model or on a velocity field derived from stacking velocity analyses. This permits its application in the very early stages of a seismic exploration project.

Special Section: Full-waveform inversion Part I

3) During optimization, GA FWI evaluates the data misfit pertaining to many different models; this allows the estimation of PPDs, the treatment of uncertainties, and thus to assess the uncertainty associated to the final velocity model.

4) However, GA FWI requires computational costs that grow exponentially with the number of unknowns; thus, the degree of resolution of the final outcome depends also on the available computer resources. In the examples we show here, we deal with 2D cases, we employ the acoustic approximation, and we aim at estimating the best velocity macromodel of the subsurface, which is a low-resolution, long-wavelength velocity model.

This last point is related to the fact that GA FWI is a computationally intensive task that requires a large number of model evaluations (in the order of tens or of hundreds of thousands) to adequately explore the multidimensional model space. In the Marmousi test with 143 unknowns, more than 40,000 model evaluations were performed in a parallel scheme — that is, 20 model evaluations were computed simultaneously in a group of five compute nodes (each compute node is a two, eight-core CPU at 2.40 GHz). This test runs in approximately three days. In another test, we reached up to 2200 unknowns and the computation of about 3×10^6 models was needed to attain convergence (Sajeva et al., 2016). Therefore, the availability of huge computing resources is a prerequisite for extending the application to 3D cases. Put in perspective, this is going to be less of an issue: computer

technology is constantly advancing, and as of today our industry can already deploy high-performance computing systems large enough to start coping with such challenges.

Moreover, besides making available powerful computing resources and applying various software optimization tricks (see e.g., Bienati et al., 2010), there is also room for further improvement of the method itself. For instance, the coarse grid of the inversion could vary with the progress of the inversion, with spatial steps that locally evolve depending on the updates of the velocity model. Or, we may try to reduce the number of forward
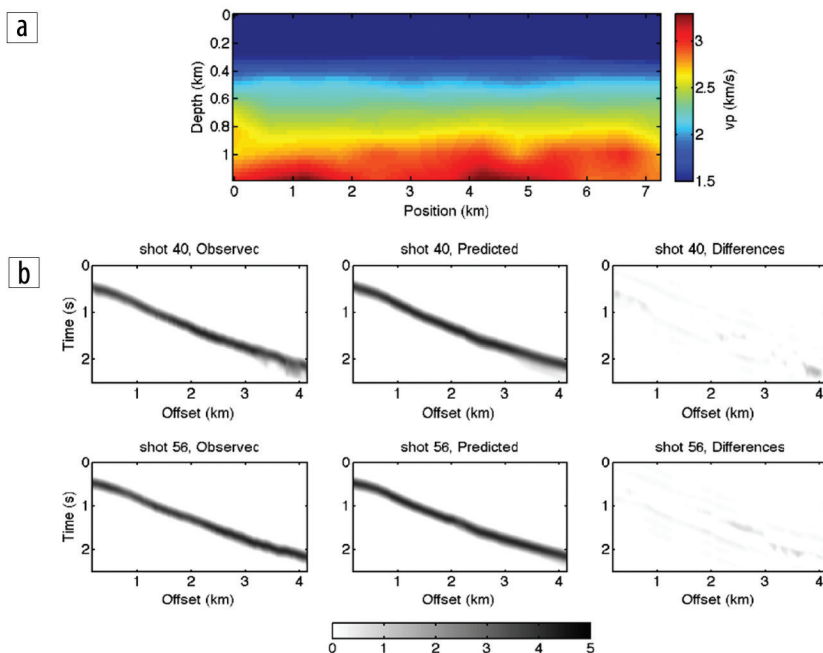
Figure 10. (a) Best velocity model (represented in the fine grid) resulting from the two-grid GA FWI of the refracted/diving waves of the marine seismic line. The inversion has reconstructed significant velocity variations, particularly in the horizontal direction. (b) Observed and predicted data and their difference (left, central, and right column, respectively) for two-shot gathers along the inline. Note the satisfactory matching between the envelopes of the observed and predicted refracted/diving waves.
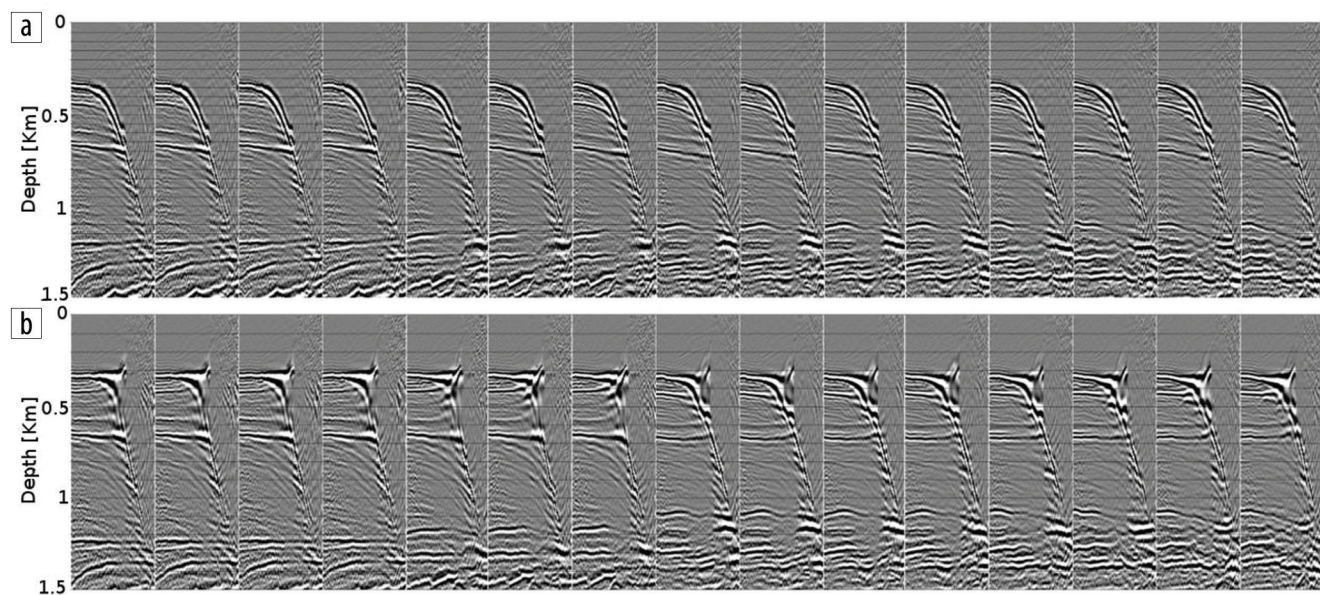
Figure 11. CIGs derived from PSDM (Kirchhoff). In (a), the migration velocity field is the 1D velocity model at the center of the search range for GA FWI. In (b), the migration velocity field is the velocity model of Figure 10a, obtained by GA FWI. The increase of the flatness of the events from (a) to (b) is satisfactory, considering the quick-look procedure with no a priori info and no interpretation or analysis effort.

modeling computations by imposing a local spatial correlation of the model parameters on the models being considered. Further studies and tests on synthetic and actual data are ongoing to further assess the applicability of the method. **TLE**

Acknowledgments

The authors wish to thank Eni for its continued support in the research and for the permission to publish this paper. The seismic data processing was carried out using the Promax software of Landmark Graphics Corporation that is gratefully acknowledged.

Corresponding author: angelo.sajeva@for.unipi.it

References

Aleardi, M., and A. Mazzotti, 2016, 1D elastic full-waveform inversion and uncertainty estimation by means of a hybrid genetic algorithm-Gibbs sampler approach: Geophysical Prospecting, http://dx.doi.org/10.1111/1365-2478.12397.

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, http://dx.doi.org/10.1190/geo2012-0104.1.

Beydoun, W. B., and A. Tarantola, 1988, First Born and Rytov approximation: Modeling and inversion conditions in a canonical example: The Journal of the Acoustical Society of America, **83**, no. 3, 1045–1055, http://dx.doi.org/10.1121/1.396537.

Billette, F., and G. Lambaré, 1998, Velocity macro-model estimation from seismic reflection data by stereotomography: Geophysical Journal International, **135**, no. 2, 671–690, http://dx.doi.org/10.1046/j.1365-246X.1998.00632.x.

Bi, H., and T. Lin, 2014, Impact of adaptive data selection on full waveform inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 1094–1098, http://dx.doi.org/10.1190/segam2014-0310.1.

Bienati, N., C. Danken, C. Calonaci, and R. Gori, 2010, Optimal implementation of Finite Difference stencils for the solution of the wave equation on multicore processors: 85th Annual International Meeting, SEG, Expanded Abstracts, 3283–3287, http://dx.doi.org/10.1190/1.3513529.

Brossier, R., S. Operto, and J. Virieux, 2009, Two-dimensional seismic imaging of the Valhall model from synthetic OBC data by frequency-domain elastic full-waveform inversion: 84th Annual International Meeting, SEG, Expanded Abstracts, 2293–2297, http://dx.doi.org/10.1190/1.3255318.

Bunks, C., F. M. Saleck, S. Zaleski, and G. Chavent, 1995, Multiscale seismic waveform inversion: Geophysics, **60**, no. 5, 1457–1473, http://dx.doi.org/10.1190/1.1443880.

Chironi, C., J. V. Morgan, and M. R. Warner, 2006, Imaging of intrabasalt and subbasalt structure with full wavefield seismic tomography: Journal of Geophysical Research, **111**, B5, 1–18, http://dx.doi.org/10.1029/2004JB003595.

Datta, D., and M. Sen, 2016, Estimating a starting model for full waveform inversion using a global optimization method: Geophysics, **81**, no. 4, R211–R223, http://dx.doi.org/10.1190/geo2015-0339.1.

Geman, S., and D. Geman, 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images: IEEE Transactions on Pattern Analysis and Machine Intelligence, **PAMI-6**, no. 6, 721–741, http://dx.doi.org/10.1109/TPAMI.1984.4767596.

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.

Ingber, L., 1989, Very fast simulated re-annealing: Mathematical and Computer Modeling, **12**, no. 8, 967–973, http://dx.doi.org/10.1016/0895-7177(89)90202-1.

Mora, P., 1988, Elastic wave-field inversion of reflection and transmission data: Geophysics, **53**, no. 6, 750–759, http://dx.doi.org/10.1190/1.1442510.

Morgan, J., M. R. 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, http://dx.doi.org/10.1093/gji/ggt345.

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

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

Plessix, R. E., and C. Perkins, 2010, Thematic set: Full waveform inversion of a deep-water ocean bottom seismometer data set: First Break, **28**, no. 1728, 71–78, http://dx.doi.org/10.3997/1365-2397.2010013.

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, http://dx.doi.org/10.1111/j.1365-2478.1990.tb01846.x.

Prieux, V., R. Brossier, Y. Gholami, S. Operto, J. Virieux, O. Barkved, and J. Kommedal, 2011, On the footprint of anisotropy on isotropic full waveform inversion: The Valhall Case Study: Geophysical Journal International, **187**, no. 3, 1495–1515, http://dx.doi.org/10.1111/j.1365-246X.2011.05209.x.

Rubinstein, R. Y., and D. P. Kroese, 2011, Simulation and the Monte Carlo method: John Wiley & Sons.

Sambridge, M., 1999a, Geophysical inversion with a neighborhood algorithm – I. Searching a parameter space: Geophysical Journal International, **138**, no. 2, 479–494, http://dx.doi.org/10.1046/j.1365-246X.1999.00876.x.

Sambridge, M., 1999b, Geophysical inversion with a neighborhood algorithm – II. Appraising the ensemble: Geophysical Journal International, **138**, no. 3, 727–746, http://dx.doi.org/10.1046/j.1365-246x.1999.00900.x.

Sajeva, A., M. Aleardi, E. Stucchi, A. Mazzotti, and B. Galuzzi, 2014a, Comparison of stochastic optimization methods on two analytic objective functions and on a 1D elastic FWI: 76th Conference and Exhibition, EAGE, Extended Abstracts, P0115, http://dx.doi.org/10.3997/2214-4609.20140857.

Sajeva, A., M. Aleardi, A. Mazzotti, N. Bienati, and E. Stucchi, 2014b, Estimation of velocity macro-models using stochastic full-waveform inversion: 85th Annual International Meeting, SEG, Expanded Abstracts, 1227–1231, http://dx.doi.org/10.1190/segam2014-1088.1.

Sajeva, A., M. Aleardi, E. Stucchi, N. Bienati, and A. Mazzotti, 2016, Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the Marmousi model: Geophysics, **81**, no. 4, R173–R184, http://dx.doi.org/10.1190/geo2015-0198.1.

Sen, M. K., and P. L. Stoffa, 1991, Nonlinear one-dimensional seismic waveform inversion using simulated annealing: Geophysics, **56**, no. 10, 1624–1638, http://dx.doi.org/10.1190/1.1442973.

Shin, C., and Y. H. Cha, 2008, Waveform inversion in the Laplace domain: Geophysical Journal International, **173**, no. 3, 922–931, http://dx.doi.org/10.1111/j.1365-246X.2008.03768.x.

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, http://dx.doi.org/10.1190/1.2953978.

Sirgue, L., O. I. Barkved, J. Dellinger, J. Etgen, U. Albertin, and J. H. Kommedal, 2010, Full waveform inversion: The next leap forward in imaging at Valhall: First Break, **28**, no. 1728, 65–70, http://dx.doi.org/10.3997/1365-2397.2010012.

Stoffa, P. L., and M. K. Sen, 1991, Nonlinear Multiparameter Optimization using genetic algorithms: inversion of plane-wave seismograms: Geophysics, **56**, no. 11, 1794–1810, http://dx.doi.org/10.1190/1.1442992.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, no. 8, 1259–1266, http://dx.doi.org/10.1190/1.1441754.

Tognarelli, A., E. Stucchi, N. Bienati, A. Sajeva, M. Aleardi, and A. Mazzotti, 2015, Two-grid stochastic full waveform inversion of 2D marine seismic data: 77th Conference and Exhibition, EAGE, Extended Abstracts, N104 01, http:/dx.doi.org/10.3997/2214-4609.201413197.

Tran, K. T., and D. R. Hiltunen, 2012, Two-dimensional inversion of full waveforms using simulated annealing: Journal of Geotechnical and Geoenvironmental Engineering, **138**, no. 9, 1075–1090, http://dx.doi.org/10.1061/(ASCE)GT.1943-5606.0000685.

Vigh, D., B. Starr, J. Kapoor, and H. Li, 2010, 3D full waveform inversion on a Gulf of Mexico WAZ data set: 80th Annual International 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, http://dx.doi.org/10.1190/1.3238367.