An innovative method to attenuate genetic drift in genetic-algorithm optimizations: examples on analytic objective functions, residual statics corrections, and 2D acoustic full-waveform inversion

Silvio Pierini, Mattia Aleardi, Alfredo Mazzotti

Abstract

Genetic algorithms (GAs) usually suffer from the so-called genetic-drift effect that reduces the genetic variability within the evolving population making the algorithm converge toward a local minimum of the objective function. We propose an innovative method to attenuate such genetic-drift effect that we name Drift-Avoidance Genetic Algorithm (DAGA). The implemented method combines some principles of Niched Genetic Algorithms (NGAs), catastrophic GAs, crowding GAs, and Monte Carlo algorithm (MCA) with the aim to maintain an optimal genetic diversity within the evolving population, thus avoiding premature convergence. The DAGA performances are first tested on different analytic objective functions often used to test optimization algorithms. In this case, the implemented DAGA approach is compared with standard GAs, catastrophic GAs, crowding GAs, NGAs, and MCA. Then, the DAGA and the NGAs approaches are compared on two well-known non-linear geophysical optimization problems characterized by objective functions with complex topologies: residual statics corrections and 2D acoustic full-waveform inversion (FWI). To draw general conclusions, we limit the attention to synthetic seismic optimizations. Our tests prove that the DAGA approach grants the convergence in case of objective functions with very complex topologies, where other GA implementations (such as standard GAs or NGAs) fail to converge. Differently, in case of simpler topologies, DAGA achieves similar performances with the other GA implementations considered. The DAGA approach may have a slightly higher or lower computational cost than

standard GA or NGA methods, depending on its convergence speed, that is on its ability to reduce the number of forward modelings with respect to the other methods.

Introduction

Geophysical optimization problems are often non-linear, multidimensional, and characterized by objective functions with complex topologies (i.e., multiple local minima). There are essentially two strategies to tackle these problems: apply linearized methods that involve the computation of the gradient of the objective function (e.g., the Gauss–Newton, conjugate-gradient; Aster et al. 2005), or apply global optimization algorithms (e.g., Genetic Algorithms, Simulated Annealing, Particle Swarm; see e.g. Sajeva et al. 2017a) that perform a direct search in the model space. On the one hand, local methods are attractive because they usually converge rapidly, but they need a good starting model to find the global minimum. On the other hand, global methods are theoretically able to exhaustively explore the entire model space and are not affected by the choice of the starting model. However, the latter are usually affected by the curse of dimensionality problem and for this reason they require many model evaluations to converge. Obviously, this issue has often prevented the application of global methods to optimization problems with many unknown model parameters and/or requiring expensive forward-modeling evaluations.

Over the last decades, the increasing computational power provided by modern computers has encouraged the applications of global optimization methods for solving geophysical problems (many examples can be found in Sen and Stoffa, 2013), and undoubtedly Genetic Algorithms (GAs) are one of the most widely applied global optimization strategies in exploration geophysics (Sen and Stoffa, 1992; Sen and Stoffa, 1996; Mallick 1999; Mallick and Dutta 2002; Padhi and Mallick 2014; Sajeva et al. 2014; Aleardi 2015; Li and Mallick 2015; Aleardi, et al. 2016; Sajeva et al. 2016; Aleardi and Ciabarri 2017; Aleardi and Mazzotti 2017; Sajeva, et al. 2017b). The basic idea of GAs is to define a population of individuals encoding candidate solutions for the problem at hand and to make evolve such population by applying some principles present in nature: inheritance, crossover, mutation,

survival of the fittest, migrations, and so on. The evolution process is driven by the fitness-function value that expresses the goodness of each candidate solution. Notwithstanding its wide applicability, it is well-known that the GA method suffers from the genetic-drift effect (Goldberg and Segrest 1987), which limits the exploration of the model space and may guide the algorithm to prematurely converge toward a local minimum of the objective function.

The first proposed global optimization method was the uniform Monte Carlo (MC) algorithm that performs an undirected and totally random search across the entire model space. Although it is inefficient and inaccurate in case of problems involving a large number of unknowns, this optimization method does not suffer from the premature convergence issue. Conversely, the convergence of the GAs to the global minimum is guaranteed by the Holland theorem (Holland, 1975) only in case of a population containing an infinite number of individuals; for finite populations, the convergence of GAs is not guaranteed. This characteristic is often called "genetic-drift". A more heuristic description of this phenomenon can be given in terms of population behavior: after some generations, the individuals of a population tend to converge toward a convex neighborhood of a minimum of the objective function (not necessarily the global minimum), and thus the exploration of other promising portions of the model space is prevented. In the worst case, the population cannot escape from such convex neighborhood, and a non-optimal solution is provided. In the past, many strategies have been proposed to reduce the genetic-drift effect. For example, a possible approach is the so-called niched genetic algorithm (NGA), in which the initial random population is divided into multiple subpopulations that are subjected to separate selection and evolution processes (Rey Horn, 1993). To maximize the exploration of the model space and to reduce the genetic drift, Aleardi and Mazzotti (2017) included into the standard NGA other evolutionary mechanisms, such as competition between subpopulations, and stretching of the fitness function. With the aim to recover the population diversity in case of premature convergence, Eldos (2008) included a catastrophic operator within the evolution of a standard GA. This operator randomly destroys a certain number of candidate solutions that are replaced by newly generated ones. Another proposed method is the so-called crowding GA

(Mengshoel and Goldberg, 2008; Mengshoel et al. 2014), in which the population diversity is preserved by pairing each offspring with a similar individual in the current population (pairing phase) and deciding which of the two will survive (replacement phase). Moreover, the genetic drift is not the only downside of GAs. Indeed, as pointed out in Sajeva et al. (2017a) GAs encounter problems in converging to the optimal solution in case of irregularly distributed minima in the objective function. To overcome or mitigate all these issues, we propose an innovative GA implementation that we name Drift-Avoidance Genetic Algorithm (DAGA), in which some of the previously mentioned optimization strategies are modified and combined (MCA, NGA, crowding GAs, catastrophic GAs).

The paper can be divided into two main sections. In the first, the performances of the DAGA approach are illustrated on analytic, multi-minima objective functions that are usually applied to test optimization algorithms. These functions mimic a wide range of possible objective functions that characterize geophysical optimization problems: for example, we consider multi-minima functions with regularly distributed local minima (the Rastrigin, and Ackley functions) and several multi-minima functions with irregularly distributed minima (the Schwefel, Eggholder, and Langermann functions). First, we demonstrate that the combination of the different principles coming from NGA, crowding GA, and catastrophic GA provide a new algorithm with superior convergence capabilities: the DAGA approach. Afterwards, we demonstrate that the performances of the DAGA approach far surpass the performances of the other GA implementations also in case of multidimensional model spaces.

In the second section, we apply the DAGA method to two geophysical optimization problems characterized by objective functions with very complex topologies: the residual statics corrections and the 2D acoustic full-waveform inversion (FWI). This part is aimed at demonstrating the superior convergence capabilities of the DAGA approach over a more standard NGA in case of highly non-linear geophysical optimization problems, and also its applicability in case of computer intensive forward-modeling evaluations.

Since the pioneering works by Rothman (1985, 1986), global optimization methods have been extensively applied to tackle the non-linearity of residual statics corrections (e.g., Wilson, and Vasudevan, 1991; Aleardi et al. 2016b; Eladj, 2016). In particular, in the context of Common-Mid-Point (CMP) consistent residual statics corrections, Sajeva et al. (2017a) demonstrated that the GA performances get worse as the number of unknown model parameters (that is the number of seismic traces within the CMP) increases.

FWI is another highly non-linear geophysical optimization problem characterized by ill-conditioning and multiple minima in the objective function (mainly related to the cycle-skipping phenomenon; Virieux and Operto, 2009). For this reason, an optimal starting model is crucial to attain convergence and to avoid entrapment in local minima when applying a gradient-based inversion. Over the last years, many strategies have been proposed to make the local FWI less influenced by the initial guess of the subsurface model (i.e., van Leeuwen et al. 2014; Choi and Alkhalifah 2015; Métivier et al. 2016; Warner and Guasch 2016). For what concerns the global approach to FWI, many authors (e.g. Datta and Sen 2016; Mazzotti et al. 2016; Sajeva et al. 2016; Galuzzi et al. 2017a; Sajeva et al. 2017b) have demonstrated the applicability of a hybrid optimization strategy in which a first run of global optimization is used to determine a suitable starting model that is subsequently used as input for the local, gradient-based inversion. However, even in these successful applications, particular care must be adopted to avoid premature convergence (e.g., adopting envelope-based objective functions or parsimonious reparameterizations of the subsurface).

In the following, we limit the attention to synthetic seismic tests in which the true model is perfectly known. This strategy will allow us to focus the discussion on the evaluation of the different algorithms and to draw essential conclusions. It has been demonstrated (e.g., Sajeva et al. 2017a) that GAs achieve good convergence rates even for strongly ill-conditioned optimization problems (i.e., problems characterized by objective function with small gradient values around the global minimum). The gradient of the objective function can also be related with the signal-to-noise ratio (SNR) of the observed data, because the gradient of the objective function decreases as the SNR decreases. In these

contexts, as confirmed by additional tests not shown here, DAGA performances are similar to those of other GA implementations. For this reason, in the following seismic experiments random noise is not added to the observed data. In addition, as in the first part on analytic objective functions, also these seismic examples are performed with different model-space dimensions and with objective functions with different topologies. As a final remark, we point out that the performances of any GA implementation may critically depend on the choice of the control parameters. For this reason, it is usually difficult to give hard and fast rules that may work with a wide range of applications, although some guidelines and rules of thumb can be dictated by experience (Aleardi and Mazzotti, 2017; Sajeva et al. 2017a); for example, the number of individuals in a population should increase with the dimension of the model space and with the complexity of the objective function. We refer the reader to De Jong (2007) for more information about the setting of GA parameters. The control parameters used in all the following examples have been determined from a trial-and-error procedure aimed at balancing the probability and rapidity of convergence with the goodness of the final solution. Some guidelines are given to illustrate how the most important DAGA parameters can be chosen.

A brief overview of genetic algorithms and niched genetic algorithms

Genetic Algorithms (Holland, 1975) are a class of global optimization methods that have been proven very effective in solving geophysical optimization problems. In GAs terminology, an individual (or chromosome) is a solution in the model space, whereas a population represents a set of individuals (i.e., an ensemble of possible solutions). Without going into mathematical details, we give in Figure 1 an outline of how the algorithm works. A very simple GA flow starts with the generation of a random population of individuals over which the fitness function (namely the goodness of each solution) is evaluated. The fitness value stochastically contributes to the selection of the best individuals for the reproduction step in which a set of new solutions (offspring) is generated by combinations of parent individuals. The offspring are mutated, their fitness is evaluated, then a new generation is created by replacing some of the parent individuals with the generated offspring. The

algorithm iterates until convergence conditions are satisfied (for example a maximum number of generations).

A more sophisticated version of GAs is the niched GA (NGA) that is inspired by the punctuated-equilibria evolutionary theory. In this case, the initial random population is divided into multiple subpopulations that undergo separated selection and evolution processes and only for given generations they can exchange some individuals (migration). It has been proved that this peculiar GA implementation is effective in attenuating the genetic drift effect (Horn, 1993).

Many different GA and NGA implementations exist. All the GA implementations adopted in this work are based on the real-coded breeder GA implementation described in Janikow and Michalewicz (1991) and Schlierkamp-Voosen and Muhlenbein (1993). More in detail, we use the roulette wheel as the selection method, the extended intermediate recombination as the crossover operator, and a real value mutation strategy. Additional information about the considered GA and NGA implementations can be also found in Sajeva et al (2017a) and in Aleardi and Mazzotti (2017). We refer the interested reader to Goldberg (1989), Mitchell (1998), and Sivanandam and Deepa (2008) for detailed theoretical information about genetic algorithms.

The DAGA method

With the aim to increase the exploration capabilities of GAs and avoid the genetic drift effect, we first hybridize a modified version of the catastrophic GA with NGAs and MCA. A catastrophe is a random phenomenon centered on a hit subpopulation that destroys a certain number of individuals, which have been stochastically selected. Depending on the radius of the catastrophe, neighbor subpopulations can be hit too. The next step aims to restore the number of individuals for each subpopulation to the original one before the catastrophic event (Figure 2). More in detail, the destroyed individuals are progressively replaced over the following generations by new ones migrating from other subpopulations, and by new individuals randomly generated with uniform probability over the entire model space (Figure 3). The probability of a single individual to migrate

is proportional to its fitness value: in other words, better individuals have higher probability to be assimilated in a subpopulation after a catastrophic event.

Let idx be the integer indexes that identify the subpopulations. The subpopulation center (idx_c) of the catastrophic event is stochastically selected based on the mean fitness values of the entire ensemble of subpopulations and using the roulette-wheel selection method (Sivanandam and Deepa, 2008): subpopulations with lower average fitness have higher probability to be hit by a catastrophic event. The radius of the catastrophe (i.e., the number of subpopulations hit by the catastrophe and centered around idx_c) is randomly determined from a Gaussian distribution with user-defined mean and variance values. The intensity (Int) of the catastrophic event (i.e., the number of individuals that will be removed from the subpopulation) for the subpopulation idx_c is randomly determined based on a Gaussian distribution with adaptive mean (N_d) and variance (σ) values:

$$Int(idx_c) = N(N_d, \sigma), \qquad (1)$$

where N identifies the Gaussian distribution, whereas N_d and σ are equal to:

$$N_d = N\_pop_0 \times e^{\left(-a\frac{gen}{gen\_max}\right)} \qquad (2),$$

$$\sigma = (N\_pop_0 - N_d) \times int_{level} \qquad (3),$$

respectively, where N_pop_0 is the original number of chromosomes in the considered subpopulation, $0 < a < 1$ is a user-defined parameter controlling the influence of the generation number on the N_d value, gen is the current generation, gen_max is the maximum number of generations, int_level is a user defined parameter controlling the intensity of the catastrophic event. Equations 2 and 3 make clear that the number of individuals destroyed by a catastrophic event decreases as the number of generations increases (i.e., as the final solution is approached). After determining $Int(idx_c)$, the intensity of the catastrophe in all the involved subpopulations is determined as:

$$Int(idx_s) = \frac{Int(idx_c)}{|idx_c - idx_s| + 1}, \qquad (4)$$

where idx_s represents the indexes of the subpopulations hit by the catastrophe. Note that in case of a catastrophic event, the best chromosomes in each subpopulation are always preserved to ensure that the optimization does not waste time re-discovering previously promising solutions (elitist strategy).

In case of genetic drift, the genetic diversity is lost, or in other words, the entire population contains several near-copies of the same individual. In this case, a considerable amount of time may be wasted in performing forward modelling for very similar models. To tackle this issue, we include in our algorithm a modified version of the crowding strategy in which random replacements are systematically performed in each generation (Figure 4). We measure the L2-norm distance for each pair of individuals in each subpopulation. Then, let t be a user-defined parameter; the algorithm identifies t couples of similar individuals for each subpopulation and select which individual must be preserved. Obviously, the probability for an individual to be preserved is proportional to its fitness values. More in detail, the roulette-wheel selection method is used to identify the individuals to be preserved. The non-selected individual is replaced with a randomly generated solution with uniform probability over the entire model space. Note that to reduce the computational cost of the algorithm, we apply the modified crowding function only if the mean distance between the chromosomes of the considered subpopulation is smaller than a previously selected threshold value. Similarly to the catastrophic event, we again preserve the best individuals. This peculiar implementation maintains the genetic diversity within each subpopulation and maximizes the exploration of the model space.

The previous considerations make it clear that the DAGA approach (schematized in Figure 5) is mainly aimed at increasing the exploration capability of a standard NGA and at efficiently exploring the most promising portions of the model space. Obviously, in applying the DAGA method a good compromise must be found between exploration and exploitation of the algorithm. In practical applications, this translates in finding (usually by a trial and error procedure) an optimal set of user-defined parameters for the problem at hand. Indeed, a too strong exploration will heavily increase the computational time and slow down the convergence, while a too strong exploitation will possibly result in premature convergence toward sub-optimal solutions.

The DAGA implementation also requires additional user-defined parameters to be set with respect to the standard GA. Among these parameters, we found that the catastrophe probability and the number of replacements play the major roles in controlling the convergence of the algorithms. On the one hand, the values for these parameters should increase as the complexity of the objective function increases. On the other hand, these values must be kept as small as possible not to increase drastically the computational cost of the DAGA optimization. However, we found that a catastrophe probability equal to 0.2-0.3, and a number of replacements between 1 and 10 are suitable for most applications. We have also found that the other parameters N_d, σ, *int_level*, *a* (see equations 1, 2, and 3) exert a much weaker influences on the DAGA performances and for this reason their values are keep fixed in all the following experiments at 2, 0.6, 0.6, and 0.9, respectively.

As a final remark, note that the inclusion of Monte Carlo principles (i.e., inclusion of new randomly generated solutions) into the genetic algorithm optimization scheme, could result in an increased computational effort related to the increased number of forward model evaluations. However, in the following we will also demonstrate that the DAGA often has a faster converge rate than, for example, the standard NGA. In other words, the DAGA approach often requires a lower number of generations (i.e. forward-model evaluations) to converge than other GA implementations.

A heuristic example of genetic drift

We now discuss a simple optimization test that offers a heuristic but clear example of genetic drift. For this test, we run the standard GA method on the Schwefel analytic function (Appendix A) considering a 2D model space. We adopt a single population formed by 10 individuals with a selection rate of 0.8 (that is the 80% of individuals of the current population are selected for reproduction), running for 100 iterations. In Figure 6a, we represent the 2D Schwefel function with superimposed the initial uniformly at random generated population, and the best GA model provided at the end of the optimization process. In Figure 6b, we represent the L2-norm distance between the current best model and the global minimum as the number of generations increases, whereas Figure 6c shows the

evolution of the average L2-norm distance between the individuals in the current population. Figure 6 clearly points out that the best model moves away from the global minimum and eventually converges toward a local minimum. In particular, the GA population remains on the same local minimum between generations 15-40, and it converges toward the final solution after 41 iterations. The significant decrease of the average distance between the chromosomes is a clear indication of genetic drift and premature convergence. This example proves that the genetic drift could constitute a major issue even in case of optimization problems with a limited number of unknowns. This characteristic is even more serious in case of field-data optimizations in which the position of the global minimum is not known a-priori. In these cases, it would be advisable performing different inversion runs from which the model resulting in the minimum objective function value is selected as the final result. Obviously, this strategy is not feasible in case of optimization problems with expensive forward modelling. In these cases, a GA implementation able to prevent the genetic drift effect is crucial to increase our confidence in the final estimated model and to maintain the computational cost affordable.

Test on analytic functions

In this section, we evaluate the convergence capabilities of the implemented DAGA approach using analytic test functions often employed to test optimization methods. The performances of the proposed approach are also compared with other GA implementations and with the MCA approach. For expositional convenience, the analytic functions we consider are described in Appendix A. In the following, we adopt the same convergence criterion for all the algorithms and for all the tests. In particular, an algorithm converges when it finds an individual (x) that satisfies the following accuracy criterion:

$$\sqrt{\frac{\sum_{i=1}^{n}(x_i - x_i^{glob})^2}{n}} < \epsilon, \quad (5)$$

where n represent the model-space dimension, $\boldsymbol{x}^{glob}$ is the global minimum, and the accuracy ϵ is set to 0.01 in all the tests. The optimization is stopped when the solution satisfies the selected accuracy criterion or when the maximum number of generations is reached (e.g., 200 generations for the 2D Schwefel function). To obtain statistically significant results, we perform 100 tests for each method and for each objective function, and for each generation we progressively count the number of tests that have attained convergence thus far. The results are represented as incremental curves showing the cumulative percentage of successful tests for each generation.

We first compare the performances of GA, NGA, crowding GA, catastrophic GA, and DAGA with the aim to demonstrate the superior convergence capabilities of the proposed method. In this case, we employ the Rastrigin, Schwefel, and Eggholder functions: The Rastrigin function has a large number of regularly distributed local minima; the Schwefel function is characterized by irregularly distributed minima and by a global minimum located at the border of the search space; the Eggholder function can be described as a combination of the Rastrigin and Schwefel functions. Note that Sajeva et al. (2017a) discussed the difficulties encountered by a standard GA in finding the global minimum when it is located at the lateral edge of the search space and distant from the other local minima. The same work also demonstrated the superior performance of GAs compared with other popular global optimizations algorithms in case of regularly distributed minima in the objective function. For this reason, the aim of the following tests is two-fold: Demonstrating the superior performances of the DAGA approach in case of irregularly distributed minima; showing that in case of regularly distributed minima the DAGA performances are comparable with those of other GA implementations. The main control parameters used in the 2D tests are given in Table 1. Note that the same parameter value is used by all the GA implementations that consider such parameter. For example, a number of chromosomes equal to 6 is used by all the considered GA implementations for optimizing the 2D Rastrigin function or a catastrophe probability value of 0.2 is used by catastrophic GA and DAGA for the 2D Eggholder function.

In the first test, we use the 2D Rastrigin function. In this case, we observe that all the considered GA implementations yield comparable performances (Figure 7a). However, even in this example we can observe that only the crowding GA and the DAGA find the global minimum in all the tests, whereas the other GA implementations (NGA, GA, and catastrophic GA) show a lower number of successes (around 95-97%). Despite these minor differences, we can claim that this example confirms the efficiency of all the considered GA implementations in finding the global minimum in case of regularly distributed minima. Figure 7b represents the average number of models evaluated by each algorithm until the end of the optimization process. We note that GA, NGA, DAGA, and crowding GA, generate a similar number of models, whereas the catastrophic GA requires a smaller number of models to find the global minimum: in other terms, catastrophic GAs achieve a faster convergence.

We now discuss the results obtained for the more challenging 2D Schwefel function (Figure 8). In this case, the standard GA converges only in 32 out of 100 tests, whereas NGA, catastrophic GA, and crowding GA outperform the standard GA, but none converges in all the 100 tests. Differently, the proposed DAGA approach achieves a 100% success rate. The average number of evaluated models (Figure 8b) shows that among all the considered GA implementations, the DAGA and the crowding GA require the lowest number of models to satisfy the convergence criterion. In other words, these two implementations achieve a faster convergence than the other GA approaches in case of objective functions with irregularly distributed minima.

Figure 9 refers to the Eggholder function. The results clearly prove that the performances of the DAGA method far surpass those of the other GA approaches, and that the proposed implementation is the only method with a 100% success rate. Similarly to the previous test on the Schwefel function, the DAGA requires a small number of models to achieve convergence with respect to the other optimization strategies. Also note that NGA, GA, and catastrophic GA are characterized by a faster exploitation than DAGA and crowding GA (i.e., note the steeper cumulative curves in the first 100 generations; Figure 9a) but they eventually get trapped into a local minimum (i.e., their success rates do not significantly change after 100 generations; Figure 9a).

All the previous examples not only prove that the DAGA outperforms all the other considered GA implementations, but also demonstrate the superior convergence capabilities of NGA, crowding GA, and catastrophic GA over the standard GA approach. For this reason, these more advanced GA approaches should be preferred instead of a standard GA in case of optimization problems characterized by objective functions with complex topologies.

The following tests are aimed at demonstrating the convergence capabilities of the DAGA approach when the dimension of the model space increases. Indeed, it is well known that the ability of a global optimization method to find the global minimum exponentially decreases as the number of unknowns increases (curse of dimensionality problem). To this aim, we employ the Rastrgin, Ackley, Schwefel, and Langermann functions. In particular, the Langermann function is extremely difficult to optimize due to the high number of local minima unevenly distributed over the model space. The Ackley function is similar to the Rastrigin, being characterized by regularly distributed minima over the entire model space. For the sake of conciseness, we focus the attention on the DAGA, NGA, and MCA approaches only. Differently from the previous tests, we now perform 50 inversion runs for each dimension and for each method. We again present the final results as incremental curves showing the percentage of successful tests until a given generation. Actually, the MCA approach is not characterized by a population evolving for a given number of generations. However, for comparability, the MCA results are represented versus the number of generations: in a single MCA generation, we evaluate the same number of models sampled by the DAGA for the same generation number. The principal control parameters used in the 2D tests are given in Table 2.

The results for the Rastrigin (Figure 10) and Ackley (Figure 11) functions confirm that both NGA and DAGA are able to converge in case of regularly distributed minima. For example, both the DAGA and NGA show comparable performances in the Ackley function for the 2D, 3D, 4D, and 10D tests. However, we can still note that the DAGA has a success rate of the 100% for all the dimensions, whereas the NGA shows a success rate around 98%. As expected the MCA approach achieves success

rates much lower than NGA and DAGA and equal, for example, to 2%, 0%, and 0% for the 3D, 4D, and 10D Rastrigin functions, respectively.

For the Langermann function (Figure 12), the DAGA approach achieves success rates of 100%, 98%, and 96%, in the 2D, 3D, and 4D tests, respectively. Differently, the NGA is characterized by a success rate of around 60% for all the dimensions, whereas the MCA successfully converge in 10, 0, and 0 tests out of 50 tests in the 2D, 3D, and 4D experiments, respectively.

Finally, the tests of Schwefel function (Figure 13) still demonstrate the superior convergence capabilities of the DAGA approach compared to NGA. For the 3D, 4D, and 10D the DAGA converges in 49, 50, and 47 out of 50 tests. Conversely, the NGA finds the global minimum in only 24, 23, and 19 tests in the 3D, 4D, and 10D cases. In this example, none of the MCA runs satisfy the convergence criterion. Figures 10-13 prove that the implemented method outperforms the NGA approach not only for limited model space dimensions but, more importantly, also for multi-dimensional model spaces and objective functions with very complex topologies. In addition, the comparison of DAGA and MCA proves that the generation of new random solutions alone does not guarantee the convergence toward the global minimum. Indeed, in the DAGA implementation the improvement of the current solution (i.e., the decrease of the objective function value associated to the best individual) is usually not simultaneous to the occurrence of a catastrophic event but is mainly related to the ability of the DAGA to process the new individuals to find a new and more promising minimum.

Residual statics corrections

In the following seismic examples, we limit the attention to the DAGA and NGA approaches only. In what follows, we perform CMP-consistent residual statics corrections on a synthetic seismic gather. Sajeva et al. (2017a) showed that this optimization problem is characterized by an objective function that shows some similarities to both the Rastrigin and Schwefel functions. For this reason,

we expect that the GA performances progressively worsen as the dimensions of the model space increase.

We use actual well-log information and a 1D convolutional forward modelling based on the exact Zoeppritz equations to generate the reference CMP gather (without residual statics). To simulate residual statics in the data, we apply to each trace in the reference CMP random time shifts uniformly distributed within the range -15/+15 ms, whereas in the subsequent optimization process we allow time shifts within the range -25/+25 ms. Similarly to the previous tests on the analytic functions, we analyses the performance of the two methods as the number of unknown model parameters increases. In this case, the unknowns are the time shifts that must be applied trace-by-trace to the CMP to maximize the energy of the associated stack trace. We discuss two cases in which the CMP gather is constituted by 50, and 80 traces, corresponding to 50, and 80 unknown time shifts (Figure 14, and Figure 15, respectively). For comparability, both the NGA and DAGA optimization starts from the same random initial population. The DAGA and NGA parameters used in the following tests are given in Table 3.

For the 50-trace example (Figure 14) the DAGA method recovers a final CMP very similar to the reference one and, more importantly, not affected by severe cycle-skipping problems. Conversely, we observe that the CMP recovered by NGA shows many misalignments in the reflections and cycle-skipped traces: this results in a stack energy much lower than that of the reference CMP (Figure 16a). Finally, in the 80-trace example (Figure 15) the DAGA approach outperforms the NGA method and again provides a final stack trace with higher energy (Figure 16c).

These examples further demonstrate that the NGA approach is prone to converge toward a local minimum in case of objective functions with complex topologies (irregularly distributed minima), and that the probability to get trapped into a local minimum increases with the dimension of the model space. Conversely, the DAGA implementation shows superior performances in all the tests, as it provides optimal solutions even for high-dimensional model spaces. The fast forward modelling associated with the residual statics corrections makes the computational costs of NGA and DAGA

very similar. Indeed, in average, in all the previous tests DAGA resulted to be only 1.11 times slower than NGA. However, note that if we select as the stopping criterion a threshold value for the energy of the stack of the corrected CMP, DAGA will probably result faster than NGA as it is characterized by a faster convergence rate. In other words, the DAGA reaches a given stack energy value in a smaller number of generations than NGA. As shown in Figures 16b and 16d, where we plot the energy of the stack trace estimated after DAGA and NGA versus the number of forward model evaluations, we always observe that for a given number of forward modelling the DAGA provides a stack trace with higher energy than NGA. For example, in Figure 16b the stack of the DAGA corrected traces reaches an energy value around 750 after 20000 sampled models, whereas the NGA requires more than 50000 forward evaluations to attain the same energy value.

As a final remark note that differently from the NGA-CMP, the final DAGA-CMPs are not affected by severe cycle-skipping issues and for this reason they could be used as valid starting models for any local optimization method for a further refinement of the residual statics estimation.

2D acoustic full-waveform inversion

In this section, we compare the DAGA and NGA performances on an acoustic global FWI aimed at retrieving the long-wavelength structure of the subsurface P-wave (Vp) velocity field. To reduce the computational cost, we limit the attention to a small portion of the acoustic Marmousi model (Figure 17a). For both NGA and DAGA, we adopt the global FWI approach described in Sajeva et al. (2016; 2017b) and Mazzotti et al. (2016), which discretizes the subsurface model with two grids: a "coarse" grid for the inversion and a "fine" grid for the modelling. Each node of the inversion grid corresponds to an unknown parameter of the global optimization: in our acoustic example the P-wave velocity at that node. Obviously, the number of cells in the coarse grid determines the number of unknowns in the global inversion. For this reason, the grid must be appropriately chosen to make the global optimization feasible. A simple bilinear interpolation is used to convert the coarse grid to the fine grid. In the following, the synthetic seismic data are generated by employing a finite-difference

acoustic code (Galuzzi et al. 2017b) with an accuracy of the fourth order in space and of the second order in time.

For the sake of consistency with the previous examples, we perform two different tests characterized by different topologies of the objective function, that in both cases is defined by the L2-norm difference between observed and predicted seismic data. In the first test, we employ a 7-Hz Ricker wavelet as the source signature (that is assumed perfectly known during the inversion) and a regular inversion grid with 55 nodes (11 along the horizontal directions and 5 along the vertical direction; Figure 17b). In the second test, we employ the same inversion grid, but we increase the dominant frequency of the source signature up to 15 Hz. Notwithstanding the same acquisition geometry (17 sources and 50 receivers equally spaced along the model and towed at 25 m below the sea level), the same number of nodes in the inversion grid, and the same 1D search ranges (Figure 17c), the different frequency ranges make these two optimization examples crucially different. Indeed, the presence of very low frequencies in the observed seismic data makes the first example less prone to cycle-skipping and for this reason we expect good convergence capabilities for both the DAGA and NGA methods. Differently, in the second test the higher dominant frequency of the source signature and the lack of low frequencies increase the probability for the optimization to get trapped into a local minimum. Then, in this case we expect that the superior exploration capabilities of the DAGA implementation make this method better suited than NGA to tackle this optimization problem. To obtain comparable results, both the NGA and DAGA start from the same initial random population. Table 4 shows the most important user-defined parameters employed in the FWI tests.

Figure 18 shows a comparison for the 7-Hz test between an observed shot gather and the same shot gather predicted by NGA and DAGA. Note that both algorithms achieve fair matches between the observed and the predicted data. As expected by the smooth and long-wavelength structure of the predicted *Vp* model, the amplitudes and the kinematics of the diving wavefield are better recovered than those of the reflected events.

Figures 19a-b represent the NGA and DAGA best models for the 7-Hz test. Both approaches achieve similar predictions that fairly represent the long-wavelength structure of the true model. As expected, the quality of the predictions decreases at the lateral edges of the model where the seismic illumination is poor. Figures 19c-d demonstrate that both algorithms yield comparable final data-misfit and model-misfit values, where the model misfit is computed as the average L2 norm distance between the current best model and the true model. Note that the higher exploration capabilities of the DAGA approach produce a more scattered trend for the model misfit evolution. The non-perfect correlation between data- and model-misfit values is mainly related to the high non-linearity of the FWI problem.

For a more quantitative assessment of the final predicted Vp macro-models, we use the DAGA and NGA outcomes to compute the amount of cycle-skipping affecting each seismic trace in the predicted shot gathers. To this end, we employ the method proposed by Shah et al. (2012) that limits the attention to the first breaks because these transmitted events are crucial to reconstruct the large and intermediate wavelengths of the subsurface model (Mora, 1989). More in detail, this method analyses the first arrivals of the data and computes at a given frequency and for each source-receiver pair the phase differences between the observed data and the data generated on the predicted subsurface model. In this residual-phase domain, cycle skipping is evidenced by a sudden 2π jump.

For the 7-Hz test the phase-residual panels are shown in Figures 19e-f. We observe similar results for the NGA and DAGA methods with no cycle-skipping issues. This confirms the suitability of both approaches to find a Vp macro-model of the subsurface in case of seismic data rich in low frequencies.

The data comparison (Figure 20) for the 15-Hz experiments shows that the subsurface model estimated by the DAGA method achieves a closer match with the observed data, especially at far source-receiver offsets. Figures 21a-b illustrate the final NGA and DAGA models predicted for the 15-Hz case. In this example we note that the DAGA yields a final model with a closer match with the long-wavelength structure of the true model, whereas the NGA provides a final prediction with several artefacts (i.e., circular low- and high-velocity anomalies). Figures 21c-d demonstrate that the

DAGA final prediction is characterized by lower data-misfit and model-misfit values than the NGA outcome. Notably, we observe that in the last NGA 50 iterations, although the data misfit slightly decreases, the associated model misfit significantly increases. This clearly indicates that the NGA is exploiting a local minimum of the objective function.

The phase-residual panels (Figures 21e-f) give a further and more quantitative demonstration that the DAGA implementation provides a final Vp field less affected by cycle-skipping issues than the model predicted by NGA. These residual cycle skipping issues could be mitigated by exploiting the stronger exploration capability of the DAGA implementation. In other terms, additional generations could be performed with the aim to sample a final predicted DAGA model closer to the optimal one. However, it likely may be more efficient to further refine the predictions by applying a local, gradient-based FWI method.

For these tests we employ a parallel, hybrid, Matlab-C-Bash code running on ten compute nodes equipped with two deca-core intel E5-2630 @2.2 GHz (128 Gb RAM). This results in a total computational cost for the DAGA method of 3.1 and 6.7 hours for the 7-Hz and 15-Hz tests, respectively, and of 2.1 and 4.5 hours for the NGA in the 7-Hz and 15-Hz experiments, respectively. Note that in these tests the stopping criterion for both algorithms is the maximum number of generations (400 and 300 for the 7-Hz and 15-Hz experiments, respectively). If we had assumed as the stopping criterion the achievement of the same data misfit for both DAGA and NGA, then the computational costs for the two experiments would had been greater for DAGA in the 7-Hz test, and lower for DAGA in the 15-Hz case. In fact, in the 7-Hz experiment, both methods reach the same data misfit in about the same number of generations (see Figure 19c), and DAGA is burdened by the extra costs of the additional operations (crowding, catastrophic operator, and so on). Instead, similarly to the residual statics example, the significantly faster convergence rate of DAGA for the 15-Hz experiment (see Figure 21c) renders it possible for DAGA to be run for a lower number of generations, thus counter-balancing its extra-cost.

Conclusions

We presented an innovative strategy to attenuate the genetic drift and to increase the exploration of the model space in a genetic algorithm (GA) optimization. The proposed Drift Avoidance Genetic Algorithm (DAGA) hybridizes different GA implementations (i.e., niched and modified versions of catastrophic and crowding GAs) and a standard Monte Carlo algorithm (MCA). The tests on analytic objective functions, demonstrated the superior convergence capabilities of the DAGA approach against other popular GA implementations (standard GA, niched GA, catastrophic GA, and crowding GA) in case of objective functions with very complex topology (i.e., many irregularly distributed local minima). Differently, in cases of simpler topologies (i.e., regularly distributed minima) the DAGA algorithm and the other GA approaches achieved very similar performances. These tests also demonstrated the capability of the DAGA approach of finding the global minimum in multi-dimensional model spaces. The seismic synthetic tests on residual statics corrections and 2D acoustic full-waveform inversion confirmed that, differently from the niched GA implementation, the DAGA approach efficiently attenuates the genetic drift effect and accurately converge toward optimal solutions also in optimization problems with a significant number of unknown model parameters and complex objective function topologies.

Due to the increased number of forward-model evaluations requested by the modified crowding and catastrophic operators, the implemented DAGA expends additional computational resources compared to the other considered GA approaches. However, the DAGA often converges faster than the other GA methods, thus requiring a fewer number of generations (and then a fewer number of forward modellings).

The DAGA approach requires the appropriate setting of additional control parameters with respect to the standard GA. However, in our tests we found that only the catastrophe probability and the number of replacements markedly influence the DAGA performances. As a rule of thumb, we suggest that a catastrophe probability around 0.2-0.3 and a number of replacements between 1 and 10, should

work in most applications. In general, the values of these parameters should increase with the expected complexity of the objective function.

A possible improvement of the actual DAGA implementation is to substitute the random replacements of individuals (e.g., after the catastrophic events) with more sophisticated strategies that exploit the already evaluated models and/or some form of a-priori information to direct the exploration toward the most promising portions of the model space. This strategy may speed up the optimization process and may further attenuate the curse of dimensionality issue.

Appendix A

Here we describe the analytical test functions used in the first part of the work.

Schwefel function

The Schwefel function can be computed as:

$$f(\mathbf{x}) = -x_i \sum_{i=1}^{n} [x_i * \sin(\sqrt{|x_i|})]. \quad (6)$$

Similarly to the Rastrigin function, the Schwefel function (Figure A1a) has a large number of local minima equal to 7^n. However, differently from the Rastrigin function, in which the local minima surround the central global minimum, in the Schwefel function, the local minima are more irregularly distributed, and important local minima are distant from the non-centred global minimum, which lies at [420.9687, …, 420.9687], or are even located at the opposite edge of the model space. The model space we consider is $[-500, 500]^n$.

Rastrigin function

the Rastrigin function can be derived as:

$$f(\mathbf{x}) = A\,n + \sum_{i=1}^{n} [x_i^2 - A\cos(2\pi x_i)], \quad (7)$$

where $A = 10$ and n is the dimension of the model space. Figure A1b shows the Rastrigin function in two dimensions. This function is a typical example of a non-convex function, with a global minimum located in $[0, …, 0]^n$ and a high number of regularly distributed local minima, which increases exponentially with the dimension of the model space. The model space we consider is $[-5, 5]^n$ in which there are 11^n local minima. The high number of local minima makes a local method inapplicable to optimize this function.

Langermann function

The Langermann function (Figure A1c) is a multimodal test function defined as follows:

$$f(\mathbf{x}) = -\sum_{i=1}^{m} c_i e^{\left(-\frac{1}{\pi}\sum_{j=1}^{n}(x_j - A_{ij})^2\right)} \cos\left(\pi \sum_{j=1}^{n}(x_j - A_{ij})^2\right), \quad (8)$$

where A and c are defined as:

$$A = \begin{bmatrix} 9.681 & 0.667 & 4.783 & 9.095 \\ 9.4 & 2.041 & 3.788 & 7.931 \\ 8.025 & 9.152 & 5.114 & 7.621 \\ 2.196 & 0.415 & 5.649 & 6.979 \\ 8.074 & 8.777 & 3.467 & 1.863 \\ 7.650 & 5.658 & 0.720 & 2.764 \\ 1.256 & 3.605 & 8.623 & 6.905 \\ 8.314 & 2.261 & 4.224 & 1.781 \\ 0.226 & 8.858 & 1.420 & 0.945 \\ 7.305 & 2.228 & 1.242 & 5.928 \end{bmatrix}, \quad (9)$$

$$c = \begin{bmatrix} 0.806 \\ 0.517 \\ 1.500 \\ 0.908 \\ 0.965 \\ 0.669 \\ 0.524 \\ 0.902 \\ 0.531 \\ 0.876 \end{bmatrix}, \quad (10)$$

and m represents the number of rows of A and c. This is an extremely difficult function to optimize because of the many and unevenly distributed local minima. In this case the position of the global minimum changes with the model space dimension: It is located in [8.06550, 9.03821], in [8.0328, 9.0927, 4.8534] and in [8.0250, 9.1520, 5.1140, 7.6210] for 2D, 3D and 4D model spaces, respectively. The model space we consider is $[0, 10]^n$.

Ackley function

The Ackley function is shown Figure A1d and is defined as:

$$f(\mathbf{x}) = -\alpha e^{\left(-\beta\sqrt{\frac{1}{n}\sum_{i=1}^{n}x_i^2}\right)} - e^{\left(\frac{1}{n}\sum_{i=1}^{n}\cos(\gamma x_i)\right)} + \alpha + e, \quad (11)$$

where $\alpha=20$, $\beta=0.2$ and $\gamma=2\pi$. It is another multidimensional function with regularly distributed minima characterized by a nearly flat outer region and a large hole at the centre. The function poses a risk for optimization algorithms, particularly hill-climbing algorithms, to be trapped in one of its

many local minima. The global minimum is located in $[0, …, 0]$ and the model space we consider is $[-32.768, 32.768]^n$.

Eggholder function

This function (Figure A1e) is only defined for a 2D space and can be computed as:

$$f(\mathbf{x}) = -(x_2 + 47)\sin\left(\sqrt{\left|x_2 + \frac{x_1}{2} + 47\right|}\right) - x_1\sin\left(\sqrt{|x_1 - (x_2 + 47)|}\right). \quad (12)$$

The global minimum is in $[512.404, 404.2319]$. This function is extremely difficult to optimize due to the many irregularly distributed local minima and the high gradient surrounding the global minimum. The model space we consider is $[-512, 512]^n$.

References

Aleardi, M., and A. Mazzotti, 2017, 1D elastic full-waveform inversion and uncertainty estimation by means of a hybrid genetic algorithm–Gibbs sampler approach: Geophysical Prospecting, **65** no.1, 64-85.

Aleardi M., A. Tognarelli, and A. Mazzotti, 2016a, Characterisation of shallow marine sediments using high-resolution velocity analysis and genetic-algorithm-driven 1D elastic full-waveform inversion: Near Surface Geophysics, **14** no.5, 449-460.

Aleardi, M., E. Stucchi, A. Sajeva, and B. Galuzzi, 2016b, Surface-consistent Residual Statics Estimation with Genetic Algorithms-An Application to a Near-surface Seismic Survey: Near Surface Geoscience 2016-22nd European Meeting of Environmental and Engineering Geophysics.

Aleardi, M., and F. Ciabarri, 2017, Assessment of different approaches to rock-physics modeling: A case study from offshore Nile Delta: Geophysics, **82** no.1, MR15-MR25.

Aleardi, M., 2015, Seismic velocity estimation from well log data with genetic algorithms in comparison to neural networks and multilinear approaches: Journal of Applied Geophysics, **117**, 13-22.

Choi, Y., and T. Alkhalifah, 2015, Unwrapped phase inversion with an exponential damping: Geophysics, **80** no.5, R251-R264.

Datta, D., and M. K. Sen, 2016, Estimating a starting model for full-waveform inversion using a global optimization method: Geophysics, **81** no.4, R211-R223.

De Jong, K., 2007, Parameter setting in EAs: a 30 year perspective, Parameter setting in evolutionary algorithms: Springer, Berlin, Heidelberg.

Eladj, S., 2016, Evaluation of Residual Static Corrections by Hybrid Genetic Algorithm Steepest Ascent Autostatics Inversion. Application southern Algerian fields: EGU General Assembly Conference Abstracts, 18, 7205.

Eldos, T., 2008, Distributed genetic algorithms: a scheme for genetic drift avoidance: Journal of Electrical Engineering, **59** no.1, 45.

Galuzzi, B., E. Zampieri, and E. Stucchi, 2017a, Global Optimization Procedure to Estimate a Starting Velocity Model for Local Full Waveform Inversion: International Conference on Optimization and Decision Science, 171-179.

Galuzzi, B., E. Zampieri, and E. Stucchi, 2017b, A local adaptive method for the numerical approximation in seismic wave modelling: Communications in Applied and Industrial Mathematics, **8** no. 1, 265-281.

Goldberg, D. E., and P. Segrest, 1987, Finite Markov chain analysis of genetic algorithms: In Proceedings of the second international conference on genetic algorithms, 1,1.

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

Li, T., and S. Mallick, 2015, Multicomponent, multi-azimuth pre-stack seismic waveform inversion for azimuthally anisotropic media using a parallel and computationally efficient non-dominated sorting genetic algorithm: Geophysical Journal International, **200** no.2, 1134-1152

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.

Mallick, S., and N. C. Dutta 2002, Shallow water flow prediction using prestack waveform inversion of conventional 3D seismic data and rock modelling: The Leading Edge, **21** no.7, 675-680.

Mazzotti, A., N. Bienati, E. Stucchi, A. Tognarelli, M. Aleardi, and A. Sajeva, 2016, Two-grid genetic algorithm full-waveform inversion: The Leading Edge, **35** no.12, 1068-1075.

Mengshoel, O. J., and D. E. Goldberg, 2008, The crowding approach to niching in genetic algorithms: Evolutionary computation, **16** no. 3, 315-354.

Mengshoel, O. J., S. F. Galán, and A. De Dios, 2014, Adaptive generalized crowding for genetic algorithms: Information Sciences, **258**, 140-159.

Métivier, L., R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2016, Increasing the robustness and applicability of full-waveform inversion: An optimal transport distance strategy: The Leading Edge, **35** no.12, 1060-1067.

Mora, P., 1989, Inversion= migration + tomography: Geophysics, **54**, 1575-1586.

Padhi, A., and S. Mallick, 2014, Multicomponent Pre-Stack Seismic Waveform Inversion in Transversely Isotropic Media Using a Non-Dominated Sorting Genetic Algorithm: Geophysical Journal International **196** no**.** 3, 1600–1618.

Rey Horn, J., 1993, Finite Markov chain analysis of genetic algorithms with niching: Forrest, **727**, 110-117.

Rothman, D.H., 1985, Nonlinear inversion, statistical mechanics, and residual statics estimation: Geophysics, **50** no.12, 2784-2796.

Rothman, D.H., 1986, Automatic estimation of large residual statics corrections: Geophysics, **51** no.2, 332-346.

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

Sajeva, A., M. Aleardi, E. Stucchi, N. Bienati, and A. Mazzotti, 2016, Estimation of acoustic macro models using a genetic full-waveform inversion: Application to the Marmousi model: Geophysics, **81** no.4, R173-R184.

Sajeva, A., M. Aleardi, B. Galuzzi, E. Stucchi, E. Spadavecchia, and A. Mazzotti, 2017a, Comparing the performances of four stochastic optimisation methods using analytic objective

functions, 1D elastic full-waveform inversion, and residual static computation: Geophysical Prospecting. **65** no. S1, 322-346.

Sajeva, A., M. Aleardi, and A. Mazzotti, 2017b, Genetic algorithm full-waveform inversion: uncertainty estimation and validation of the results: Bollettino di Geofisica Teorica ed Applicata, **58** no.4, 395-414.

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

Sen, M. K., and P. L. Stoffa, 1992, Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion: Geophysical Journal International **108** no.1, 281-292.

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.

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: 2012 SEG Annual Meeting. Society of Exploration Geophysicists, 1-5.

Sivanandam, S.N., and S. N. Deepa, 2008, Genetic algorithm optimization problems. Introduction to Genetic Algorithms: Springer.

Van Leeuwen, T., F. J. Herrmann, and B. Peters, 2014, A New Take on FWI-Wavefield Reconstruction Inversion: 76th EAGE Conference and Exhibition 2014.

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

Warner, M., and L. Guasch, 2016, Adaptive waveform inversion: Theory: Geophysics, **81** no.6, R429-R445.

Wilson, W. G., and K. Vasudevan, 1991, Application of the genetic algorithm to residual statics estimation: Geophysical Research Letters, **18** no.12, 2181-2184.

Figures, Tables and Captions
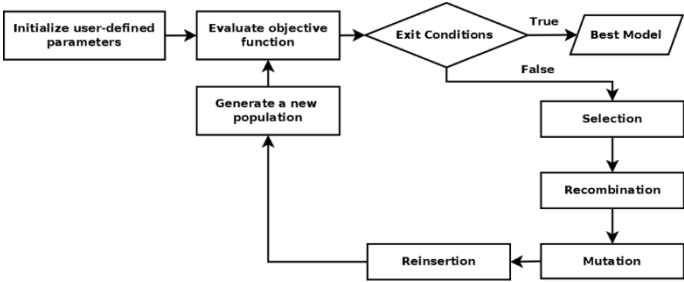
Figure 1: Example of a standard GA flow.
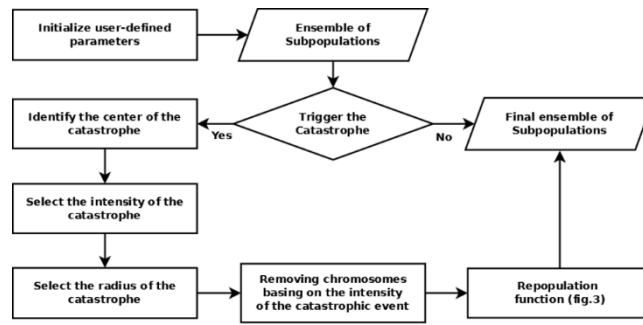
Figure 2: Scheme representing the implemented catastrophic function.
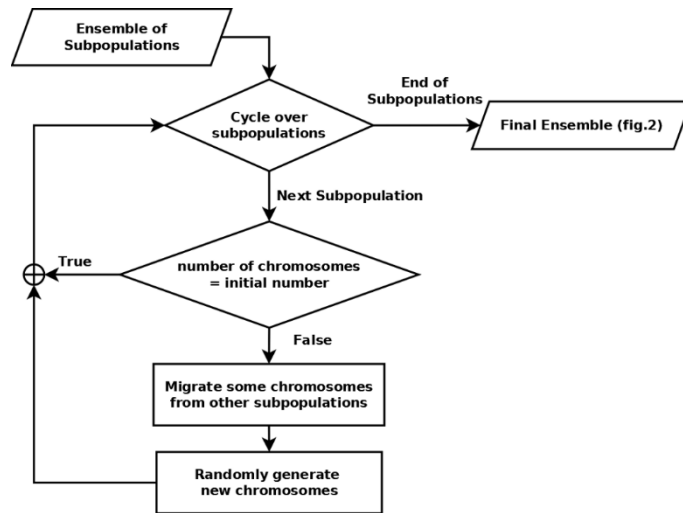
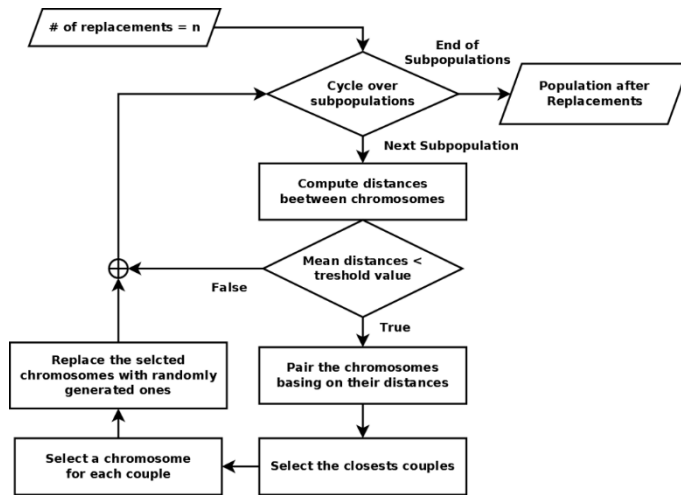*Figure 3: Scheme representing the repopulation strategy after a catastrophic event.*

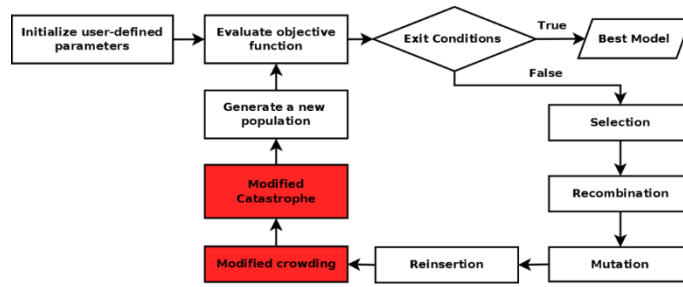*Figure 4: Scheme representing the implemented crowding function.*

Figure 5: Scheme representing the of DAGA approach. The red rectangles identify the DAGA functions added to the standard GA flow.
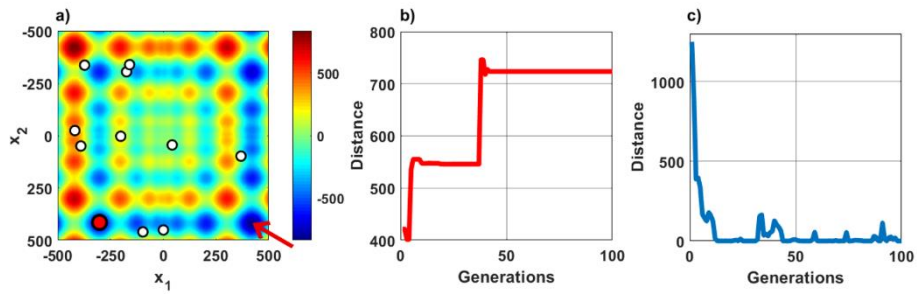
Figure 6: GA test on the 2D Schwefel function. a) 2D Schwefel function with superimposed the initial random population (white dots), and the final solution (red dot). The red arrow points toward the global minimum located at [420.9687, 420.9687]. b) Evolution of the L2-norm distance between the best model and the global minimum. c) Average-L2 norm distance between the chromosomes within the evolving population.
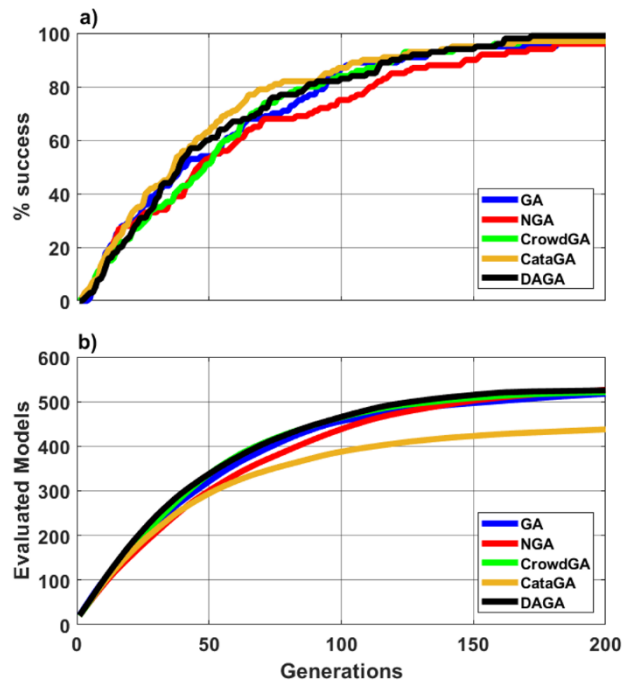
Figure 7: Test on the 2D Rastrigin function. a) Cumulative curves representing the percentage of successful tests over 200 generations. b) Number of evaluated models by each considered GA implementation.
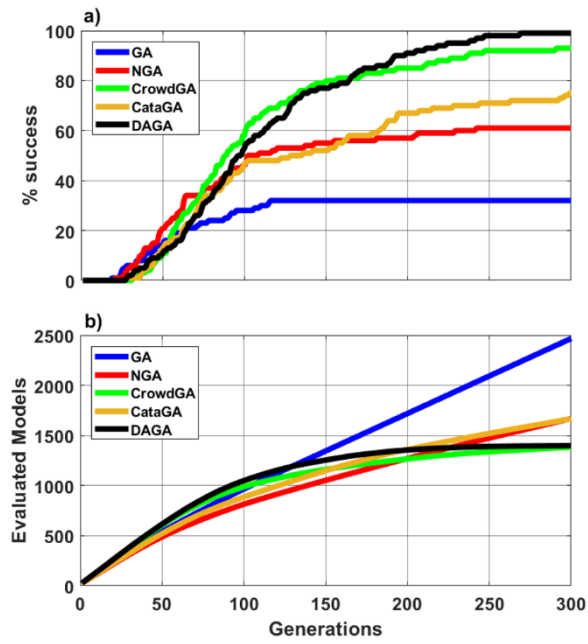
Figure 8: Test on the 2D Schwefel function. a) Cumulative curves representing the percentage of successful tests over 300 generations. b) Number of evaluated models by each considered GA implementation.
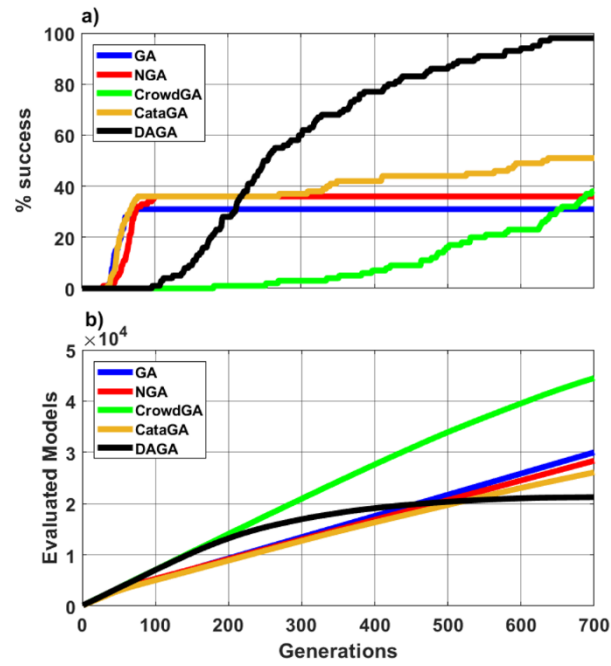
Figure 9: Test on the 2D Eggholder function. a) Cumulative curves representing the percentage of successful tests over 700 generations. b) Number of evaluated models by each considered GA implementation.
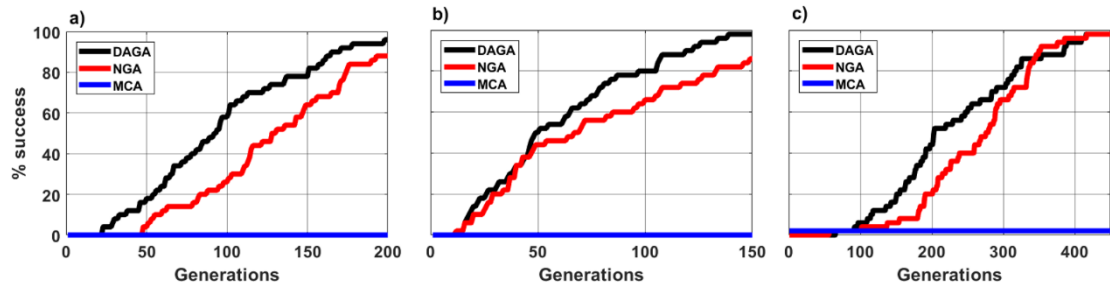
Figure 10: Cumulative curves representing the percentage of successful tests for the 3D, 4D, and 10D Rastrigin functions (a, b, and c, respectively).
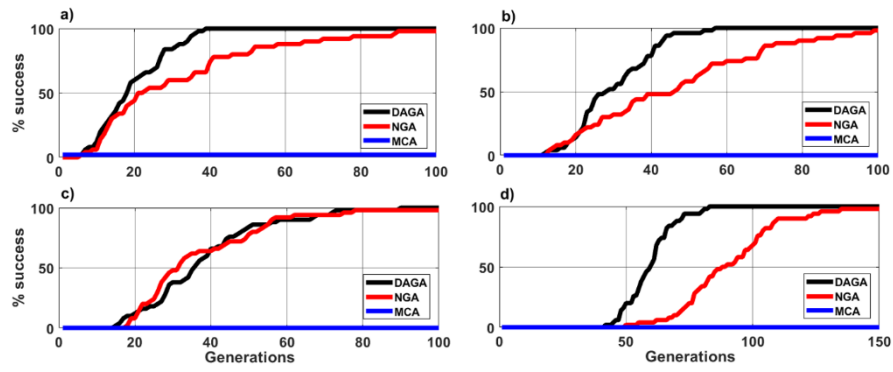
Figure 11: Cumulative curves representing the percentage of successful tests for the 2D, 3D, 4D, and 10D Ackley functions (a, b, c, and d respectively).
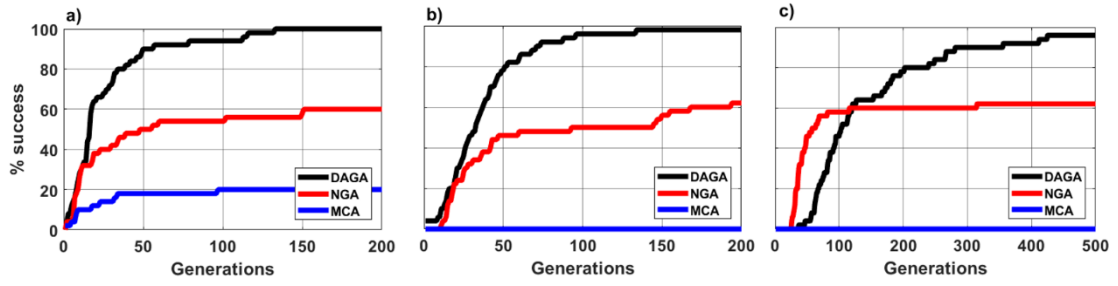
Figure 12: Cumulative curves representing the percentage of successful tests for the 2D, 3D, and

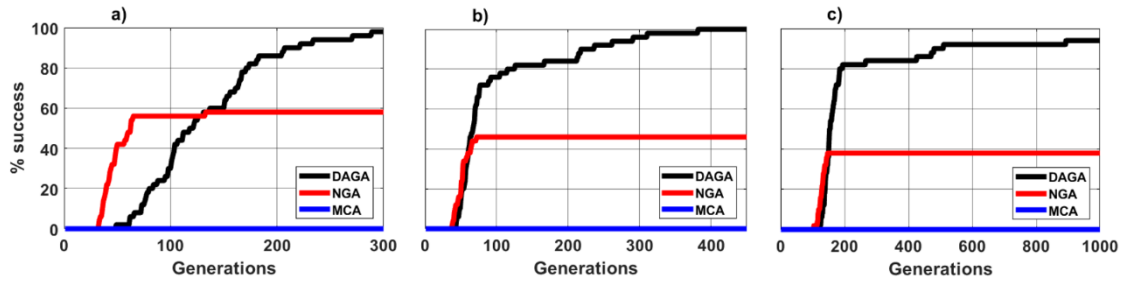*4D Langermann functions (a, b, and c, respectively).*

Figure 13: Cumulative curves representing the percentage of successful tests for the 3D, 4D, and

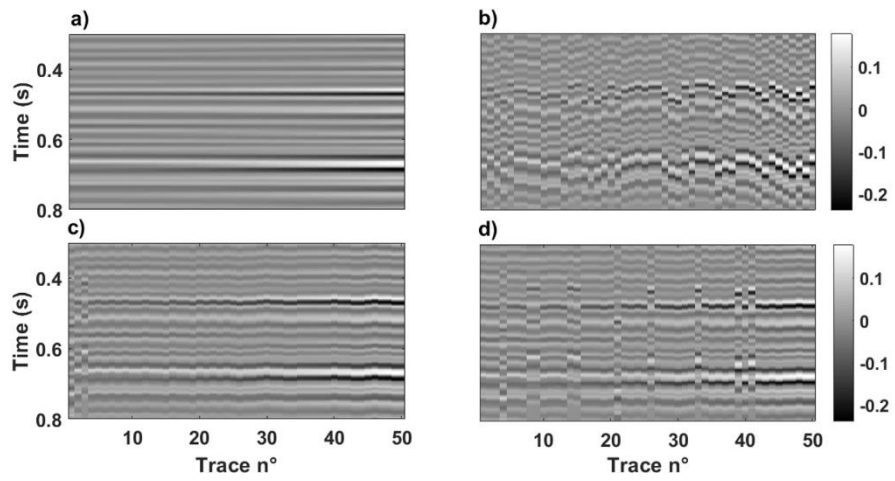*10D Schwefel functions (a, b, and c, respectively).*

Figure 14: Results for residual statics corrections with a 50-trace CMP. a) Reference CMP. b)

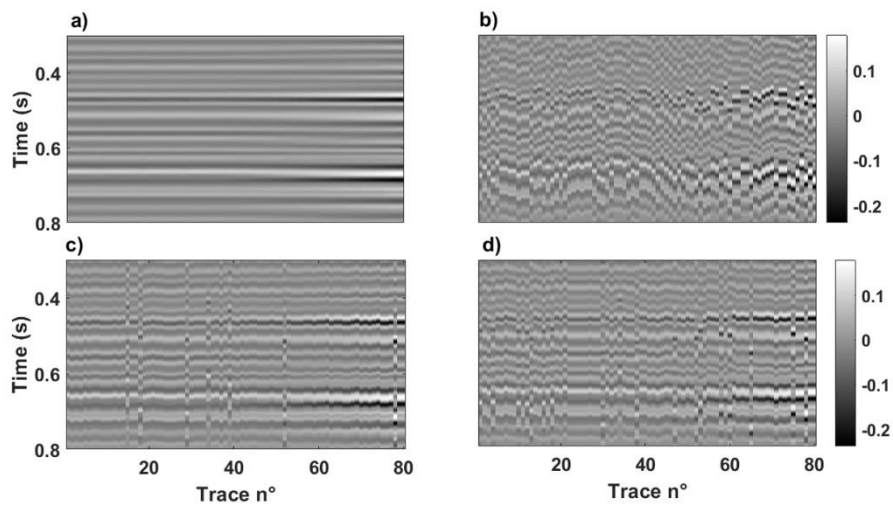*Time-shifted CMP. c) DAGA result. d) NGA result.*

Figure 15: Results for residual statics corrections with a 80-trace CMP. a) Reference CMP. b) Time-shifted CMP. c) DAGA result. d) NGA result.
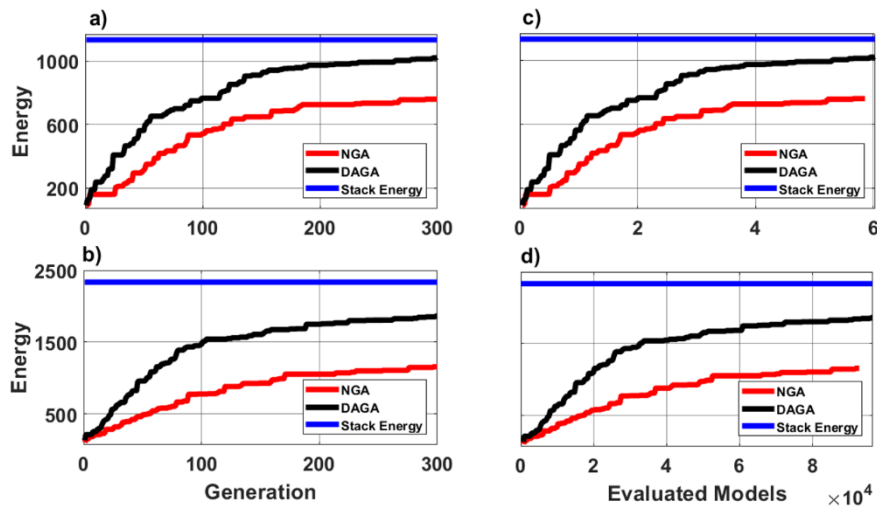
Figure 16: Evolution of the energy of the stack trace associated to the current best model for the DAGA and NGA optimizations (black and red curves, respectively). The blue curves show the energy of the stack trace associated to the reference CMP. a), b) represent the evolution for the 50-, and 80-trace examples, respectively, versus the number of iterations. c), and d) show the evolution for the 50-, and 80-trace examples, respectively, versus the number of forward modellings.
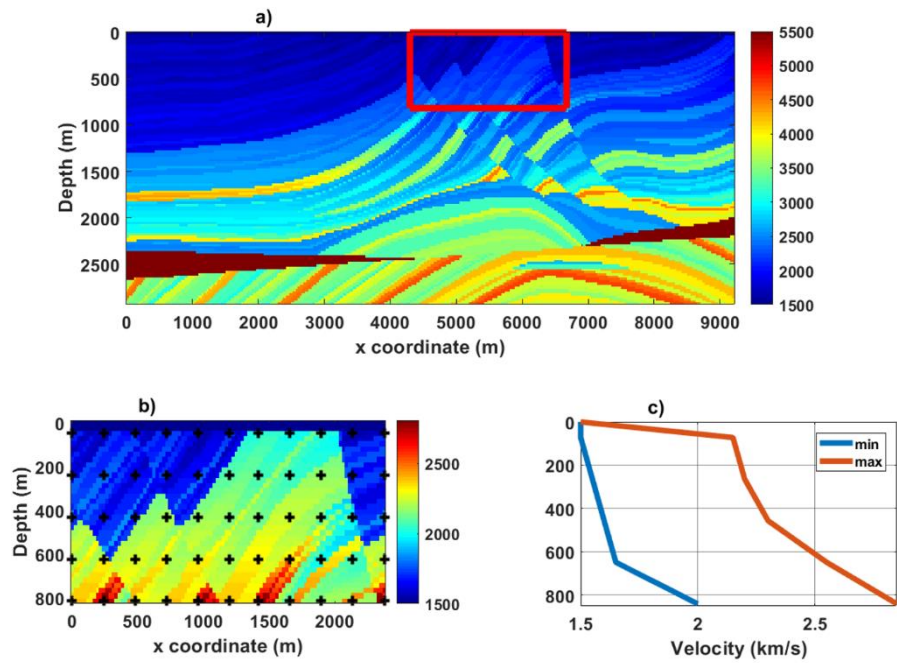
Figure 17: a) Entire acoustic Marmousi model. The red rectangle encloses the portion considered in the FWI tests. b) Close-up representing the model considered in the FWI optimizations. The black crosses represent the nodes of the inversion grid. c) The admissible ranges for the Vp values considered in the FWI tests.

Figure 18: Comparison of a shot gathers for the 7-Hz test. a) Is an example of observed shot gather. b) and c) represent the corresponding shot gather generated on the best NGA model, and the difference between a) and b), respectively. d), and e) represent the corresponding shot gather generated on the best DAGA model, and the difference between a) and d), respectively.

Figure 19: Results for the 7-Hz tests. a), and b), Final models predicted by NGA and DAGA, respectively. c) and d) Evolution of data misfit and model misfit for the DAGA and NGA optimizations, respectively. e), and f) phase-residual panels associated with the best NGA and DAGA models respectively. The colormaps code the phase residual in radians.
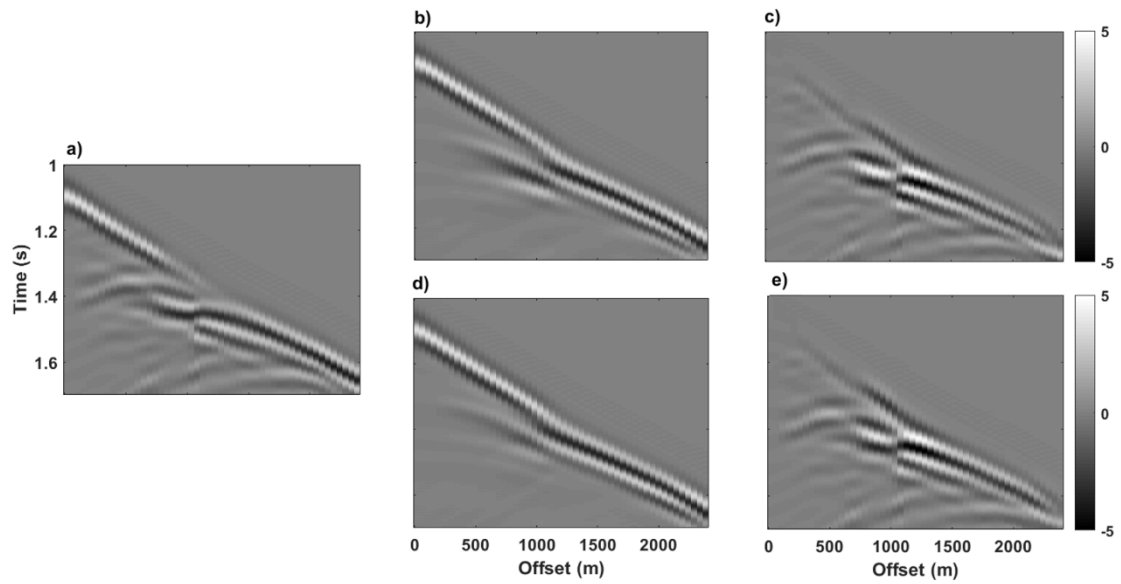
Figure 20: Comparison of a shot gathers for the 15-Hz test. a) Is an example of observed shot gather. b) and c) represent the corresponding shot gather generated on the best NGA model, and the difference between a) and b), respectively. d), and e) represent the corresponding shot gather generated on the best DAGA model, and the difference between a) and d), respectively.
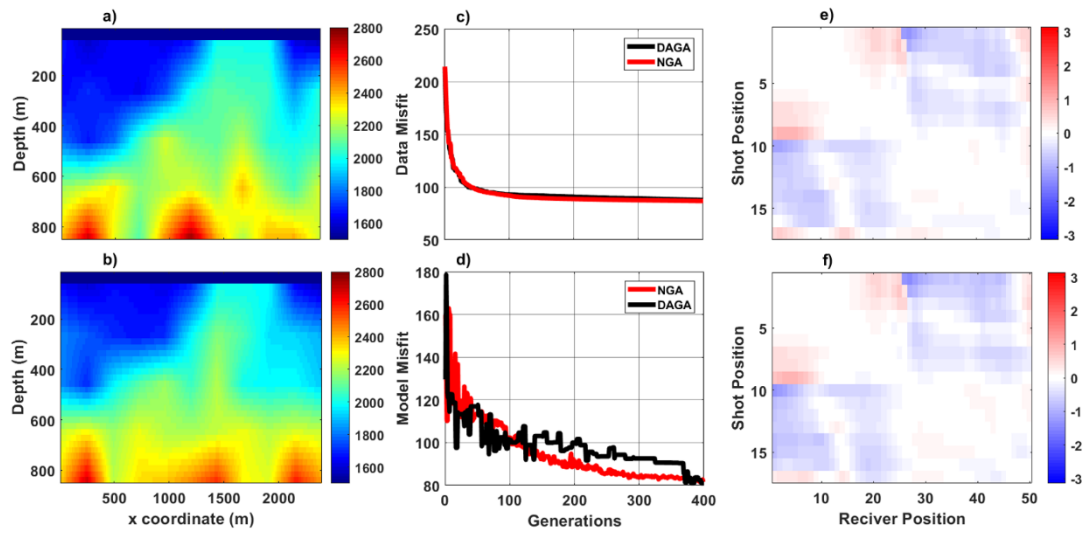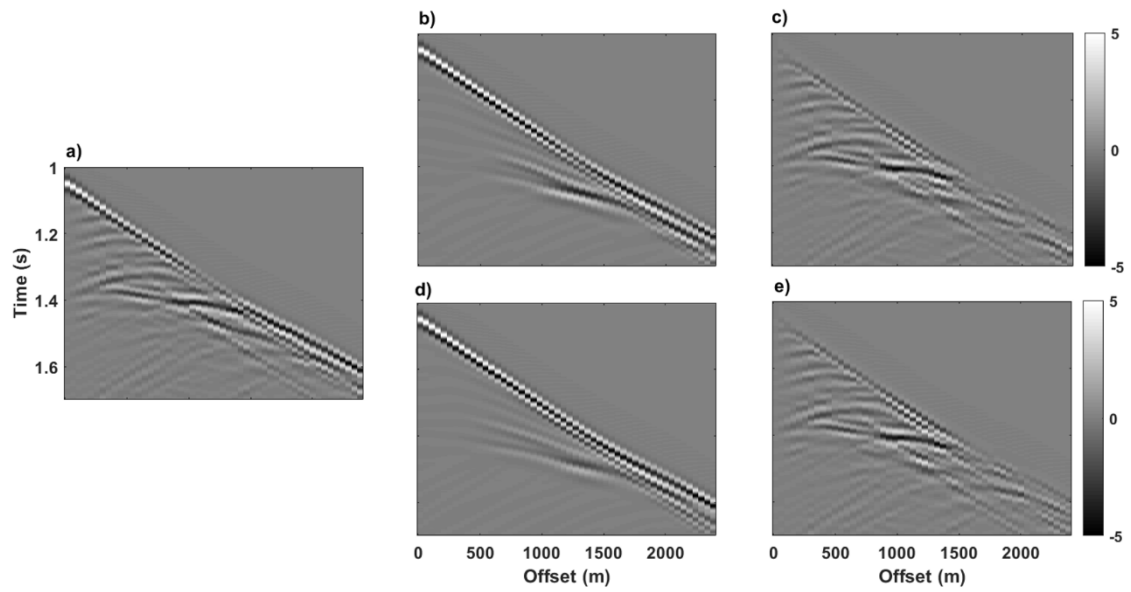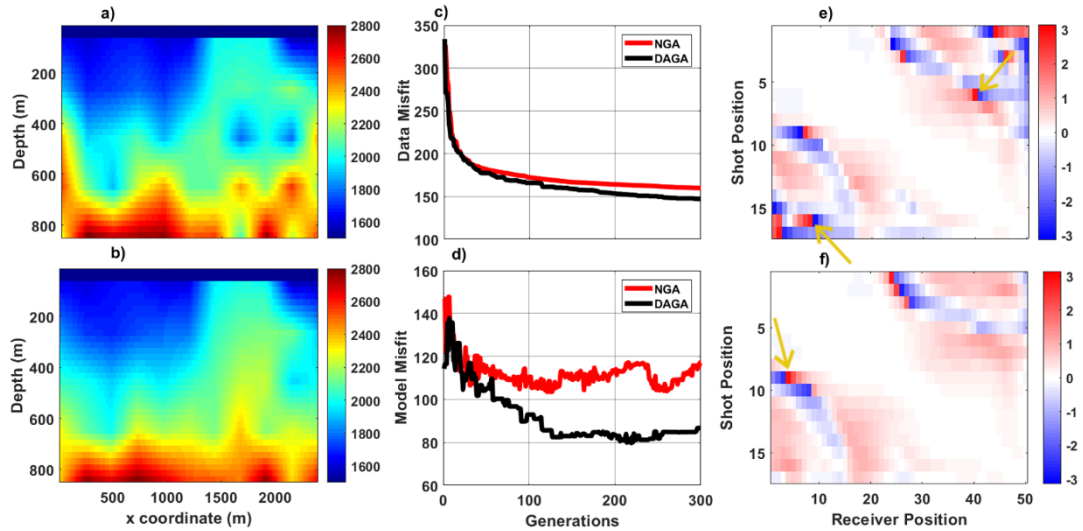
Figure 21: Results for the 15-Hz tests. a), and b), Final models predicted by NGA and DAGA, respectively. c) and d) Evolution of data misfit and model misfit for the DAGA and NGA optimizations, respectively. e), and f) phase-residual panels associated with the best NGA and DAGA models respectively. The colormaps code the phase residual in radians. The yellow arrows highlight some examples of cycle-skipped traces.
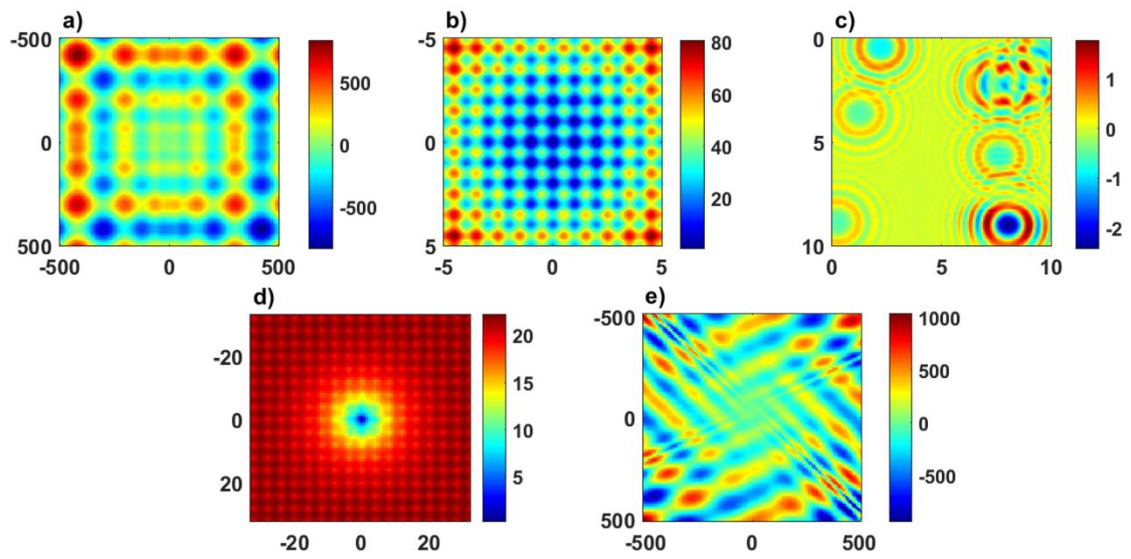
Figure A1: a) 2D Schwefel function. b) 2D Rastrigin function. c) 2D Langermann function. d) 2D Ackley function. e) 2D Eggholder function.

	2D Rastrigin	2D Schwefel	2D Eggholder
# chromosomes	6	7	76
# subpopulations	2	2	2
Selection rate	0.8	0.8	0.8
Maximum # of generations	200	300	700
# of replacements	1	1	5
Catastrophe probability	0.2	0.2	0.2

Table 1: Principal control parameters employed in the 2D tests.

	Rastrigin			Ackley				Langermann			Schwefel		
# of dimensions	3	4	10	2	3	4	10	2	3	4	3	4	10
# chromosomes	15	18	72	8	10	12	21	10	20	60	24	60	300
# subpopulations	3	3	3	2	2	2	3	2	2	3	2	2	4
Selection rate	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
Maximum generations	150	200	450	100	100	100	150	200	200	500	300	450	1000
# replacements	1	1	3	1	1	1	1	1	2	5	3	5	12
Catastrophe probability	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2

Table 2: Principal control parameters used in the optimization tests with different model-space

dimensions.

	50 traces	80 traces
# chromosomes	250	400
# subpopulations	2	2
Selection rate	0.8	0.8
Maximum # of generations	300	300
# of replacements	3	3
Catastrophe probability	0.3	0.3

Table 3: Principal control parameters used in the residual statics corrections.

	7-Hz test	15-Hz test
# chromosomes	240	320
# subpopulations	3	4
Selection rate	0.8	0.8
Maximum # of generations	400	300
# of replacements	14	14
Catastrophe probability	0.2	0.25

Table 4: Principal control parameters employed in the FWI tests.