A new parallel simulated annealing algorithm for 1.5D acoustic full-waveform inversion

Xin Fu and Kris Innanen

ABSTRACT

Full-wave inversion (FWI) based on deterministic optimization (DO) methods is an appealing tool to detect the physical properties of the subsurface media, and increasing successful examples have been reported. However, the DO FWI is highly model-dependent, its success relies on a good starting model. To solve this problem, some researchers resort to the stochastic global optimization (SGO) methods that have shown the potential to alleviate the suffering of the model-dependent problem in FWI. Whereas, the SGO methods also have their own drawback that it needs to solve a great number of forward problems This is dramatically computationally expensive for the wave-equation-based FWI. In our study, we use a heuristic SGO algorithm, the very fast simulated annealing (VFSA) algorithm, to implement the constant-density acoustic FWI. To save computational time, we develop a new parallelization VFSA, in which the serial structure of VFSA is changed to some degree. Instead of updating the model parameter one by one in the same thread in a conventional serial VFSA, the parallel VFSA updates the N model parameter separately on N threads, in which the maximum efficient processor number is N. Since performing a 2D VFSA FWI directly without using any parameterization to reduce parameter number (dimension) is still prohibitive, we test the different FWIs (the DO FWI, the conventional serial VFSA FWI, and the new parallel VFSA FWI) on a 1.5D model, and for both VSP (vertical seismic profile) and surface seismic data. For each data, we use both an unbiased starting model crossing the true model and a biased starting model far away from the true model with the depth increase to investigate how the different FWIs rely on the starting model. The tests show that all FWIs for VSP seismic data are not too model-dependent, but the DO FWI for surface seismic data is mode-dependent and the VFSA FWIs can solve this problem well. Furthermore, to further save the computational cost, the data used for VFSA FWIs are multisource shot gathers. And all seismic data used are in time domain.

INTRODUCTION

Full-wave inversion (FWI) based on deterministic optimization (DO) methods (Lailly et al., 1983; Tarantola, 1984) is an appealing tool to detect the physical properties of the subsurface media, and increasing successful examples have been reported. However, the DO FWI that is a local optimization method is highly model-dependent (Virieux and Operto, 2009), its success relies too much on a good starting model. If the starting model is not good enough, FWI will easily be stuck at a local minimum that is often caused by the cycle skipping. Naturally, avoiding cycle skipping and building a good starting model for FWI have been two important areas that researchers are going deep into.

To address cycle skipping, several approches are developed. The most common use is the multiscale technique (Bunks et al., 1995; Sirgue and Pratt, 2004), in which the low- and high-frequency components sequentially participate in the inversion. Since the low-frequency component has a larger time period, it has less chance of cycle skipping.

However, the presence of low-frequency components is still the premise of successful applications (Dessa et al., 2004; Brossier et al., 2009; Fichtner et al., 2013) of the multiscale technique, and real field seismic data usually cannot meet this premise. Although Wu et al. (2014) take the envelope of seismic data to recover the low-frequency components, it is still physically mysterious. Warner and Guasch (2016) use Wiener filters that transform the predicted seismic data into the observed data to avoid the cycle skipping in the case of lacking the low-frequency component. And the analogous approaches can be found in Van Leeuwen and Mulder (2010), Routh et al. (2011), and Luo and Sava (2011).

Traping into the local minimum is not only associated with cycle skipping, the other causes including the strong nonlinearity of FWI should also take responsibility. So establishing a good starting model for FWI is still the main tune, which can not only avoid the cycle skipping but also mitigate the other causes. A good starting model should contain the low-wavenumber velocity structure. Many approaches have shown the talent for building the low-wavenumber model. Shin and Cha (2008) obtain low-wavenumber model in the Laplace domain. Ma and Hale (2013) utilize the dynamic image warping to estimate the time shifts between the observed data and synthetic data during the inversion to obtain the low-wavenumber model, then back to the conventional FWI for the high-wavenumber model. Wu et al. (2014) use the envelope fluctuation and decay of seismic data to estimate the low-wavenumber velocity structure. Zhu and Fomel (2016) propose a misfit function based on adaptive matching filtering in waveform inversion to generate a reasonable starting model for FWI. Chen et al. (2016) build the starting model by interpolating the velocities from well logs.

The researches mentioned above are based on the DO methods, recently, more and more researchers resort to the stochastic global optimization methods to alleviate the suffering of the model-dependent problem in FWI. Due to that a great number of time-consuming forward modelings are needed in the global-optimization-based FWI, currently, using only the global optimization method to obtain accurate models are mostly reported for 1D models (Afanasiev et al., 2014; Aleardi and Mazzotti, 2016), and for 2D models, many researchers use the global optimization method to produce a starting model for DO FWI. (Datta and Sen, 2016; Sajeva et al., 2016; Mazzotti et al., 2016; da Silva et al., 2019; Visser et al., 2019).

In our study, we will use a heuristic SGO algorithm, the very fast simulated annealing (VFSA) algorithm, to implement the constant-density acoustic FWI. To save computational time, we will develop a new parallelization VFSA, in which the serial structure of VFSA is changed to some degree. Instead of updating the model parameter one by one in the same thread in a conventional serial VFSA, the parallel VFSA updates the N model parameter separately on N threads, in which the maximum efficient processor number is N. Since performing a 2D VFSA FWI directly without using any parameterization to reduce parameter number (dimension) is still prohibitive, we will test the different FWIs (the DO FWI, the conventional serial VFSA FWI, and the new parallel VFSA FWI) on a 1.5D model, and for both VSP (vertical seismic profile) and surface seismic data. For each data, we will use both an unbiased starting model crossing the true model and a biased starting model far away from the true model with the depth increase to investigate how the different FWIs rely on the starting model. Furthermore, to further save the computational cost, the data

used for VFSA FWIs are multisource shot gathers. And all seismic data used are in time domain.

DETERMINISTIC OPTIMIZATION FULL-WAVEFORM INVERSION

DO FWI starts from a given model $\mathbf{m}_0$ and uses a DO method to search a model $\mathbf{m}$ that makes the synthetic data $\mathbf{d}_{syn}(\mathbf{m})$ match the observed data $\mathbf{d}_{obs}$ best. Usually, this is achieved by minimizing the L2 norm of data residual $\delta\mathbf{d}$ $(\mathbf{d}_{syn}(\mathbf{m}) - \mathbf{d}_{obs})$ given by

$$E(\mathbf{m}) = \frac{1}{2}\delta\mathbf{d}^T\delta\mathbf{d}, \tag{1}$$

where T denotes the transpose of a matrix. For constant-desity acoustic FWI, the model $\mathbf{m}$ in equation 1 represents the presure wave velocity model $v(\mathbf{x})$ in which $\mathbf{x}$ is the coodinate vector, and $\mathbf{d}_{syn}$ is the pressure field $P(\mathbf{x}, t)$ at receiver positions, which depends on $\mathbf{x}$ and time t. In this paper, $P(\mathbf{x}, t)$ is obtained by soveling the time-domain constant-desity acoustic wave equation given by

$$\frac{1}{v^2(\mathbf{x})}\frac{\partial^2 P(\mathbf{x}, t)}{\partial t^2} - \bigtriangledown^2 P(\mathbf{x}, t) = s(t)\delta(\mathbf{x} - \mathbf{x}_s), \tag{2}$$

where $\bigtriangledown^2$ is the Laplace operator, $s(t)$ is the source, and $\mathbf{x}_x$ is the source position. A finite-difference method (eighth order in space and second order in time) and a perfectly matched layer (PML) boundary condition are used to solve the wave equation in our study as a whole.

Many DO methods have been developed for FWI, including Newton-type optimizations (e.g., full Newton and Gauss-Newton methods), gradient-based optimizations(e.g., steepest-descent [SD] and non-linear conjugate-gradient [NCG] methods), Quasi-Newton optimizations (e.g., BFGS and l-BFGS methods), truncated-Newton optimations, and so on. In this work, we typically apply the SD FWI method. According to the adjoint method (Tarantola, 1984; Bunks et al., 1995; Plessix, 2006) and the preconditioning of deconvolution imaging condition which can compensate the geometric spread effect of wavefields (Margrave et al., 2011; Pan et al., 2014; Fu et al., 2019), the model perturbation can be expressed as

$$\Delta v(\mathbf{x}) = -\mu \sum_{r=1}^{ng}\sum_{i=1}^{ns} \frac{2}{v(\mathbf{x})^3} \frac{\int_0^{t_{max}} dt[\ddot{P}_f(\mathbf{x}, t; \mathbf{x}_s)P_b(\mathbf{x}, t; \mathbf{x}_r)]}{\int_0^{t_{max}} dt[P_f(\mathbf{x}, t; \mathbf{x}_s)P_f(\mathbf{x}, t; \mathbf{x}_s) + \lambda I_{max}]}, \tag{3}$$

where μ is the step length obtained by linear search; ng, ns are the number of receivers and shots, respectively; t_{max} is the maximum forward/backward propagating time t of wavefields; $\mathbf{x_r}$ is receiver positions; $P_f(\mathbf{x}, t; \mathbf{x}_s)$ is the forward wavefield due to the source at $\mathbf{x}_s$ and $\ddot{P}_f(\mathbf{x}, t; \mathbf{x}_s)$ is its the second derivative with respect to t; $P_b(\mathbf{x}, t; \mathbf{x}_r)$ is the backward/time-reversal wavefield due to the data residual $\delta\mathbf{u}$ at position $\mathbf{x}_r$; $I_{max} = \max_{\mathbf{x},t}[P_f(\mathbf{x}, t; \mathbf{x}_s)P_f(\mathbf{x}, t; \mathbf{x}_s)]$ is the square of maximum absolute value in forward propagation wavefield; λ is a damp factor.

SIMULATED ANNEALING FULL-WAVEFORM INVERSION

Very fast simulated annealing

Very fast simulated annealing (VFSA) (Ingber and Rosen, 1992), also known as the adaptive simulated annealing (ASA)(Ingber, 2000), is a popular method in the simulated annealing (SA) family (Metropolis et al., 1953; Kirkpatrick et al., 1983; Geman and Geman, 1984; Cerby, 1985; Pincus, 1970) which mimics the annealing process of a physical crystallization. During the process, the temperature of the heated material is lowered gradually, thus the system energy is reduced. Until the system reaches the equilibrium state, namely the system energy becomes stable, this annealing process is finished. To crystallize (i.e., reach the global minimum of the energy function) successfully and quickly, a suitable annealing schedule is necessary. Too fast annealing could lead to crystallization failure, while too slow annealing will increase the computational cost. Especially for the wave-equation-based FWI, the computation cost may raise to an unacceptable level.

Different from DO methods, VFSA is a heuristic stochastic global optimization method. A new candidate model $\mathbf{m}$ in the parameter i at temperature k is given by a Cauchy distribution expressed as

$$\mathbf{m}_{k+1}^{i} = \mathbf{m}_{k}^{i} + y^{i}(\mathbf{m}_{max}^{i} - \mathbf{m}_{min}^{i}), \tag{4}$$

where the superscript i is the model parameter or dimention index and the subscript k is the iteration or time index, $\mathbf{m}_{k}^{i}$ is the number i parameter of present model $\mathbf{m}_k$ constrained by $\mathbf{m}_k \in [\mathbf{m}_{min}, \mathbf{m}_{max}]$, and y^i is generated by

$$y^{i} = sgn(u^{i} - 0.5)T_k[(1 + 1/T_k)^{|2u^{i}-1|} - 1], \tag{5}$$

u^i is a random number from the uniform distribution $U[0, 1]$, and $T_k = \alpha^k T_0$ is the temperature at iteration k, in which T_0 is the starting temperature and α is a positive coefficient to decay the temperature. Typically, α is less than 1, depends on the complexity of the inverse problem, e.g., 0.9 for an easy problem and 0.99 for a tough one. y^i depends on the temperature, as the temperature decays, it has a bigger chance to generate a relatively small perturbation. Equation 5 is also the feature that distinguishes VFSA from other SA algorithms.

Next, using Metropolis criterion to accept or reject the new candidate model. It can be expressed as

$$p = \begin{cases} 1 & if\ E(\mathbf{m}_{k+1}) < E(\mathbf{m}_k) \\ \exp(-\dfrac{E(\mathbf{m}_{k+1}) - E(\mathbf{m}_k)}{T_k}) & if\ E(\mathbf{m}_{k+1}) \geq E(\mathbf{m}_k), \end{cases} \tag{6}$$

where p is acceptance probability, E is an energy function that will be discussed in the next subsection. When the new candidate model reduces the energy, it is unconditionally accepted, but when the energy increases, it is accepted according to the probability p. With the decaying of temperature, p decreases.

Energy function

In FWI, the energy function is defined as the misfit between the observed shot gather data $\mathbf{d}_{obs}$ and synthetic data $\mathbf{d}_{syn}$ obtained by solving the wave equation. In our study, we use the normalized L1 norm of data residual as the misfit, so the energy function is given by

$$E(\mathbf{m}) = \frac{\| \mathbf{d}_{obs} - \mathbf{d}_{syn}(\mathbf{m}) \|_1}{\| \mathbf{d}_{obs} \|_1}. \tag{7}$$

Normalization can enhance the adaptability of the chosen annealing schedule. Since the data residual can be impacted by many factors, for instance, the amplitude of the source, the number of shots, the length of recording time, and so on. For a non-normalized energy function, even in the case of the same annealing schedule and candidate model, when any of the above factors change, the acceptance probability could change. Whereas, the normalized energy function can avoid or mitigate such a situation. As for the L1 norm, we empirically use it. A deeper investigation may be requested outside of this paper, to compare the pros and cons of different types (e.g., L2 norm type and crosscorrelation type) of energy functions for VFSA FWI.

Multisource waveform inversion

The multisource method is an efficient way to release the computational burden in FWI. In a DO FWI, multishot gathers can cause crosstalk artifacts in the final inverted results. To solve this problem, an additional technique, such as the encoded multisource technique (Krebs et al., 2009), must be used. Fortunately, for global optimization FWI, such as VFSA FWI, we do not have the trouble of crosstalk artifacts involving multishot gathers. So we can directly and safely use the multisource shot gather data.

Parallel scheme

SA is a single point stochastic search algorithm where a new candidate model is a neighbor model of the present model, so the computational cost is directly and positively associated with the model size. For FWI with long-computation-time forward modeling and large model size, the computational cost of using SA will be very expensive, even for a 1D or 1.5D model. Parallelization is a good way to alleviate the computation burden. Parallel SA algorithms can be simply classified into two categories: single-chain parallel SA and multi-chain parallel SA (Onbaşoğlu and Özdamar, 2001). However, the efficiency of single-chain parallel SA is limited by the serial nature of SA. Multi-chain parallel SA makes the SA algorithm better adapt to parallelization by changing the serial structure. It runs different single-chain SAs in different processors and compares the results in different chains at some times, depending on the method used, to find the best result and send it back to each chain. Or different chains are completely independent, and finally, their results are compared to choose the best model.

However, multi-chain parallel SA is not specifically designed for the problem of the model size which is key to the effect of SA. And it often requests multi starting models (each chain needs a starting model), which is not suitable for FWI. In FWI, we start from

a reasonable initial, for instance, the model from the velocity analysis, rand than a true random model. A good starting model can effectively improve the performance of the SA algorithm, save computational cost. You can also start the multi chains from the same starting model, but it could increase the chance of repeat search in model space that will compromise the computational efficiency.

The performances of most present parallelism methods are highly problem-dependent, thus design a suitable parallel SA for FWI (1.5D acoustic FWI in our work) is necessary. In this paper, we design a new parallel scheme based on a single-chain VFSA, in which each parameter is updated simultaneously rather than sequentially in a conventional VFSA with no parallelization. Its pseudocode is given in Table 1. Instead of updating the model parameter one by one in the same thread in a conventional serial SA, this parallel VFSA updates the N model parameter separately on N threads. In this parallel VFSA, the maximum effective number of processors is the same as that of model parameters.

Table 1. Pseudocode of the parallel VFSA

Notation: M_{max} is the length of Markov chain, k_{max} is the maximum temperature iteration, N is the parameter number.
Input: $\mathbf{m}_0$, $\mathbf{m}_{min}$, $\mathbf{m}_{max}$, $E(\mathbf{m}_0)$, T_0, k_{max}, M_{max}, α
Output: $\mathbf{m}$, E
Initialization: $k = 0$, $j = 0$, $\mathbf{m}_k = \mathbf{m}_0$
For $k < k_{max}$
1. **For** $j < M_{max}$
   1.1. **Parallel for** $i = 1, 2, ..., N$ // only update parameter i in thread i
      1.1.1. Generate candidate model $\mathbf{m}_k^i = \mathbf{m}_0^i + y^i(\mathbf{m}_{max}^i - \mathbf{m}_{min}^i)$;
      1.1.2. Constraint $\mathbf{m}_k^i$ by $\mathbf{m}_k^i \in [\mathbf{m}_{min}^i, \mathbf{m}_{max}^i]$;
      1.1.3. Calculate $E(\mathbf{m}_k)$;
      1.1.4. Accept or reject $\mathbf{m}_k^i$ according to the Metroplois criterion;
      End
   1.2 Assemble the updated N parameters to obtain the new $\mathbf{m}_k$;
   1.3 $\mathbf{m}_0 = \mathbf{m}_k$;
   1.4 Calculate $E(\mathbf{m}_0)$;
   1.6 $j = j + 1$;
   End
2. $k = k + 1$;
3. $T_0 = \alpha T_0$;
End

1.5D ACOUSTIC MODEL EXAMPLE

To have a comparison between the deterministic FWI and VFSA FWIs (VFSA FWI with no parallelization and VFSA FWI with parallelization) and illustrate the feasibility of the new parallel VFSA, we use a 10-meter-spacing 60-by-30 1.5D acoustic model without lateral variation to test them. The true model is displayed in Figure 1a. We not only test the

surface data observed by the surface receivers in Figure 1a, but also test the VSP (vertical seismic profile) data observed by the VSP receivers in 1a. Three shots (white asterisks at the surface of Figure 1a) are excited separately for the deterministic FWI to avoid the crosstalk noise and excited simultaneously for VFSA FWIs to save computing time. The source used here is a 20Hz Ricker wavelet and the time sampling of the data is 2 ms. The observed multisource data are displayed in Figure 1b and c. For both data, each is used with two different starting models, unbiased and biased, displayed in Figure 1d. The unbiased model crosses the whole true model, and the biased model gets further away from the true model with depth increasing. These two models are used to investigate the dependence of the inverted result of FWI on the starting model. During the inversion of VFSA FWI, the velocity of each row of the model is treated as one parameter, and 60 parameters for the model as a whole, thus there is no lateral variation for inverted results of the VFSA FWIs. The annealing schedule for this model is: $T_0 = 0.001$, $k_{max} = 250$, $M_{max} = 50$, $\alpha = 0.98$, $\mathbf{m}_{min} = 0.8\mathbf{m}_0$, $\mathbf{m}_{max} = 1.2\mathbf{m}_0$, in which the constraint is a very loose one.

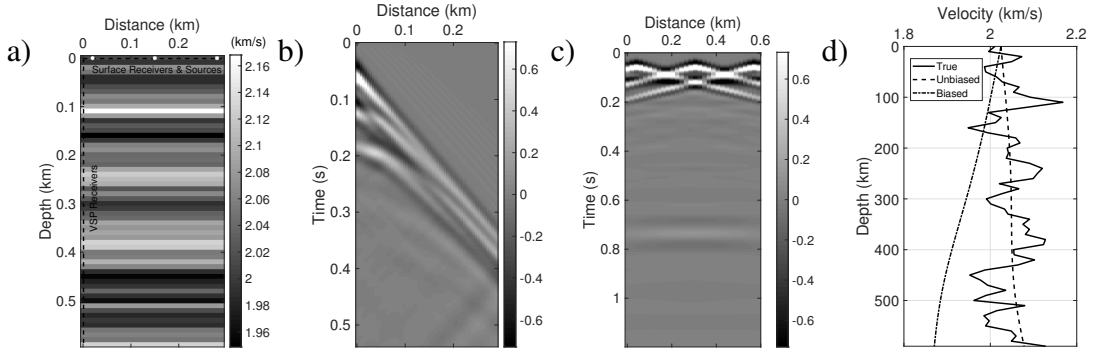

FIG. 1. (a) The true 1.5D P-wave velocity model and acquisition geometries. (b) Observed multisource VSP data. (c) Observed multisource surface data. (d) The black solid line is a trace in FIG 1a, dash and dash-dot lines are unbiased and biased starting models, respectively.

VSP seismic data test

The inverted results of VSP data inversions from the three FWIs are displayed in Figure 2a-d and Figure 3a-d. Figure 2a and 3a are the results of deterministic FWI, the quality becomes worse with the distance increasing. We extracted the traces of the best results from the Figures 2a and 3a, and displayed them in the Figures 2b and 3b respectively. The all inverted results of three FWIs using two different starting models are pretty good, and the results of the VFSA FWIs are slightly better than that of the deterministic FWI. These illustrate that VSP data inversion is not too model-dependent.

Surface seismic data test

The inverted results of surface data inversions from the three FWIs are displayed in Figure 4a-d and Figure 5a-d. Figure 4a and 5a are the results of deterministic FWI, the best results are at the center. We extracted the traces of the best results from the Figures 4a and 5a, and displayed them in the Figures 4b and 5b respectively. The results of the VFSA FWIs are better than that of the deterministic FWI. And the results of the VFSA FWI with no parallelization is slightly better than that of the VFSA FWI with parallelization. The results also show that the deterministic FWI can obtain a good result when starting

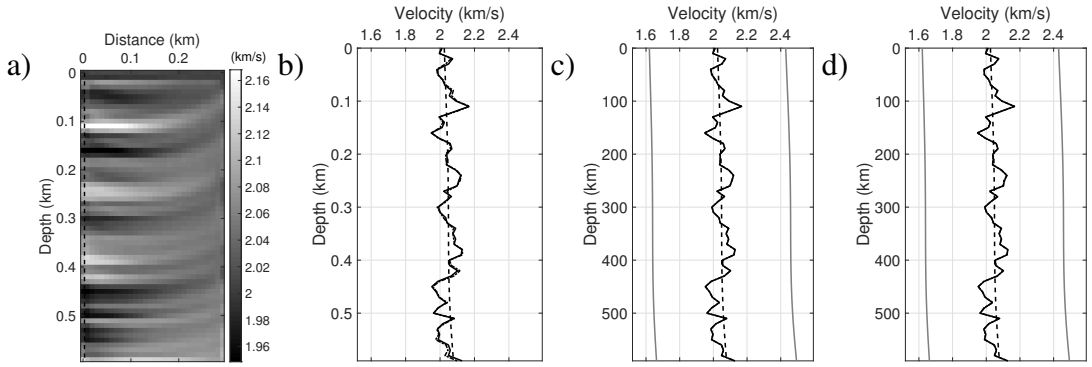

FIG. 2. VSP data inversion with the unbiased starting model. (a) The inverted result of the deterministic FWI. (b) The dash-dot line that overlaps the black solid line is the trace at the dash line in FIG 2a. (c) The dash-dot line that cannot be distinguished with the black solid line is the inverted result of VFSA FWI without parallelization. (d) The dash-dot line that cannot be distinguished with the black solid line is the inverted result of VFSA FWI with parallelization. The black solid lines are the true model, the dash lines are the starting model, and the gray solid lines are the constraints.

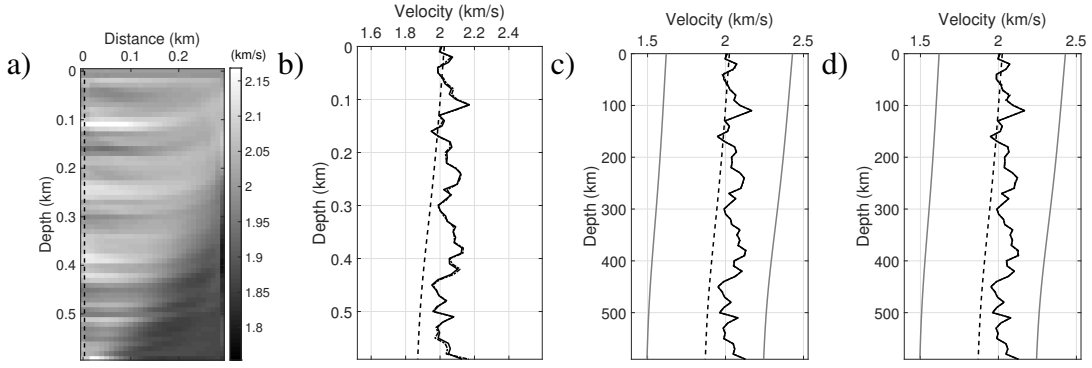

FIG. 3. VSP data inversion with the biased starting model. The legends are the same as that in FIG 2.

from an unbiased model but it cannot handle well when the starting model is biased. For a biased starting model, the deterministic FWI can recover the relative variation of the velocity model, but it cannot accurately recover the velocity values. Compared with the deterministic FWI, the VFSA FWIs do not rely much on the starting model, and good results can be obtained even if the starting model is biased.

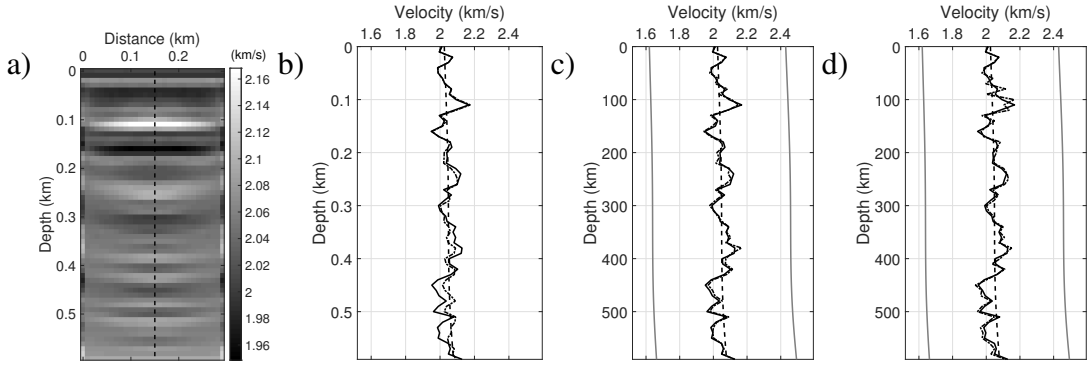

FIG. 4. Surface data inversion with the unbiased starting model. The legends are the same as that in FIG 2.

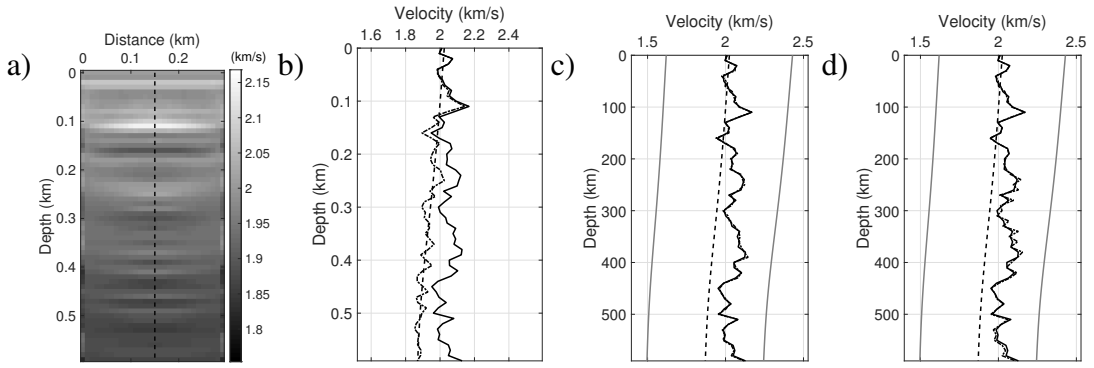

FIG. 5. Surface data inversion with the biased starting model. The legends are the same as that in FIG 2.

Parallelization performance

In the VFSA FWI with parallelization, the number of total wave equation forward modeling times is $k_{max} * M_{max} * (N + 1)$ where the $k_{max} * M_{max} * N$ times are parallelized. For the VFSA FWI with no parallelization which is serial, the number of wave equation forward modeling times is $k_{max} * M_{max} * N$. If we do not consider the communication time between processors, the computing time of VFSA FWI with no parallelization should be $(P * N)/(N + P)$ times of the computing time of VFSA FWI with parallelization using P processors. When P is much smaller than N, $(P * N)/(N + P)$ approximates to P. The maximum of $(P*N)/(N+P)$ is $P/2$ that appears when P equals to N. These calculations are based on the assumption that each forward modeling time is fixed, but practically this time will slightly float according to the state of the computer.

In our tests, all FWIs are performed on a desktop with Intel(R) Core(TM) i7-4770 CPU @ 3.4GHz 3.4GHz. We realize the parallelization by Parallel Computing Toolbox in Matlab, in which 4 workers/processors are used. Parallelization effect is displayed in Figure 6a-d. From the 6a, c and e, we can see that the VFSA FWI with no parallelization has lower energy than that with parallelization at the beginning, but they converge to a similar energy level at the end. The average computing time of VFSA FWI with no parallelization is about 3.5 times of the computing time of VFSA FWI with parallelization, which is a little lower than the 3.75 according to $(P * N)/(N + P)$. A reason for the difference is that the consuming time in the innermost loop in Table 1 depends on the longest time thread rather than the average time of all threads.

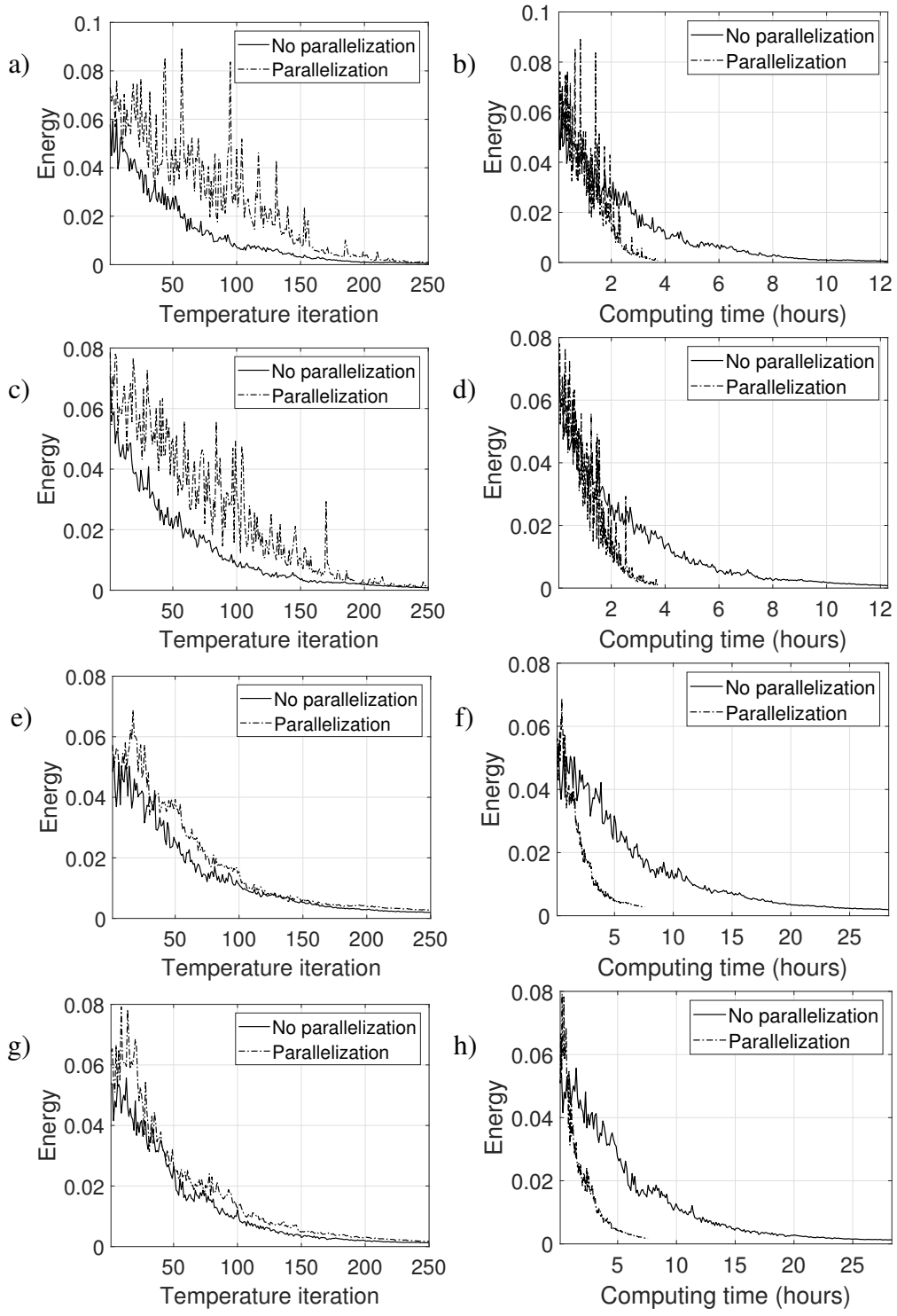

FIG. 6. The left figures display curves of energy versus temperature iteration, and the right figures display curves of energy versus computing time. (a) and (b) corresponds to the results in FIG 2c and d. (c) and (d) corresponds to the results in FIG 3c and d. (e) and (f) corresponds to the results in FIG 4c and d. (g) and (h) corresponds to the results in FIG 5c and d.

DISCUSSION

In our work we do not consider the lateral variation of the model, in the future work, we will apply the new parallelization VFSA for 2D models by combining it with a proper parameterization strategy.

CONCLUSIONS

To save computational time, we develop a new parallelization VFSA, in which the serial structure of VFSA is changed to some degree. Instead of updating the model parameter one by one in the same thread in a conventional serial VFSA, the parallelization VFSA updates the N model parameter separately on N threads, in which the maximum efficient processor number that can be used is N. Since performing a 2D VFSA FWI directly without using any parameterization to reduce parameter number is still prohibitive, we test the different FWIs (the DO FWI, the conventional serial VFSA FWI, and the new parallelization VFSA FWI) on a 1.5 model, and for both VSP (vertical seismic profile) and surface seismic data. For each data, we use both an unbiased starting model crossing the true model and a biased starting model far away from the true model with the depth increasing to investigate how the different FWIs rely on the starting model. The tests show that all FWIs for VSP seismic data are not too model-dependent, but the DO FWI for surface seismic data is mode-dependent and the VFSA FWIs can solve this problem well. Furthermore, to further save the computational cost, the data used for VFSA FWIs are multisource shot gathers.

ACKNOWLEDGEMENTS

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors, NSERC (Natural Science and Engineering Research Council of Canada) through the grants CRDPJ 461179-13 and CRDPJ 543578-19. Partial funding also came from the Canada First Research Excellence Fund.

REFERENCES

Afanasiev, M. V., Pratt, R. G., Kamei, R., and McDowell, G., 2014, Waveform-based simulated annealing of crosshole transmission data: A semi-global method for estimating seismic anisotropy: Geophysical Journal International, **199**, No. 3, 1586–1607.

Aleardi, M., and Mazzotti, A., 2016, 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.

Brossier, R., Operto, S., and Virieux, J., 2009, Robust elastic frequency-domain full-waveform inversion using the l1 norm: Geophysical Research Letters, **36**, No. 20.

Bunks, C., Saleck, F. M., Zaleski, S., and Chavent, G., 1995, Multiscale seismic waveform inversion: Geophysics, **60**, No. 5, 1457–1473.

Cerby, V., 1985, Thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm: Journal of optimization theory and applications, **45**, 41–51.

Chen, Y., Chen, H., Xiang, K., and Chen, X., 2016, Geological structure guided well log interpolation for high-fidelity full waveform inversion: Geophysical Supplements to the Monthly Notices of the Royal Astronomical Society, **207**, No. 2, 1313–1331.

da Silva, N. V., Yao, G., and Warner, M., 2019, Semiglobal viscoacoustic full-waveform inversionsemiglobal viscoacoustic inversion: Geophysics, **84**, No. 2, R271–R293.

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

Dessa, J.-X., Operto, S., Kodaira, S., Nakanishi, A., Pascal, G., Virieux, J., and Kaneda, Y., 2004, Multiscale seismic imaging of the eastern nankai trough by full waveform inversion: Geophysical Research Letters, **31**, No. 18.

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., and Villasenor, A., 2013, Multiscale full waveform inversion: Geophysical Journal International, **194**, No. 1, 534–556.

Fu, X., Romahn, S., and Innanen, K., 2019, Waveform inversion combining one-way and two-way wave-equation migration: the 31st Annual Research Report of the CREWES Project.

Geman, S., and Geman, D., 1984, Stochastic relaxation, gibbs distributions, and the bayesian restoration of images: IEEE Transactions on pattern analysis and machine intelligence, , No. 6, 721–741.

Ingber, L., 2000, Adaptive simulated annealing (asa): Lessons learned: arXiv preprint cs/0001018.

Ingber, L., and Rosen, B., 1992, Genetic algorithms and very fast simulated reannealing: A comparison: Mathematical and computer modelling, **16**, No. 11, 87–100.

Kirkpatrick, S., Gelatt, C. D., and Vecchi, M. P., 1983, Optimization by simulated annealing: science, **220**, No. 4598, 671–680.

Krebs, J. R., Anderson, J. E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A., and Lacasse, M.-D., 2009, Fast full-wavefield seismic inversion using encoded sources: Geophysics, **74**, No. 6, WCC177–WCC188.

Lailly, P., Bednar, J. et al., 1983, The seismic inverse problem as a sequence of before stack migrations: Conference on Inverse Scattering, Theory and Application, Society for Industrial and Applied Mathematics, Expanded Abstracts, 206–220.

Luo, S., and Sava, P., 2011, A deconvolution-based objective function for wave-equation inversion, *in* SEG Technical Program Expanded Abstracts 2011, Society of Exploration Geophysicists, 2788–2792.

Ma, Y., and Hale, D., 2013, Wave-equation reflection traveltime inversion with dynamic warping and full-waveform inversion: Geophysics, **78**, No. 6, R223–R233.

Margrave, G., Yedlin, M., and Innanen, K., 2011, Full waveform inversion and the inverse hessian: the 23rd Annual Research Report of the CREWES Project.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E., 1953, Equation of state calculations by fast computing machines: The journal of chemical physics, **21**, No. 6, 1087–1092.

Onbaşoğlu, E., and Özdamar, L., 2001, Parallel simulated annealing algorithms in global optimization: Journal of global optimization, **19**, No. 1, 27–50.

Pan, W., Margrave, G. F., and Innanen, K. A., 2014, Iterative modeling migration and inversion (immi): Combining full waveform inversion with standard inversion methodology, *in* SEG Technical Program Expanded Abstracts 2014, Society of Exploration Geophysicists, 938–943.

Pincus, M., 1970, Letter to the editor—a monte carlo method for the approximate solution of certain types of constrained optimization problems: Operations research, **18**, No. 6, 1225–1228.

Plessix, R.-E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, No. 2, 495–503.

Routh, P., Krebs, J., Lazaratos, S., Baumstein, A., Chikichev, I., Lee, S., Downey, N., Hinkley, D., and Anderson, J., 2011, Full-wavefield inversion of marine streamer data with the encoded simultaneous source method, *in* 73rd EAGE Conference and Exhibition incorporating SPE EUROPEC 2011, European Association of Geoscientists & Engineers, cp–238.

Sajeva, A., Aleardi, M., Stucchi, E., Bienati, N., and Mazzotti, A., 2016, Estimation of acoustic macro models using a genetic full-waveform inversion: Applications to the marmousi modelgenetic fwi for acoustic macro models: Geophysics, **81**, No. 4, R173–R184.

Shin, C., and Cha, Y. H., 2008, Waveform inversion in the laplace domain: Geophysical Journal International, **173**, No. 3, 922–931.

Sirgue, L., and Pratt, R. G., 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, No. 1, 231–248.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, No. 8, 1259–1266.

Van Leeuwen, T., and Mulder, W., 2010, A correlation-based misfit criterion for wave-equation traveltime tomography: Geophysical Journal International, **182**, No. 3, 1383–1394.

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

Visser, G., Guo, P., and Saygin, E., 2019, Bayesian transdimensional seismic full-waveform inversion with a dipping layer parameterization: Geophysics, **84**, No. 6, R845–R858.

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

Wu, R.-S., Luo, J., and Wu, B., 2014, Seismic envelope inversion and modulation signal model: Geophysics, **79**, No. 3, WA13–WA24.

Zhu, H., and Fomel, S., 2016, Building good starting models for full-waveform inversion using adaptive matching filtering misfit: Geophysics, **81**, No. 5, U61–U72.