Global Optimization for Full Waveform Inversion: Understanding Trade-offs and Parameter Choices

Gregory Ely, Alison Malcolm, and David Nicholls

ABSTRACT

In this paper, we use a fast Helmholtz solver with global optimization methods to estimate an initial velocity model for full-waveform inversion based on raw recorded waveforms. More specifically, we combine the field expansion method for solving the Helmholtz equation with a reduced parameterization of the velocity model allowing for extremely fast forward solves of realistic velocity models. Unlike conventional Full Waveform Inversion that uses gradient based inversion techniques, global optimization methods are less sensitive to local minima and initial starting models. However, these global optimization methods are stochastic and are not guaranteed to converge to the global minima. Because our adaptation of the field expansion method is so fast, we can study the convergence of these algorithms across parameter choice and starting model. In this paper we compare two commonly used global optimization methods, particle swarm optimization, and simulated annealing, and examine the limitations of these algorithms and their dependence on choice of parameters. We find that PSO outperforms SA and that PSO converges reliably for noisy data provided the number of parameters remains small. In addition, using this methodology we are able to estimate reasonable FWI starting models with higher frequency data.

INTRODUCTION

Most seismic techniques such as imaging, rely on a velocity model inverted from noisy data through a non-linear inverse problem. These inverted velocity models may be inaccurate and lead to incorrect interpretations of the subsurface. For example, an erroneously fast or slow section of a velocity model could cause a syncline structure to appear as an anticline and be incorrectly interpreted as a potential trap. Full Waveform Inversion (FWI) and other gradient based methods will converge to local minima (Warner et al., 2013) and will lead to inaccurate velocity models if initiated with an inaccurate starting model (Virieux and Operto, 2009). In recent years Uncertainty Quantification (UQ) methods have been applied to the FWI problem in order to mitigate the impact of local minima and provide a more probabilistic interpretation of the sub surface (Stuart et al., 2016; Zhu and Gibson, 2016; Ray et al., 2016). See Osypov et al. (2013) for a discussion of importance of risk assessment and UQ in seismic imaging. These methods allow for an informed decision about the reliability of the subsurface image and aid in risk estimates for drilling a potential well. However, these methods either require hundreds of thousands of wave equation forward solves to converge to accurate error estimates of subsurface parameters (Ely et al., 2018), or use local approximations of a Gaussian to approximate the uncertainty (Fang et al., 2018). The number of forward solves required for MCMC is likely infeasible for many seismic imaging problems and forward models. In addition, in low noise scenarios there many only be a single region of high probability and MCMC techniques may be unnecessary. Instead, global optimization methods are an attractive alternative as they require far fewer forward solves and provide an estimate of the best fitting model (maximum Likelihood) at a fraction of the number of forward solves needed for UQ and yet can still avoid being trapped by local minima.

In recent years global optimization methods such as Simulated Annealing (SA) (Datta and Sen, 2016; Galuzzi et al., 2017; Sajeva et al., 2016) and Particle Swarm Optimization (PSO) (Ely et al., 2015; Shaw and Srivastava, 2007) have become popular methods of calculating an initial velocity model. These methods can be an alternative or supplement to conventional methods of building velocity models such as tomography, NMO velocity analysis or very low frequency FWI (Woodward et al., 2008). Conventional initial velocity building methods typically rely on gradient based minimization and often require the manual picking of arrival times or picks in semblance space. By contrast, global optimization methods, although not guaranteed to converge to the global minima, are far more robust to local minima and frequently do not require the calculation of a gradient, reducing memory constraints and algorithmic complexity. However, starting models, frequency, and other parameter choice can have a significant impact on the convergence of these algorithms. Although these algorithms require significantly fewer iterations than UQ algorithms, rigorous comparison requires numerous global optimization runs and thousands or tens of thousands of forward solves. In addition, due to the stochastic nature of these algorithms, two different runs with the same starting model can converge to different finals models, requiring even more simulations to understand performance. Due to the computational cost of most forward solvers, previous work has been unable to accurately characterize these impacts and benchmark these algorithms as significant simplifications were necessary to control this computational cost. For example, Sajeva et al. (2017) use analytical solutions to test problems as well as the 1D wave equation to compare global optimization algorithms. In another attempt to make the problem computationally tractable, previous work has typically shown performance with one or only a few starting models when tens or hundreds of runs are necessary to generate a statistical model of convergence and accurately compare one algorithm to another (Rios and Sahinidis, 2013). Studies in the global optimization literature use test functions that are extremely fast to evaluate but are not representative of local minima found in FWI and seismic imaging. Other work combining global optimization methods and FWI have lacked a sufficiently fast test function that accurately accounts for the physics of seismic inversion in 2D.

In this paper, we use the reduced model parameterization of velocity models introduced in Frazer and Sen (1985); Zelt and Smith (1992) and Datta and Sen (2016) combined with two global optimization methods to estimate an initial velocity model. Due to the rapid speed of our forward model and we can quantitatively characterize the (1) limits on the number of degrees of freedom, (2) importance of initial guesses, (3) choice of tuning parameters, (4) comparison of different global optimization methods and (5) impact of source frequency. These comparisons, would be impossible with slower conventional finite difference solvers.

From the numerical studies presented in this paper we are able to draw several conclusions about the performance of global optimization methods for FWI and what may impact the likelihood of finding a good initial model. Although the FWI problem is known to be more prone to local minima at higher frequency (Warner et al., 2013), we find that source frequency does not significantly decrease the likelihood of convergence and a good starting model can be found with a relatively high frequency of 5 or 8 Hz from raw waveform data. This suggests that we could build initial background velocity models with much higher frequencies than are typically used. We also find that the accuracy of the initial starting models weakly impacted the likelihood of converging to the correct model. From our results we find that only having an extremely accurate initial model improved the likelihood of convergence. This suggests that only if the initial model is in the basin of attraction of the global minima are we guaranteed to converge to the true solution and outside of this range the global optimization methods are equally likely to find an accurate solution independent of starting model. In addition, we find that our implementation of Particle Swarm Optimization (PSO) out preforms Simulated Annealing (SA) and is able to find a more accurate initial model with the same number of forward solves.

The remainder of the paper is organized as follows. We first describe a reduced parameterization of the velocity model that is compatible with the field expansion forward solver. We then demonstrate that this first forward model and parametrization can yield comparable results to finite difference simulations if modifications are made to the forward solver. Second, we describe the two global optimization methods used in the paper: Particle Swarm Optimization (PSO) and Simulated Annealing (SA). Finally, we demonstrate our inversion on several synthetic models of varying complexity to determine the limitations on number of degrees of freedom, choice of initial model, and iterations needed. The field expansion method achieves significant computational saving through restricting the velocity model to consist of a series of non-overlapping layers. Although this parameterization severely restricts the velocity models we can simulate, we find that the model is sufficiently accurate to estimate initial models for FWI as demonstrated later in our paper.

FORWARD MODEL

Unlike gradient based minimization methods, global optimization and uncertainty quantification methods require far fewer degrees of freedom, tens versus thousands. In this section we briefly describe the reduced parametrization previously used by the authors in Ely et al. (2018) for uncertainty quantification. The reduced parameterization is compatible with the field expansion method (Malcolm and Nicholls, 2011a), allowing for extremely rapid forward modeling of the Helmholtz equation. We briefly describe the field expansion method and the modifications we make to the forward solver to mitigate several of the artifacts inherent to the field expansion method.

Reduced parameterization

The field expansion method is able to achieve rapid forward solves by heavily restricting the velocity model to consists of a number of non-overlapping piecewise constant layers as shown in Figure 1. Although the Earth has often been approximated by series of layers, this parameterization of the velocity model is likely too restrictive for most applications. In addition, parameterizing each layer individually as a variable would require many more degrees of freedom than is compatible with global optimization methods. A more realistic parameterization introduced in Frazer and Sen (1985) is to approximate the Earth as a series of curved interfaces with a velocity gradient between the many interfaces. This parameterization, illustrated in Figure 2 allows for complex velocity models with only a few degrees of freedom and has successfully been used for global optimization (Datta and Sen, 2016), tomography (Zelt and Smith, 1992), and for uncertainty quantification (Ely et al., 2018). Datta and Sen (2016) use a similar parameterization for global optimization with finite difference methods and thus are significantly limited in computation time. Through combining this parameterization with the field expansion, we can invert the velocity models in a fraction of the time. Ely et al. (2018) argues that the field expansion forward solver is roughly 30 to 40 times faster than finite difference methods depending on the model complexity.

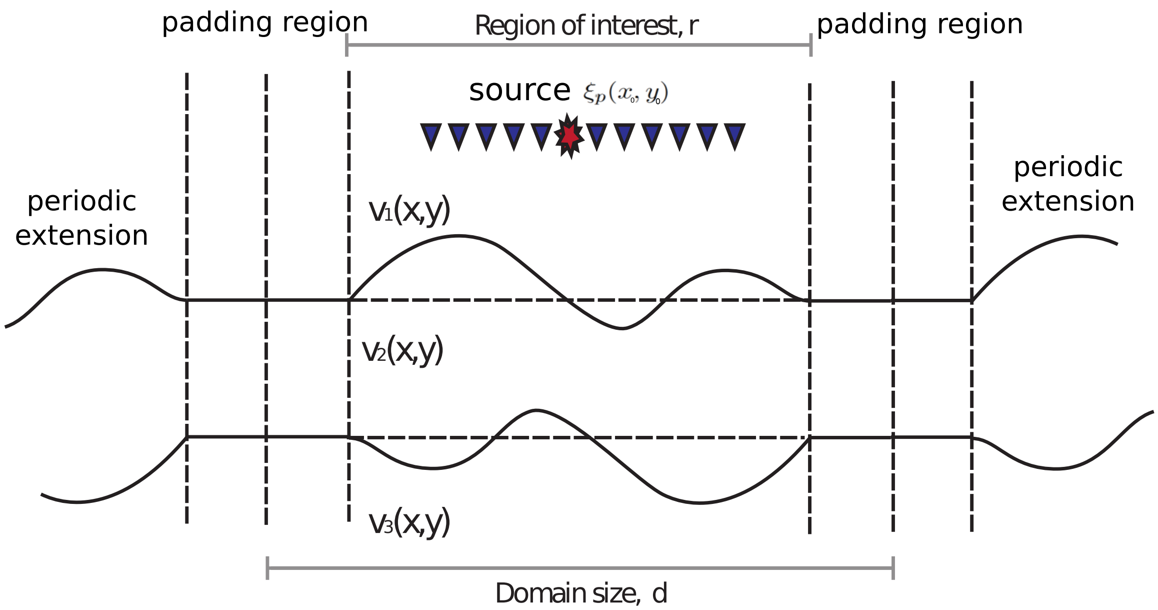

Figure 1: Diagram of field expansion velocity model and repeating boundary conditions.

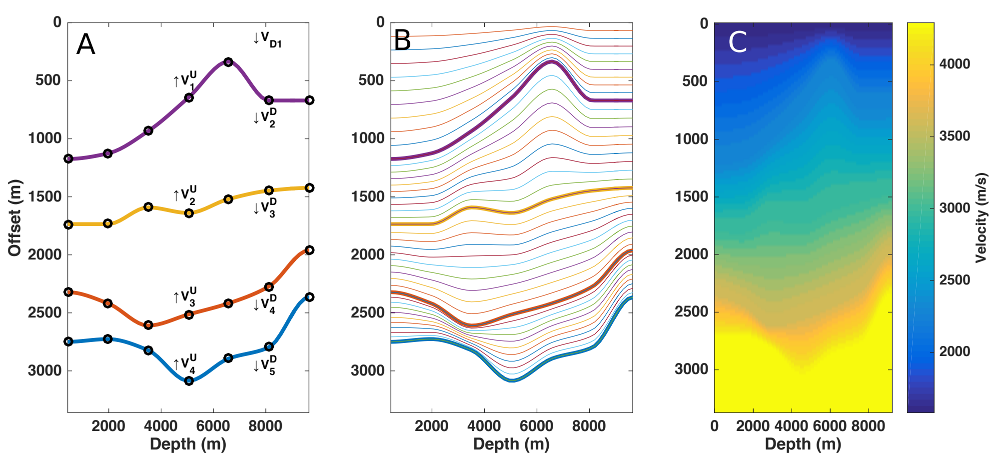

Figure 2: **A:** Reduced parameterization diagram. **B:** Interpolated layer interfaces based on the reduced parameterization. **C:** Pixelized velocity model derived from reduced parameterization consisting of piecewise constant layers.

We build a perturbed layered velocity model following from Ely et al. (2018) which is illustrated in Figure 2A. The velocity model consists of M master interfaces each with N_q control points which control the relative height of an interface across the horizontal position. The full position of the interfaces is interpolated with cubic splines between the control points. The velocity inside layer i, from interface $i-1$ to i, is a linear gradient from $\mathbf{V}_i^u$ to $\mathbf{V}_i^d$ increasing with depth. Although we cannot achieve a true gradient with the piecewise constant constraint of the field expansion, we can approximate this gradient by dividing each layer into numerous constant velocity sub-layers as shown in Figure 2B. This parameterization allows us to generate fairly complex velocity models with a limited number of parameters. For a model with 4 layers and 7 control points per interface we require only 32 degrees of freedom compared to the tens of thousands needed for a pixelized velocity model as shown in Figure 2C.

Field Expansion

In this section we briefly describe the field expansion method for solving the Helmholtz equation, see Malcolm and Nicholls (2011b) and Ely et al. (2015) for a more detailed description of the method. The field expansion method solves the Helmholtz equation for periodic boundary conditions as illustrated in Figure 1. The method achieves significant computational saving by taking the analytic solution to the scattered field for a series of flat horizontal layers and then using series expansions to calculate the scattered field for non-flat interfaces. The source and model repeat infinitely in the horizontal direction. These boundary conditions and source configurations vary significantly from the boundary conditions typically used in seismic imaging. In the following section we discuss how we modify the velocity model and field expansion method to make the results comparable to finite difference methods with perfectly matching layers frequently used in seismology.

In the field expansion method the field $v_m(x,y)$ within the m_{th} layer is subject to a periodic point source given by

$$\xi_p(x,y) = \frac{1}{2id}\frac{e^{i(\alpha_p(x-x_0)+\beta_p|y-y_0|)}}{\beta_p}, \tag{1}$$

where,

$$\alpha_p = \alpha + (2\pi/d)p, \qquad \beta_{j,p} = \begin{cases} \sqrt{k_m^2-\alpha_p^2} & \alpha_p^2 < k_m^2, \\ i\sqrt{\alpha_p^2-k_m^2} & \alpha_p^2 > k_m^2, \end{cases} \tag{2}$$

d is inter domain spacing or spatial period, p is the spatial mode number and k is the wavenumber of the source, $k = \frac{2\pi f}{v_1}$. From Equations 1 & 2 we see that the source becomes singular when β_p approaches zero which occurs at particular combinations of wavenumber and spatial mode. This singularity corresponds to a physical resonance in periodic boundary condition known as Rayleigh or Wood's anomalies, see e.g. Maystre (2012). These Wood's anomalies, generate high energy artifacts at specific frequencies that do not resemble anything in seismic imaging. These anomalies are resonances excited by the periodic source and velocity model.

To mitigate these artifacts we increases the domain spacing to be several times the size of interest. For example, if our velocity model is 3 km in offset, we set the domain size to be 15km. This padding of the domain decreases the amount of energy leaked from one repeating domain to the other. In addition, we make the velocity model slightly absorptive and dispersive by adding a complex term to the velocity. This prevents the singularity in Equations 1, and the Wood's anomalies are no longer present.

To demonstrate the effectiveness of this strategy we generate a synthetic velocity model consisting of flat layers and synthetic data at 3,5,8, and 10 Hz as shown in Figure 3. We then use the field expansion method to generate synthetic data in which we add a dispersion term (negative complex component) to the velocity model from 10^{-4} to 10^{-1} of the true velocity model and compare the field expansion simulation to data generated with PySIT's Helmholtz solver (Hewett and Demanet, 2013). Figure 4 shows the mean squared error between the finite difference and field expansion data as a function of the dispersive term. From Figure 4 we see that a large misfit is apparent if no dispersive term is added and the misfit is minimized with a dispersive term of approximately $10^{-2.5}$ for a frequency of 3 Hz. From this figure we also see that this misfit tends to increase as a function of frequency and therefore it may be more useful to use the field expansion method at lower frequencies. These discrepancies appear to be comparable or less than the differences reported between several time domain finite difference solvers (Symes et al., 2009). In addition, although the Normalized Mean Square Error (NMSE) increases as a function of frequency, we see from Figure 3 that the field expansion phase agrees well with the PySIT forward solver and only varies in amplitude. Instead of using a naive NMSE as an objective function, using an objective function that only measures difference in phase would likely provide more accurate inversions such as the zero-lag cross correlation objective function used in Liu et al. (2016).

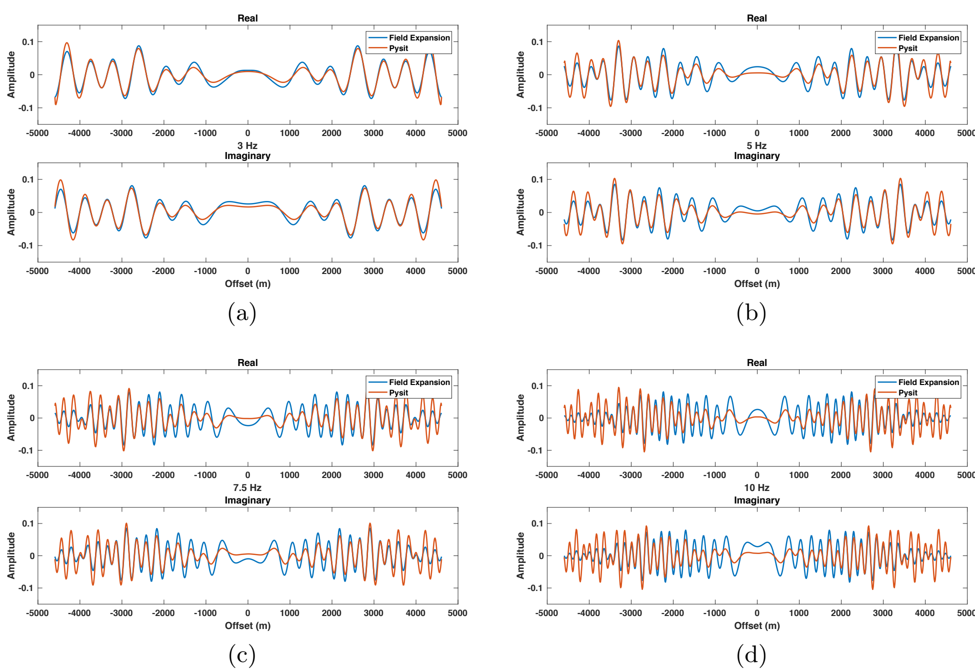

Figure 3: Real and imaginary components of the PySIT and field expansions fields at several frequencies. We generate these results with a dispersion term of 2.5×10^{-2}. **A:** 3 Hz. **A:** 5 Hz. **A:** 7.5 Hz. **A:** 10 Hz.

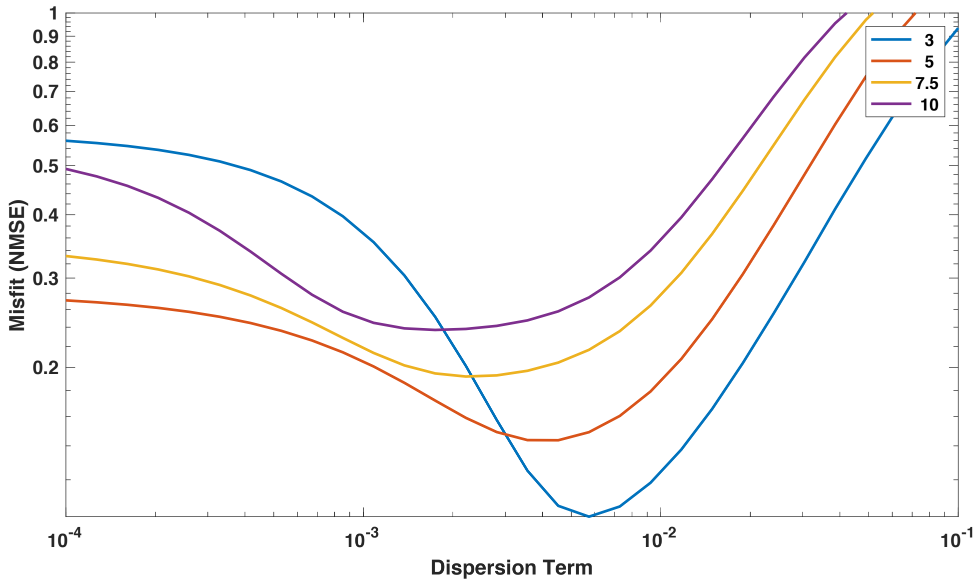

Figure 4: Misfit as a function of dispersion term for several source frequencies.

GLOBAL OPTIMIZATION METHODS

Although there are numerous varieties of global optimization methods (Sen and Stoffa, 2013), variants of Simulated Annealing (SA) and Particle Swarm Optimization (PSO) have gained popularity for seismic velocity model inversion. In their most basic form, PSO and SA solve a global optimization problem with a fixed number of continuous variables and do not attempt to optimize over degrees of freedom or solve a combinatorial problem. PSO is a parallel algorithm in contrast to SA which is necessarily a serial or sequential process, however, parallel variants of SA exist (Ram et al., 1996). In this section we briefly describe the two algorithms.

Particle Swarm Optimization

In this section we briefly describe the global particle swarm optimization method for estimating a velocity model. The PSO algorithm is based on the behavior of flocks of animals (Eberhart and Kennedy, 1995). In the algorithm, a collection of agents move through search space and communicate their results to one another to ideally guide the search towards a global minima.

In PSO, the agents are given a random starting position in parameter space with a random agent velocity at which point they move through the search space. At every iteration, each agent evaluates the forward model with parameters according to its current position and calculates a cost function between the measured and simulated field. In our case, we use the normalized mean squared error between the measured and observed field but any cost function could be used. Each agent, as it moves through parameter space, keeps track of the lowest scoring or best solution (personal best) it has visited and the best solution visited by all agents (global best) for all previous iterations. Once the score of the new position is updated, the agent's velocity is calculated based on the locations of the updated personal and global bests such that the agent is accelerated towards the local and global bests. Algorithm 1 summarizes the PSO algorithm for a swarm consisting of N_s agents with their position vector $\mathbf{x}$ and agent velocity vector $\mathbf{v}$. Note that in this algorithm the evaluation of each agent's cost function is independent from one another and the forward solves can be trivially parallelized across all agents. To apply the global optimization algorithms to the reduced forward model parameterization we must convert the field expansion velocity model to a vector of values ranging from -1 to 1. This is achieved by setting the minimum and maximum permissible velocities and layer depths to the range of -1 to 1. In the expression for the agent velocity update in Algorithm 1, γ is the inertia term controlling how much the previous agent velocity is maintained from the prior iteration and a_g and a_p are the acceleration terms that determine how much the personal best and neighborhood best alter each agent's velocity at each iteration. For all of the simulations performed in this paper the following parameters are used: $\gamma = .9$, $a_g = 1.49$, and $a_p = 1.49$. If the agent velocity exceeds .05 of the total search space then it is limited to .05 of the search space. This hard limiting of the speed of exploration prevents the search space from being explored too granularly and prevents the global optimization algorithm from missing a potential minima.

Algorithm 1 PSO

Initialization
for $i = 1$ to $\mathrm{N_s}$ **do**
  Initiate personal best score to infinity: $p_i^s = \infty$
  Initiate swarm position from a distribution across all of model space: $\mathbf{x}_i = \mathbf{U}[-1, 1]$
  Initiate swarm velocity from a uniform distribution in the range: $\mathbf{v}_i = \mathbf{U}[-vel_{max}, vel_{max}]$
end for
for $i = 1$ to maxiterations **do**
  for $i = 1$ to $\mathrm{N_s}$ **do**
    Evaluate forward solve and calculate agent's score: $\mathbf{G}(\mathbf{X}_i)$
    Record personal best score and location if better than current best: **if** $p_i^s < \mathbf{G}(\mathbf{X}_i)$ **then** $p_i^s < \mathbf{G}(\mathbf{X}_i)$ and $\mathbf{p}_i = \mathbf{x}_i$
    Determine neighborhood best location. $\mathbf{g}_i \leftarrow min(p_i^s..., p_{N_s}^s)$.
    Calculate new position: $\mathbf{x}_i = \mathbf{x}_i + \mathbf{v}_i$
    Clamp velocities: If any entries of v_i exceed vel_{max} or $-vel_{max}$ then set the entries to vel_{max} or $-vel_{max}$
  end for
end for

Simulated Annealing

Simulated Annealing is a sequential global optimization method that is based on the Metropolis Hastings (MH) algorithms used for Markov Chain Monte Carlo (MCMC) estimates for posterior distributions (Van Laarhoven and Aarts, 1987). Unlike MCMC methods, simulated annealing does not strictly provide uncertainty estimates and only provides a maximum likelihood estimate (or global minima) from a series of proposed velocity models

$\mathbf{m}_0, ..., \mathbf{m}_N$ (Tarantola, 2005). Much like the standard MH sampler, an initial model ,$\mathbf{m}_{cur}$, and proposal model $\mathbf{m}_P$ is selected from a Gaussian distribution centered about the current model and their likelihoods are calculated by,

$$L(\mathbf{m}) \equiv p(\mathbf{d} \mid \mathbf{m}) \propto \exp\left[-\frac{1}{2}\big(f(\mathbf{m}) - \mathbf{d}\big)^{\mathsf{T}} \Sigma^{-1} \big(f(\mathbf{m}) - \mathbf{d}\big)\right], \tag{3}$$

where f is the forward solver, $\mathbf{d}$, is the measured data, $\mathbf{m}$ is the proposed model, and Σ is the measured noise covariance. In the case of a noiseless observation we can set Σ to be the identity matrix. If the proposed model's likelihood, l_*, is greater than the current model likelihood, l_{i-1}, then the model is accepted. Otherwise the less likely proposed model is accepted with a probability of,

$$\frac{l_*/l_{i-1}}{t_i} \tag{4}$$

where t_i is the annealing temperature that follows an exponential cooling schedule,

$$t_i = t_{i-1} T_r. \tag{5}$$

where t_1 is the initial temperature that controls the chance of a bad solution being accepted, and T_r is the cooling that controls how quickly bad solutions are rejected as a function of iteration. From the acceptance criteria, we see that when the temperature is higher the sampler is more likely to accept proposals that poorly fit the data and potentially climb out of local minima. As the temperature decreases the sampler is less and less likely to accept a proposal that is worse than the current one. The standard SA algorithm given in Algorithm 2 is identical to the MH procedure except that the acceptance probability changes as a function of iteration, under Equation 5 and the proposal distribution. Numerous variations on cooling schedules and proposal distributions (Ingber, 1989) exist, and several have been used to estimate initial velocity models in seismic imaging (Sen and Stoffa, 2013; Datta and Sen, 2016; Xue et al., 2011). In this paper we choose simulated annealing with an exponential cooling schedule and Gaussian proposal distribution.

RESULTS

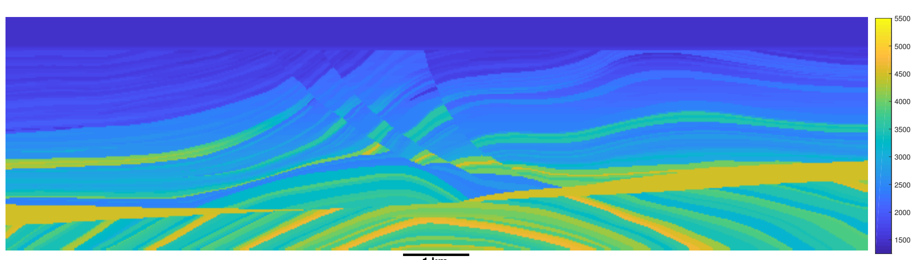

Figure 5: Marmousi constant density velocity model.

In this section we present the results of using PSO and SA to invert for an unknown

Algorithm 2 Simulated Annealing

Initialize starting model: $\mathbf{m}_0$
Initialize starting temperature: t_0
Initialize cooling ratio: T_r
for $i = 1$ to maximum iterations **do**
  $\mathbf{n} \leftarrow \text{Normal}(0, \mathbf{C}_i)$ proposed model perturbation
  $\mathbf{m}_* \leftarrow \mathbf{m}_{i-1} + \mathbf{n}$
  $l_* \leftarrow L(\mathbf{m}_*)$ proposed model score
  $\alpha_i \leftarrow \frac{l_*/l_{i-1}}{t_i}$ acceptance probability
  $coin \leftarrow \text{Bernoulli}(\alpha_i)$ "unfair" coin flip
  $t_{i+1} = t_I \times T_r$ Decrease temperature.
  if $coin = 1$ Accept/reject proposed model based on flip result **then**
    $\mathbf{m}_i \leftarrow \mathbf{m}_*$
  else
    $\mathbf{m}_i \leftarrow \mathbf{m}_{i-1}$
  end if
end for

seismic velocity model. We first test the inversion methods on a variety of simple flat layer velocity models to test the limitation on the number of degrees of freedom, reliability of inversion, importance of source frequency, and noise. Based on the values gleaned from these simplified runs, we test the limitations of these algorithms on more complex velocity models and perform an inversion on data generated with a finite difference forward solver from a gridded Marmousi velocity model (Versteeg, 1994) shown in Figure 5.

Simple models

To test the performance and limitations of the global optimization methods, we generate a set of flat layered velocity models using 2 to 10 master layers with 10 sub-layers per gradient. This velocity model is a rough approximation of a one dimensional gradient of the Marmousi velocity model shown in Figure 5. Using, a flat layer model allows us to rapidly test the dependence of our approach on numerous factors such as: total degrees of freedom, accuracy of initial models, and source frequency. Using a flat layered velocity model allows us to compute forward solves extremely quickly (the Taylor series order can be set to zero) permitting us to easily test and tune parameters.

Number of layers

To test the limitations on the number of layers, we generate a set of true velocity models consisting of flat layers roughly based on the Marmousi velocity model consisting of 2-10 master layers with 9,200 meters of offset at a frequency of 5 Hz. The dashed line in Figure 6 shows the true velocity profile for a 4 layer example.

For each number of master layers, we initiate PSO with 48 agents from a uniform distribution centered around the true velocity model. The velocity gradients, $\mathbf{v}_n^u$ & $\mathbf{v}_n^d$, in the distribution are limited to +/- 2,500 m/s around the true velocity model, and the master interfaces, $\mathbf{d}_n$, were limited to +/- 1,000 m of the true velocity model. Once initialized, we run the inversion for 250 iterations resulting in $48 \times 250 = 12,000$ evaluations of the forward model. The velocity at the surface for each agent is set to the true value of 1,500 m/s to reflect the fact that the surface velocity is generally known or can easily be determined from the direct arrival. Figure 6 shows the initial velocity models and true velocity model as well as the initial best fitting and final best fitting velocity model for a single run. From this figure, we see that although the best initial velocity model is far from the true velocity, after 250 iterations the model has converged to the true model. PSO is a stochastic algorithm and performance can vary from run to run and initial starting models. We therefore repeated this inversion procedure 25 times for each number of master layers drawing different initial models from the same uniform distribution, keeping the number of layers fixed for each set of 25 runs at each number of master layers. Figure 7 shows the median error for each of the 25 runs as a function of number of master layers, and the two red lines show the misfit of the 20 top/bottom percentile solutions (the 5th best and 5th worst fitting out of a possible 25 models according to the NMSE of the velocity model). We find that for up to 4 master layers nearly all PSO inversions converge to the true model independent of the initial model. However, beyond 4 master layers the likelihood of convergence to the true velocity significantly degrades.

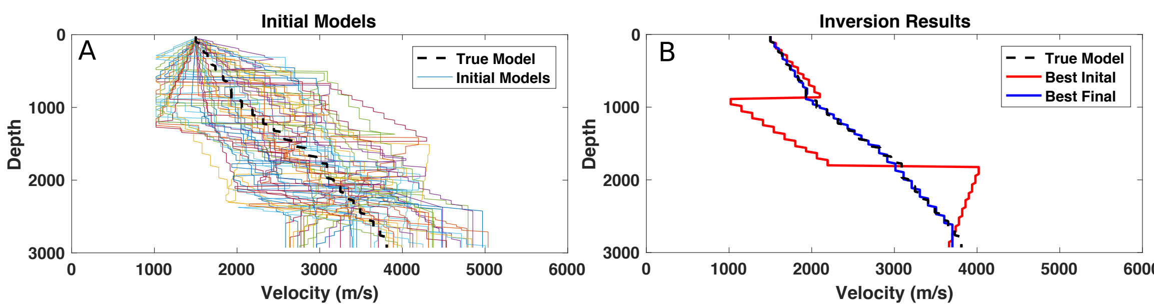

Figure 6: Single run of PSO inversion with 4 master layers. **A:** All initial models and the true velocity profile. **B:** Initial best fitting and final best velocity profile after 250 iterations.

Simulated Annealing

To fairly compare SA to PSO we took all of the 1,200 (48 agents $\times$ 25 runs) initial models for each number of master layers and ran SA for the same number of iterations (250). For the inversion we choose a cooling rate of .99, an initial temperature of 1×10^{-7} and a Gaussian step-size of $.5 \times 10^{-3}$ of the total search space of each parameter. The initial temperature was chosen based on ensuring that nearly all initial proposed models are accepted as suggested in Datta and Sen (2016). Although, a more rigorous comparison between SA and PSO involving optimization of all possible tuning parameters is beyond the scope of this paper, this comparison does give insight into the performance differences between the two types of global optimization algorithms. Figure 8 shows the median and top/bottom 20 percentile errors for SA inversions. The SA inversions appear to result in a much larger overall error compared to PSO with only a slight dependence on the number of layers. There also appears to be a slight decrease in accuracy with 3 or more master layers. However, the NMSE for

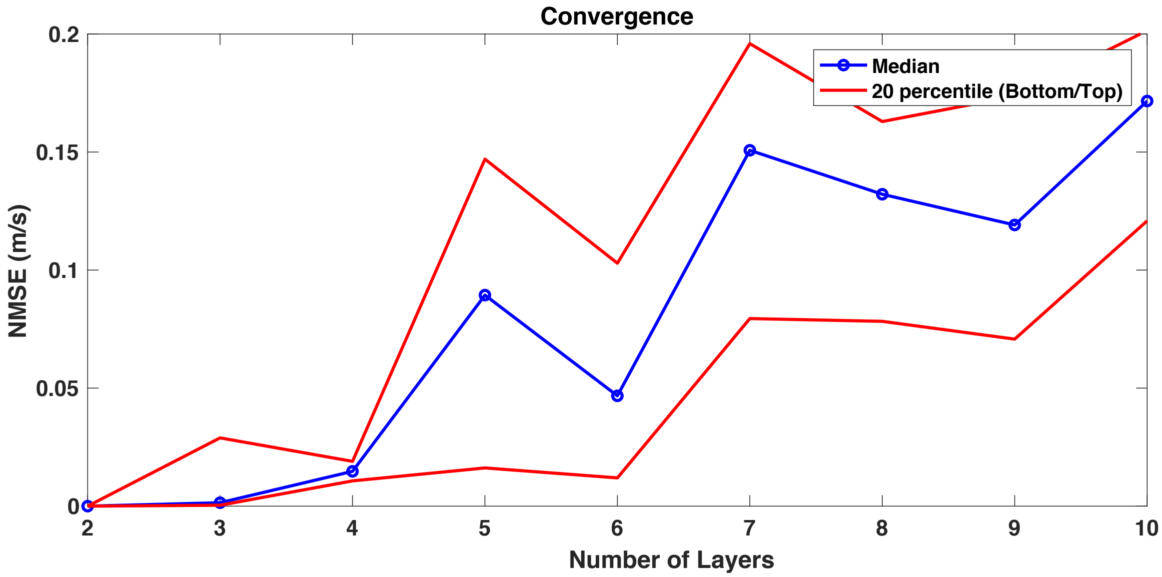

Figure 7: Particle Swarm Optimization Convergence as a function of number of layer.

all of the inversions is significantly larger than PSO and for the remainder of the results we focus on the performance of PSO.

Source Frequency

To test the impact of source frequency, we run 25 inversions of the velocity model with 2-8 master layers and source frequencies of 3, 5, and 8Hz. For this example we use a swarm size of 40 agents and run the inversion for 400 iterations to accommodate the computational constraints of our computer cluster. The initial models are drawn from a uniform random distribution centered about the true velocity model with a range of +/- 1250 m/s for the velocity gradients and +/- 500 m for the depth of the master interfaces. Figure 9 shows the convergence rates as a function of the number of master layers for the three source frequencies. Much like the convergence plot shown in the previous section, the likelihood of convergence significantly decreases for all models with more than 4 master layers. However, across frequency there appears to be very little difference in performance. This suggests that if low frequency information is unavailable due to high noise, our approach could be used to invert an initial velocity model from higher frequency data.

Number of Agents

A major factor driving the computational cost of particle swarm optimization is the size of the swarm, or number of agents. To test this limitation we run an inversion with swarm sizes of between 10 and 100 for 400 iterations with a 5 master layer velocity model at 5 Hz. As before, the initial models are drawn from a uniform random distribution centered about the true velocity model with a range of +/- 1250 m/s for the velocity gradients and +/- 500 m for the depth of the master interfaces. Figure 10 summarizes the median and top/bottom 20^{th} percentile convergence performance. From Figure 10 we see that the number of agents

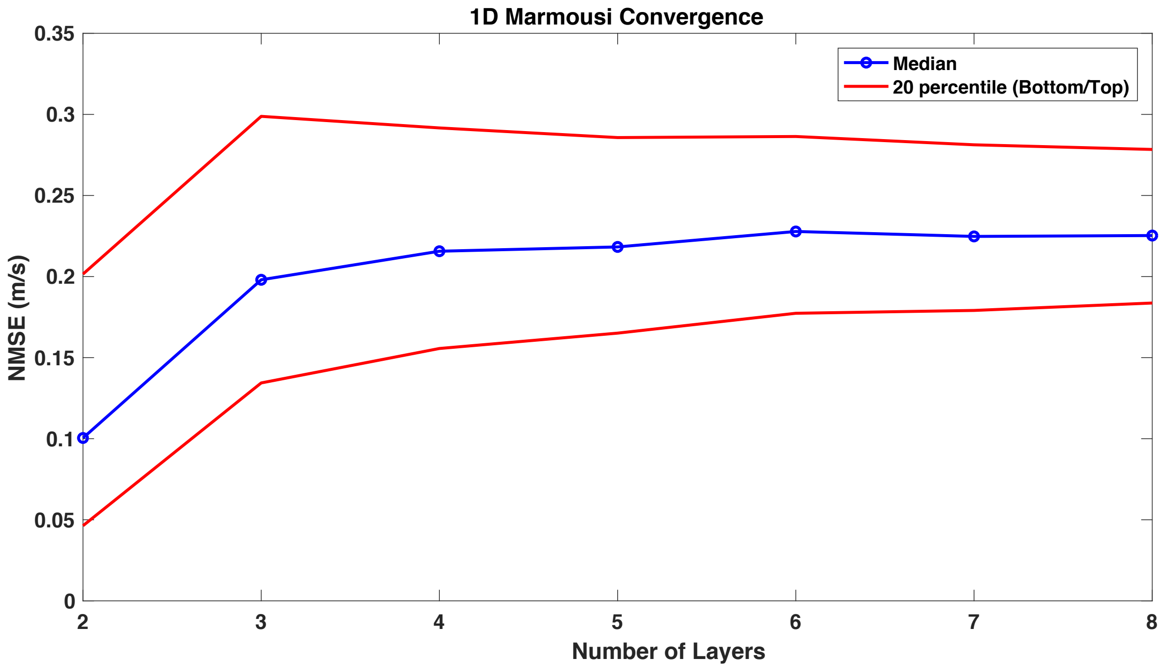

Figure 8: Simulated Annealing convergence results.

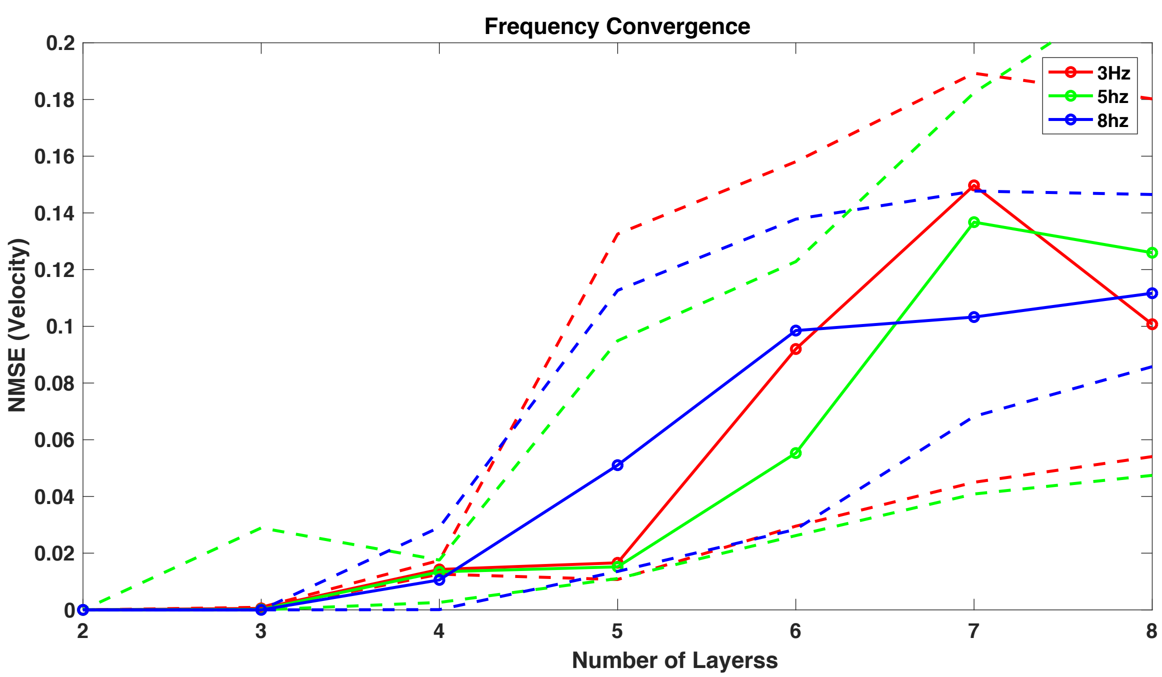

Figure 9: Convergence as a function of source frequency and number of master layers. Median values are shown with a solid line and top/bottom 20^{th} percentile solutions are shown with dashed lines.

does not have a large effect on the convergence and so we chose to use a relatively modest 20-40 agents from this point forward.

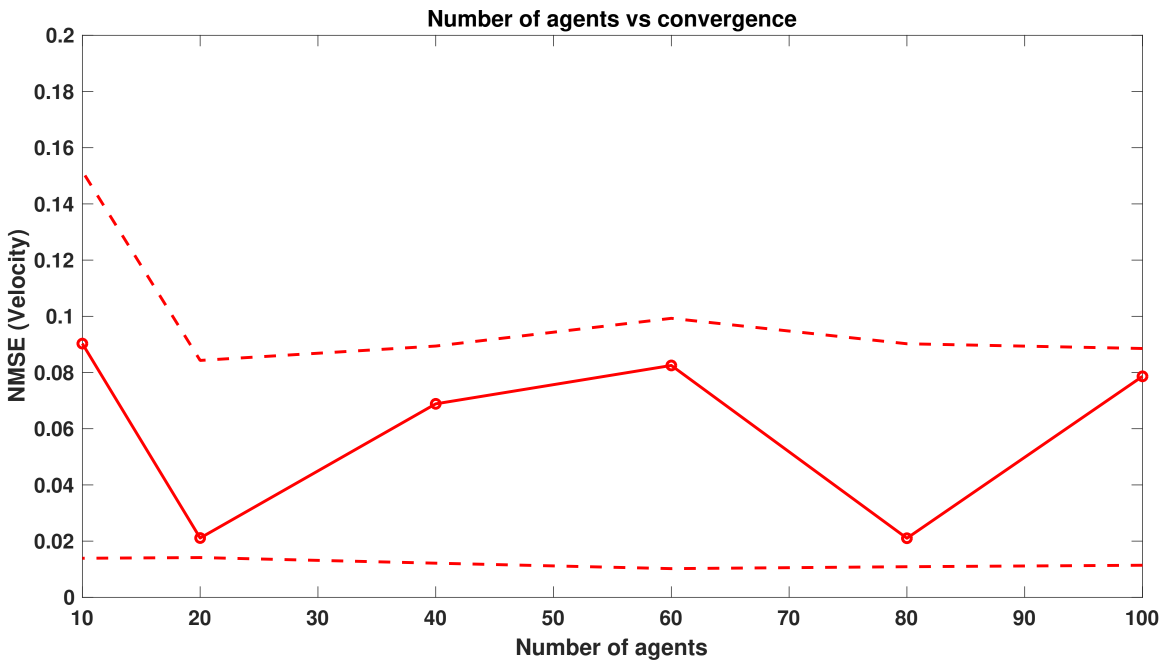

Figure 10: Convergence as a function of number of agents. Median values are shown with a solid line and top/bottom 20^{th} percentile solutions are shown with dashed lines.

Number of Iterations

For the majority of the simulations above, we run for 400 iterations. However, this number of iterations may not be optimal and the inversion may require more to convergence. To test the impact of the number of iterations on convergence, we run 25 inversions on a 5 master layer model for 50-1,000 iterations with a swarm size of 20 agents and source frequency of 5Hz. We initiate each agent position with a velocity model of +/- 1,250 m/s for the velocity gradient and +/- 500 m for master interface depth of the true model. Figure 11 shows the closeness of the forward model as a function of iteration. From this figure we see that there is a slight increase in the likelihood of convergence when more iterations are used up to a maximum of 400. However, beyond this there is little impact to the likelihood of convergence and additional iterations may be unnecessary.

Quality of Initial Guess

In the previous examples we initiated each swarm with a velocity model of +/- 1,250 m/s for the velocity gradient and +/- 500 m/s for the interface depths of the true velocity model and initiated the inversion. For all inversions the agents' master layer depths are initialized from a uniform distribution +/- 125 m centered about the true interface depth and the inversion is run for 400 iterations. For real inversions, some prior information about the geology of the region (e.g. prior velocity analysis) could influence the initial guess and its accuracy. From

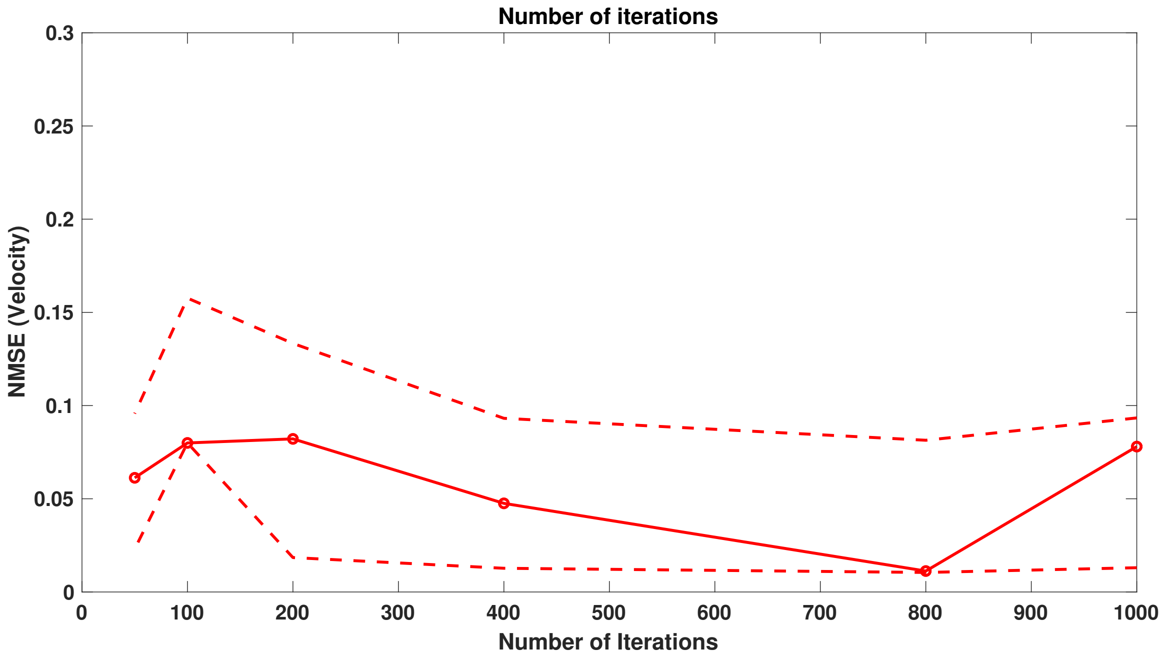

Figure 11: Convergence as a function of maximum iteration. Median values are shown with a solid line and top/bottom 20^{th} percentile solutions are shown with dashed lines.

previous studies (Ely et al., 2018), choice of initial guess significantly impacts inversion quality. To test the effect on initial guess quality we generate 25 inversions, varying the range of initial guesses from +/- 125 to +/- 1,000 m/s of the true velocity model. Figure 12 summarizes the convergence results for this experiment. From this figure we see that only when initial guess quality is very good (less than +/- 500 m/s) do the inversions have a high likelihood of converging to the true model. However, if the initial guess quality is poor (+/- 500 m/s or more) then the likelihood of convergence appears to not significantly depend on the initial guess quality. Although the median of the 25 inversions increases steadily as function of initial guess (with the exception of the outliers at +/- 750 m/s) the 20th percentile results decrease dramatically beyond +/- 500 m/s. Because the range of values is so large after +/- 500 m/s the median outlier at +/-750 m/s is likely due to the large range of misfits observed at these wider initial guesses. These results suggest that unless the initial guess is within the base of attraction to the local minima (+/- 500 m/s) the PSO inversion quality appears to independent of initial guess.

Convergence and noise

In order to test the impact of noise on convergence of the PSO algorithm, we generate a synthetic trace using a 4 master layer velocity model. We chose this model because the previous inversions in this configuration reliably converge to the true model without any noise. We then use the velocity model to generate a synthetic trace at 5 Hz and add Gaussian white noise at 10 different magnitudes such that the Signal to Noise Ratio (SNR) (where SNR is the ℓ_2 norm of the signal divided by the ℓ_2 norm of the noise) ranges from 0.11 to 53. Figure 13 illustrates the observed trace over a range of SNRs. At each of the SNRs we

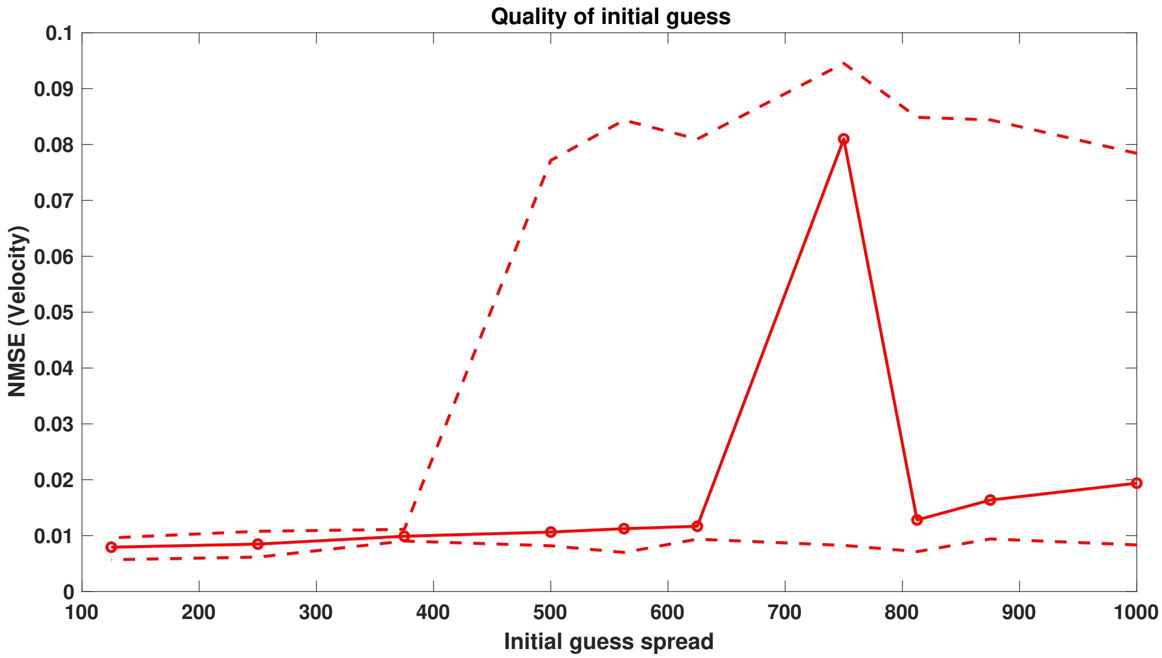

Figure 12: Model misfit as a function of initial guess quality. Median values are shown with a solid line and top/bottom 20^{th} percentile solutions are shown with dashed lines.

perform the PSO inversion initiating the starting models from a uniform distribution +/- 1250 m/s the true velocity gradient and +/- 500 m of the true master layer interface, and run our inversion scheme for 400 iterations. We repeat this procedure 25 times for each noise level, generating a new noise realization at each run. Figure 14 shows the model misfit as a function of SNR. From the figure we see that the inversion accuracy decreases as we increase the noise level, and, for SNRs below 0.5, the inversion is extremely unreliable. However, for even fairly low SNRs the inversion scheme is able to produce reliable solutions. This is likely due to the fact that our reduced parameterization highly regularizes the inversion scheme, and probably improves the robustness to noise compared with conventional FWI schemes that have many more degrees of freedom and can overfit noise.

Complex Examples

In the previous examples we used an extremely simple forward model consisting of flat layers that does not fully capture the Earth's complex subsurface. In this section we demonstrate our inversion algorithm on two complex velocity models that more accurately represent real-world geology.

Perturbed Field Expansion Model

In the first example we generate several perturbed layered velocity models, shown in the left-most panel of Figure 15, consisting of between 2-5 master layers and 7 control points per interface. These velocity models are roughly based on the Marmousi velocity model

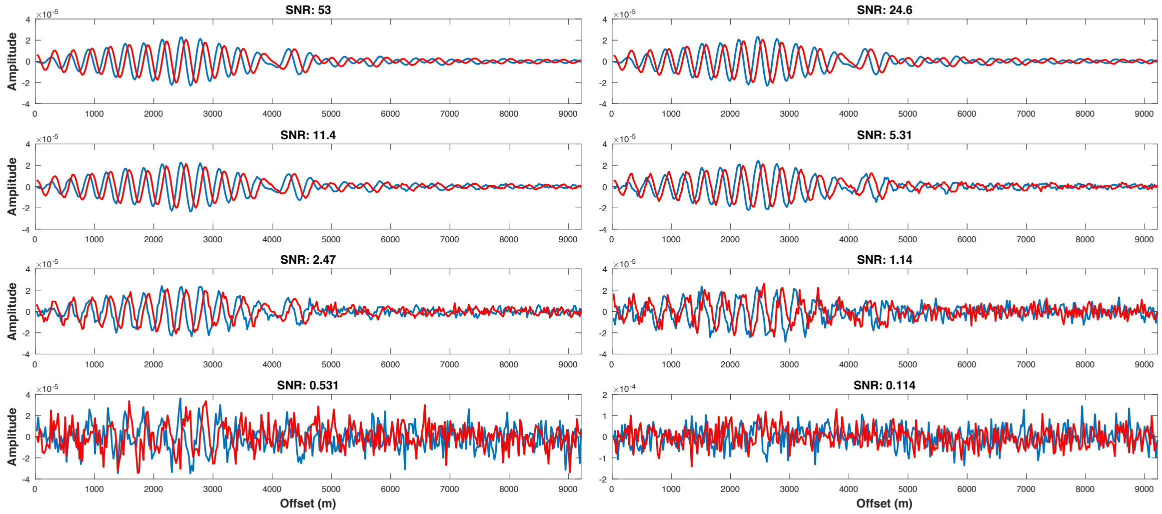

Figure 13: Illustrations of several signal to noise ratios at 5Hz. The real component of the field are shown in blue and the imaginary in red.

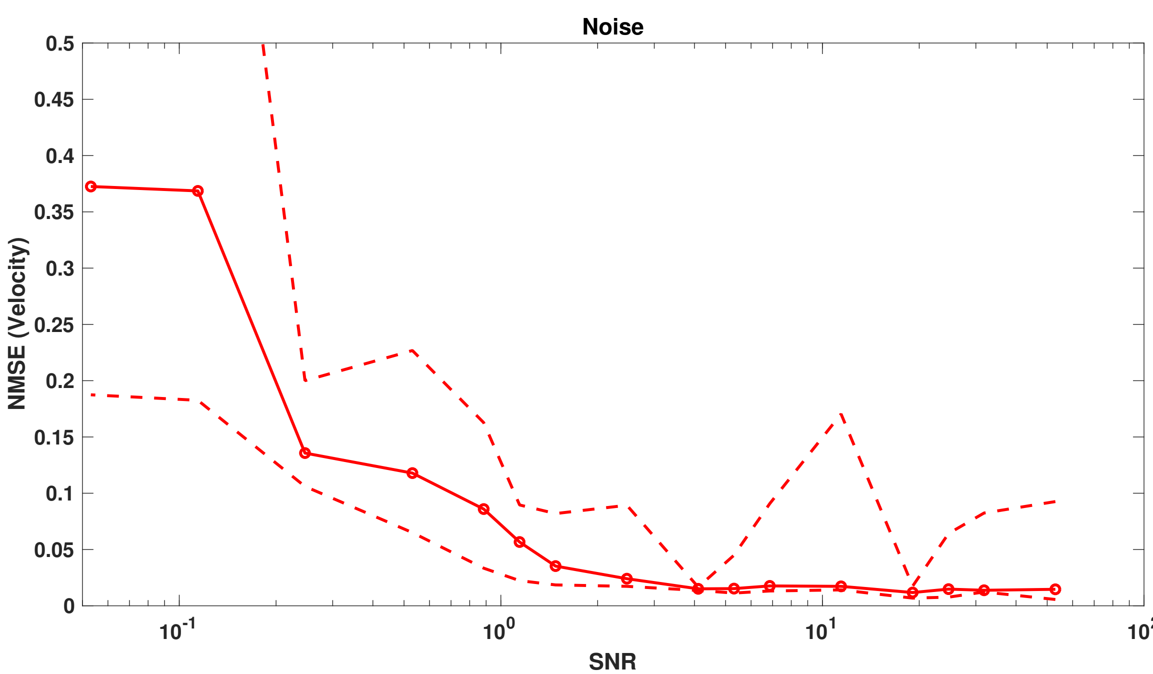

Figure 14: Model misfit as a function of Signal to Noise Ratio. Median values are shown with a solid line and top/bottom 20^{th} percentile solutions are shown with dashed line.

(Versteeg, 1994) shown in Figure 5. We then initiate a swarm of 20 agents from a uniform distribution about the true model with +/- 1250 m/s of the velocity gradient, +/ 1000 m for depths of the master layer interfaces. The locations of the control points are randomly chosen from a uniform distribution centered about 0 with +/- 100 m. We then run 25 inversions for 250 iterations at 3 Hz for each of the four velocity models. As with the flat layered velocity models used in the previous section, convergence likelihood decreases as the number of layers increases, becoming unreliable after 5 or more layers, see Figure 16. Figure 15 shows the best fitting model and the 5th best model (or 20th percentile solution) compared to the true velocity model. Although, the inversions tend to match the upper structure of the velocity model accurately, the bottom sections are more inconsistent particularly as the number of master layers increases.

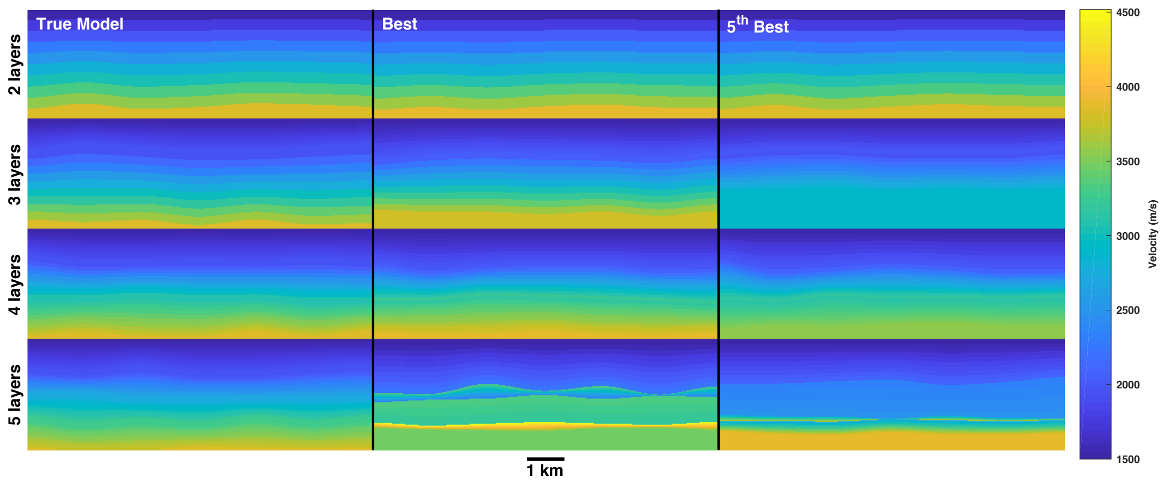

Figure 15: **Left:** True perturbed velocity models consisting of 2-5 master layers. **Middle:** the best fitting, lowest model NMSE. **Right:** The 5^{th} best fitting (20th percentile) velocity model.

Marmousi Inversion

To test the effectiveness of our inversion scheme on more realistic data, we generate 3 and 5 Hz traces using the Marmousi velocity model for a single source shown in Figure 5 with Pysit's Helmholtz solver (Hewett and Demanet, 2013). We then take the 1D vertical approximation of the pixelized velocity model and generate a profile similar to the profile shown in Figure 6. We then initiate the swarm with a velocity gradient +/- 1,250 m/s about the true profile and set the master layers to be equally spaced as a function of depth. The perturbation on each layer are initiated randomly from a uniform distribution from $0 \pm 50\ m$ and we run the PSO algorithm for 250 iterations. As with the prior experiments, we repeat this procedure 25 times for each number of master layers from 2 to 5. Figure 17 summarizes the data and model misfit between the Pysit simulated data and the inverted data using the field expansion method. Although the final measurement error is significantly higher for the 5Hz data, the error between the inverted models is comparable for both 3 and 5 Hz, suggesting that both inversion frequencies could be used for finding an initial starting model. Unlike, the results in the previous section, the convergence does not monotonically increase as with the field expansion generated truth data. The difference between the data

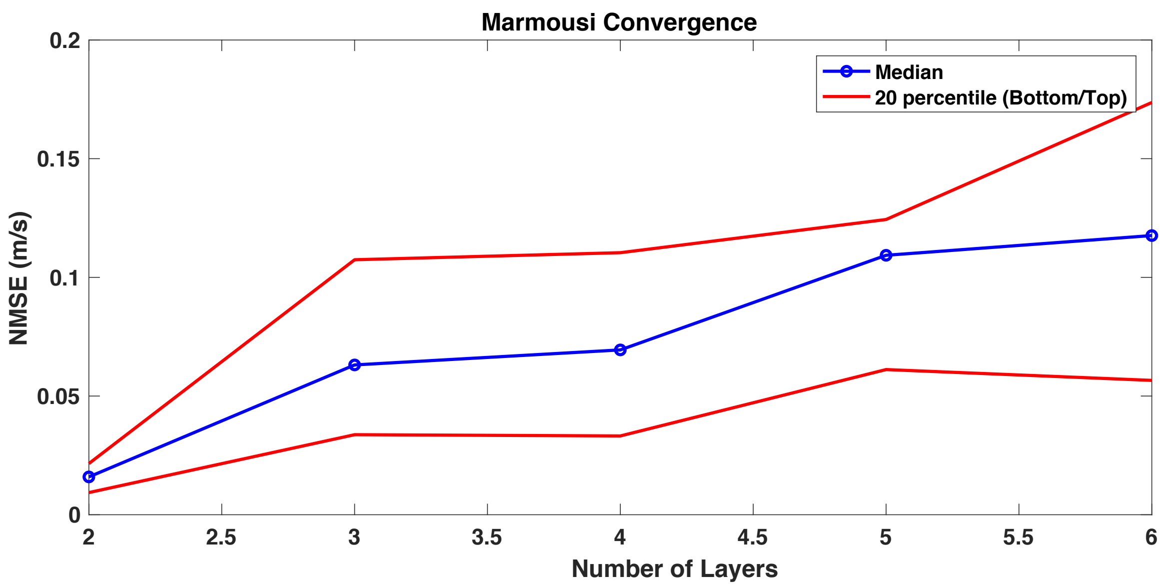

Figure 16: Convergence results for Marmousi like inversion as function of number of master layers.

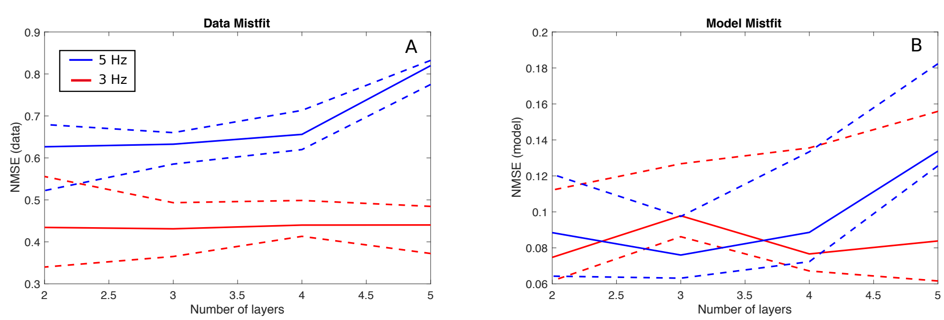

Figure 17: **A:** Data misfit between PSO inversions and the field generated from the Marmousi model at 3Hz and 5Hz as a function of number of master layers. **B:** Model misfit for the PSO inversion between the smoothed Marmousi model shown in Figure 18A left and the smoothed inverted velocity models. For both plots the median values are shown with solid lines and the top and bottom 20th percentile solutions are shown with dashed lines.

is likely due to the fact that the field expansion method agrees more closely with the finite difference data at lower frequencies as shown in Figure 4.

To further test the effectiveness of the starting models we take the 3 inversions that have the lowest data misfit for the 3Hz and 5Hz inversions and use these models as initial starting models for conventional FWI. We apply a Gaussian blur to each of the field expansion velocity models and then insert the known water layer into the initial model. Without the application of the Gaussian blur, the FWI inversions fail due to a number of high frequency artifacts. Figure 18BCD left show the 3 initial starting models at 3Hz and 19BCD left show the 3 initial models at 5Hz. For comparison we also ran FWI using the true Gaussian blurred velocity model shown in Figures 18A & 19A and the resulting inversion is shown on the right. We run all the FWIs with 64 shots for 10 iterations at each of the frequencies 3Hz, 4Hz, 5Hz, 6.5Hz, 8Hz and 10Hz using Pysit's Helmholtz solver. From the final models shown in the right of Figures 18 & 19 we see that the majority of PSO inverted initial models result in good final velocity models. Although there are artifacts present in some of the inversion results, the upper structure of the velocity model is well recovered and, for many of the models, the deeper structure agrees well with the inversion from the true baseline (see Figure 18A right).

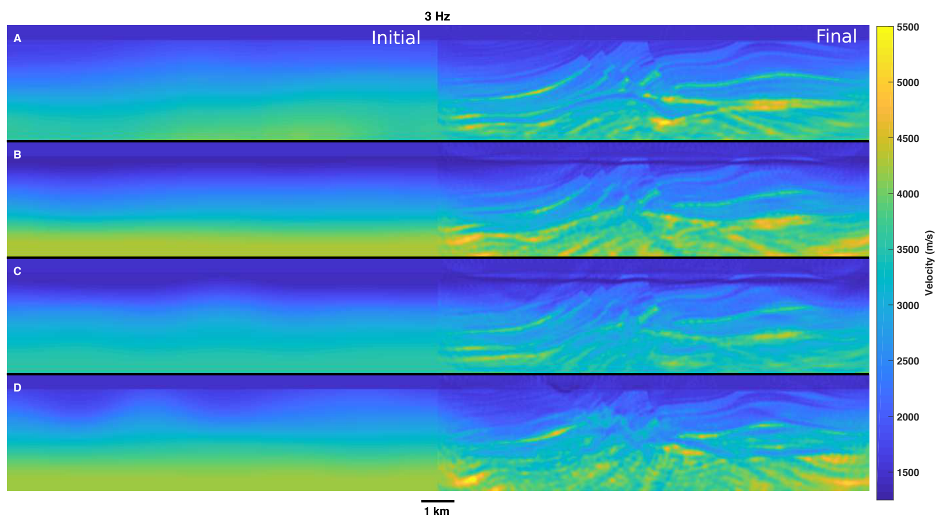

Figure 18: **Left:** Initial models used for FWI. **Right:** Final models after FWI. **A:** Inversion using the true smoothed Marmousi Velocity model. **BCD:** The initial models based on the best fitting, lowest data measurement error, PSO inversion at 3Hz and the resulting final models from FWI.

DISCUSSION

From these results we find not only that it possible to estimate a good starting model from medium frequency data, 3-8Hz, but also that the ultra-low frequencies used to estimate starting models based on raw data may be unnecessary. The difference between conven-

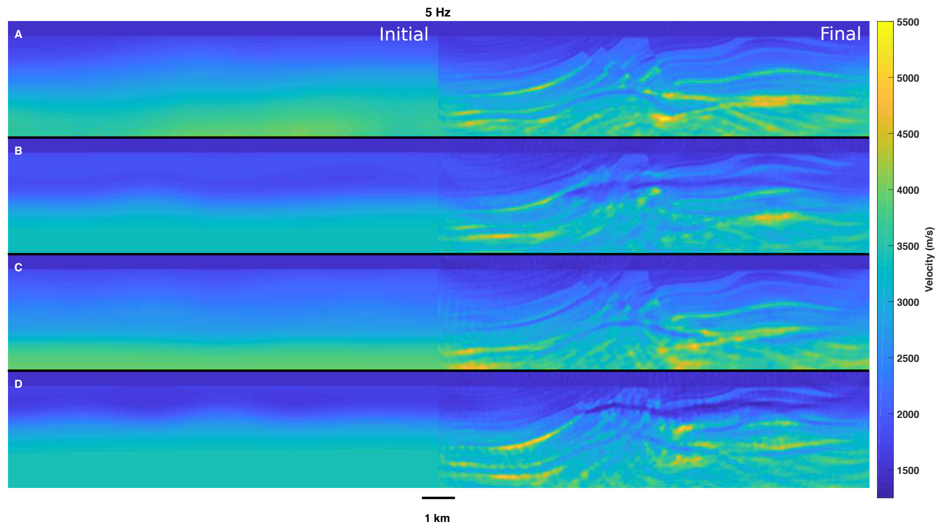

Figure 19: **Left:** Initial models used for FWI. **Right:** Final models after FWI. **A:** Inversion using the true smoothed Marmousi Velocity model. **B,C,D:** The initial models based on the best fitting, lowest data measurement error, PSO inversion at 5Hz and the resulting final models from FWI.

tional Helmholtz solvers and the field expansion method appears to increase as a function of frequency. Modifying the minimized cost function to measure phase differences could significantly decrease these discrepancies. Additional work is needed to study the performance of the field expansion based inversions at higher frequencies and with alternative cost functions. Although we are able to estimate good starting models at 5Hz, we still need the 3Hz frequency data to generate a good inversion. Running the inversion starting at 5 Hz, even for the smoothed true velocity model shown in Figure 18A, resulted in an unusable model with numerous artifacts. Other approaches, such as low frequency interpolation (Li and Demanet, 2016; Jin et al., 2018), could be used to fill in the lower (3-5 Hz) frequencies needed for FWI once a starting model is built.

From the results shown in Figures 18 & 19 we see that FWI can result in models that vary significantly for quite similar starting models. Although some of the artifacts are easily identifiable in the final image, some structures in the final image are significantly distorted or less pronounced. Little numerical work has been done on the impact of erroneous starting models and it would be fruitful to examine the performance of FWI across a population of initial models. However, it is difficult to determine what should be considered a plausible starting model. One potential approach would be use the uncertainty quantification framework presented in Ely et al. (2018) to generate a distribution of velocity models, select a few representative models, and perform FWI with these initial models. This set of initial models could then be used to benchmark various FWI algorithms against one another and give more realistic bounds of performance than using a single initial model. Although we looked only at a few global optimization algorithms and the impacts of parameters, the approach in this paper could easily be tested on a wide variety of global optimization algorithms currently being studied in the geophysics community. Using our reduced parameterization and the field expansion method should allow for objective comparisons among algorithms that would be impossible with slower conventional solvers.

CONCLUSION

In this paper we developed a platform for comparing global optimization methods for seismic velocity estimation. Although we did not exhaustively test all possible combinations of tuning parameters and cost functions, this study can guide parameter choice for other forward models where it would be too expensive to explore these trade-offs in solution space. In addition, the framework presented in this paper could easily be extended to test and benchmark a variety of global optimization methods currently used by the seismic community, such as conventional FWI. For instance, one could quantitatively compare the performance of algorithms by testing their performance over a population of reasonable starting models instead of single velocity model.

REFERENCES

Datta, D., and M. K. Sen, 2016, Estimating a starting model for full-waveform inversion using a global optimization method: Geophysics, **81**, R211–R223.

Eberhart, R. C., and J. Kennedy, 1995, A new optimizer using particle swarm theory: Proceedings of the sixth international symposium on micro machine and human science, New York, NY, 39–43.

Ely, G., A. Malcolm, and D. Nicholls, 2015, Combining Global Optimization and Boundary Integral Methods to Robustly Estimate Subsurface Velocity Models: Presented at the , Society of Exploration Geophysicists.

Ely, G., A. Malcolm, and O. V. Poliannikov, 2018, Assessing uncertainties in velocity models and images with a fast nonlinear uncertainty quantification method: Geophysics, **83**, R63–R75.

Fang, Z., C. Da Silva, R. Kuske, and F. J. Herrmann, 2018, Uncertainty quantification for inverse problems with weak pde-constraints: Geophysics, **83**, 1–107.

Frazer, L. N., and M. K. Sen, 1985, Kirchhoff-; helmholtz reflection seismograms in a laterally inhomogeneous multi-layered elastic mediumi. theory: Geophysical Journal International, **80**, 121–147.

Galuzzi, B., E. Zampieri, and E. Stucchi, 2017, Global optimization procedure to estimate a starting velocity model for local full waveform inversion: International Conference on Optimization and Decision Science, Springer, 171–179.

Hewett, R., and L. Demanet, 2013, the pysit team, 2013: PySIT: Python seismic imaging toolbox v0, **5**.

Ingber, L., 1989, Very fast simulated re-annealing.

Jin, Y., W. Hu, X. Wu, and J. Chen, 2018, Learn low wavenumber information in fwi via deep inception based convolutional networks, *in* SEG Technical Program Expanded Abstracts 2018: Society of Exploration Geophysicists, 2091–2095.

Li, Y. E., and L. Demanet, 2016, Full-waveform inversion with extrapolated low-frequency data: Geophysics, **81**, R339–R348.

Liu, Y., J. Teng, T. Xu, Y. Wang, Q. Liu, and J. Badal, 2016, Robust time-domain full waveform inversion with normalized zero-lag cross-correlation objective function: Geophysical Journal International, **209**, 106–122.

Malcolm, A., and D. P. Nicholls, 2011a, A field expansions method for scattering by periodic multilayered media: The Journal of the Acoustical Society of America, **129**, 1783–1793.

———, 2011b, A field expansions method for scattering by periodic multilayered media: The Journal of the Acoustical Society of America, **129**, 1783–1793.

Maystre, D., 2012, Theory of Woods Anomalies, *in* Plasmonics: Springer Berlin Heidelberg, Springer Series in Optical Sciences, No. 167, 39–83.

Osypov, K., Y. Yang, A. Fournier, N. Ivanova, R. Bachrach, C. E. Yarman, Y. You, D. Nichols, and M. Woodward, 2013, Model-uncertainty quantification in seismic tomography: method and applications: Geophysical Prospecting, **61**, 1114–1134.

Ram, D. J., T. Sreenivas, and K. G. Subramaniam, 1996, Parallel simulated annealing algorithms: Journal of parallel and distributed computing, **37**, 207–212.

Ray, A., A. Sekar, G. M. Hoversten, and U. Albertin, 2016, Frequency domain full waveform elastic inversion of marine seismic data from the alba field using a bayesian transdimensional algorithm: Geophysical Journal International, **205**, 915–937.

Rios, L. M., and N. V. Sahinidis, 2013, Derivative-free optimization: a review of algorithms and comparison of software implementations: Journal of Global Optimization, **56**, 1247–1293.

Sajeva, A., M. Aleardi, B. Galuzzi, E. Stucchi, E. Spadavecchia, and A. Mazzotti, 2017, Comparing the performances of four stochastic optimisation methods using analytic objective functions, 1d elastic full-waveform inversion, and residual static computation: Geophysical Prospecting, **65**, 322–346.

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 modelgenetic fwi for acoustic macro models: Geophysics, **81**, R173–R184.

Sen, M. K., and P. L. Stoffa, 2013, Global Optimization Methods in Geophysical Inversion: Cambridge University Press.

Shaw, R., and S. Srivastava, 2007, Particle swarm optimization: A new tool to invert geophysical data: Geophysics, **72**, F75–F83.

Stuart, G., W. Yang, S. Minkoff, and F. Pereira, 2016, A two-stage markov chain monte carlo method for velocity estimation and uncertainty quantification, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists, 3682–3687.

Symes, W., I. S. Terentyev, and T. Vdovina, 2009, Getting it right without knowing the answer: Quality control in a large seismic modeling project, *in* SEG Technical Program Expanded Abstracts 2009: Society of Exploration Geophysicists, 2602–2606.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: siam.

Van Laarhoven, P. J., and E. H. Aarts, 1987, Simulated annealing, *in* Simulated annealing: Theory and applications: Springer, 7–15.

Versteeg, R., 1994, The marmousi experience: Velocity model determination on a synthetic complex data set: The Leading Edge, **13**, 927–936.

Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration: Geophysics, **74**, WCC1–WCC26.

Warner, M., T. Nangoo, N. Shah, A. Umpleby, and J. Morgan, 2013, Full-waveform inversion of cycle-skipped seismic data by frequency down-shifting, *in* SEG Technical Program Expanded Abstracts 2013: Society of Exploration Geophysicists, 903–907.

Woodward, M., D. Nichols, O. Zdraveva, P. Whitfield, and T. Johns, 2008, A decade of tomography: GEOPHYSICS, **73**, VE5–VE11.

Xue, Y., M. K. Sen, and Z. Deng, 2011, A new stochastic inference method for inversion of pre-stack seismic data, *in* SEG Technical Program Expanded Abstracts 2011: Society of Exploration Geophysicists, 2757–2761.

Zelt, C., and R. Smith, 1992, Seismic traveltime inversion for 2-d crustal velocity structure: Geophysical journal international, **108**, 16–34.

Zhu, D., and R. Gibson, 2016, Seismic inversion and uncertainty analysis using a trans-dimensional markov chain monte carlo method, *in* SEG Technical Program Expanded Abstracts 2016: Society of Exploration Geophysicists, 3666–3671.