MCMC-based time-lapse full-waveform inversion

Xin Fu and Kris Innanen

ABSTRACT

In this study, we have proposed a Bayesian time-lapse full-waveform inversion (FWI) based on the Markov chain Monte Carlo (MCMC) algorithm and a new method to estimate the data error standard deviation for the time-lapse data according to its feature. To achieve the MCMC-based time-lapse FWI, we have employed the inversion strategies including the double-difference time-lapse FWI (DDFWI), the time-domain multi-source data, the local-updating target-oriented inversion, calculating model covariance with the adaptive Metropolis algorithm, and the new data error standard deviation estimation method. The MCMC algorithm applied is a random walk Metropolis-Hastings MCMC, a typical stochastic global optimization method. In the conventional deterministic optimization (DO) DDFWI containing a baseline inversion for the baseline model and a monitoring inversion for the monitoring model, both inversions are performed by the DO FWI. In the MCMC DDFWI proposed in this work, we keep the DO FWI for the baseline inversion but employ the MCMC algorithm for the monitoring inversion. And the final time-lapse model is the difference between the inverted monitoring model and the baseline model. Synthetic data tests using a 2D acoustic model have demonstrated the feasibility of MCMC DDFWI on both time-lapse model inversion and uncertainty qualification. We also have compared the MCMC DDFWI with the conventional DO DDFWI, which shows that the inverted average time-lapse model of MCMC DDFWI can provide the results with clearer edges of the nonzero time-lapse model change and fewer coherent errors.

INTRODUCTION

Full-wave inversion (FWI) (Lailly et al., 1983; Tarantola, 1984; Virieux and Operto, 2009) based on the wave equation has been employed extensively in geophysics. Time-lapse FWI that can detect time-lapse property changes of the subsurface with a high resolution has become an important tool. Conducting time-lapse FWI normally contains twice inversions, a baseline inversion for the baseline model and a monitoring inversion for the monitoring model, and the time-lapse model is produced by subtracting the baseline model from the monitoring model. Only consider how data and starting models are used, inversion strategies of the time-lapse FWI can be classified into three basic categories, including the parallel difference FWI (using baseline data and monitoring data independently, using the same starting model for twice inversions), the sequential difference FWI (using baseline data and monitoring data independently, using the inverted baseline model for monitoring inversion) (Oldenborger et al., 2007; Routh and Anno, 2008), and the double-difference FWI (DDFWI) (Watanabe et al., 2004; Onishi et al., 2009; Denli and Huang, 2009; Zheng et al., 2011; Asnaashari et al., 2011; Routh et al., 2012; Raknes et al., 2013; Maharramov and Biondi, 2014; Raknes and Arntsen, 2014; Yang et al., 2016) adopted in this work.

In DDFWI, the first inversion is the baseline inversion that is the same as the other strategies, in which the input elements are the baseline data and a reasonable starting model. But in the second monitoring inversion, DDFWI uses a composited data as an alternative

of the monitoring data, which is the difference data (the difference between the monitoring data and the baseline data) plus the synthetic data of the inverted baseline model. Since FWI is extremely easy to be trapped into a local minimum, it means the twice inversions in both the parallel difference FWI and sequential difference FWI will yield different convergences on inverted models, it causes coherent errors on the final inverted time-lapse model (Asnaashari et al., 2014; Yang et al., 2015). Nevertheless, DDFWI is using the difference data, which helps it to focus on the target time-lapse area, thus DDFWI has fewer coherent errors in the inverted time-lapse model. Although DDFWI also is of the shortcoming of requiring good repeatability of baseline and monitoring surveys (Yang et al., 2015), with the efforts of researchers, some demerits have been improved, for instance, Hicks (2002) and Yang et al. (2016) use the interpolation technique to resample baseline and monitoring data to the same grids, Fu et al. (2020) develop a double-wavelet DDFWI method to solve the case that the wavelets of baseline data and monitoring data are different.

FWI based on deterministic optimization (DO) methods is a very common 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 the starting model. If the starting model is not good enough, FWI is easy to fail, especially when lacking low-frequency data. Recently, more and more researchers resort to stochastic global optimization methods for not only model inversion but also uncertainty qualification. Due to that a great number of time-consuming forward modelings are needed in the global-optimization-based FWI, at present using only the global optimization method to directly obtain the inverted models are mostly reported for 1D models (Hong and Sen, 2009; Afanasiev et al., 2014; Aleardi and Mazzotti, 2016; Ray et al., 2016), and for 2D models, researchers often need to combine some other strategies with the global optimization method to alleviate the computing burden. Datta and Sen (2016) reduce the parameters by sparsely parameterizing the velocity as several interfaces and use an very fast simulated annealing algorithm. Sajeva et al. (2016) propose a two-grid technique, that is a coarse grid for the subsurface model but a finer grid for the forward modeling, based on the genetic algorithm to lower the model dimension, also see in Mazzotti et al. (2016). Biswas and Sen (2017) parameterize the velocity model with Voronoi cells and represent the parameters with certain nuclei points, and utilize a Reversible Jump Hamiltonian Monte Carlo algorithm developed by themselves. Ely et al. (2018) employ a fast field expansion method to simulate the wavefield. Also, you can see da Silva et al. (2019) adopt a quantum particle-swarm optimization for a sparse Q (quality factor) macro model. And Visser et al. (2019) present each layer in a narrow model with velocity, thickness, and lower interface dip angle under the frame of Bayesian transdimensional (trans-d) Markov chain Monte Carlo (MCMC), then stitch the narrow models together as a large 2D model. Totally, almost all methods of using a global optimization for 2D model need a sparse parameterization and then input the inverted model as a starting model of the DO FWI. Although some researchers also use the global optimization method to invert the 2D models directly, it only works for some special cases. For example, Stuart et al. (2019) combine a two-stage MCMC with a coarse-grid filter to enhance the acceptance rate of MCMC, and layer models are used in their work. Another example is given by Gebraad et al. (2020), Bayesian elastic FWI based on Hamiltonian Monte Carlo. In our work, we do not use any sparse tool to perform the stochastic

global optimization FWI for a 2D acoustic time-lapse model.

In this study, we will propose a Bayesian time-lapse FWI based on the MCMC algorithm. Performing the stochastic global optimization FWI with time-lapse data is rarely reported, you can only see in Kotsi et al. (2020) currently. In their work, they are using a single shot and a single frequency data based on a Metropolis-Hastings algorithm also adopted in this work and a local acoustic solver for the forward modeling. Differently, we will use a time-domain multisource shot gather data and a local updating strategy. Moreover, we will propose a new data error prior estimation especially for the time-lapse data according to its feature, and we will have a comparison between the conventional DO time-lapse FWI and the proposed MCMC time-lapse FWI.

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.

BAYESIAN INFERENCE AND MCMC

Bayesian inference

Inversion based Bayesian inference (Tarantola, 2005) can combine both the model prior information and data prior information, and allow the uncertainty estimation of solutions. The prior information of the model $\mathbf{m}$ is expressed by the prior probability density function (PDF) $p(\mathbf{m})$ that should be available before the inversion starting. And the prior information on the observed data $\mathbf{d}_{obs}$ under a given model $\mathbf{m}$ is decribed by the conditional PDF $p(\mathbf{d}_{obs}|\mathbf{m})$, also called likelihood function. With respect to the Bayesian inference frame, the two pieces of prior information guide to the posterior PDF $p(\mathbf{m}|\mathbf{d}_{obs})$ expressed as

$$p(\mathbf{m}|\mathbf{d}_{obs}) = \frac{p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})}{p(\mathbf{d}_{obs})} \propto p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m}), \tag{4}$$

where $p(\mathbf{d}_{obs}) = \int p(\mathbf{d}_{obs}|\mathbf{m})p(\mathbf{m})d\mathbf{m}$ is the marginal likelihood or model evidence, which includes all possibly available information of $\mathbf{m}$ given $\mathbf{d}_{obs}$, and is a constant value normalizing $p(\mathbf{m}|\mathbf{d}_{obs})$. In this study, we do not take $p(\mathbf{d}_{obs})$ into account, since it does not impact the shape of $p(\mathbf{m}|\mathbf{d}_{obs})$.

In our work, we typically assume the prior PDF $p(\mathbf{m})$ and likelihood function $p(\mathbf{d}_{obs}|\mathbf{m})$ conform to Gaussian distributions expressed as

$$p(\mathbf{m}) \propto \exp\{-\frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T\mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0)\}, \tag{5}$$

$$p(\mathbf{d}_{obs}|\mathbf{m}) \propto \exp\{-\frac{1}{2}(\mathbf{d}_{syn}(\mathbf{m}) - \mathbf{d}_{obs})^T\mathbf{C}_d^{-1}(\mathbf{d}_{syn}(\mathbf{m}) - \mathbf{d}_{obs})\}, \tag{6}$$

where $\mathbf{C}_m$ and $\mathbf{C}_d$ are, respectively, prior model and data error covariances. Putting equation 5 and 6 in to equation 4, we rewrite the posterior PDF as

$$p(\mathbf{m}|\mathbf{d}_{obs}) \propto \exp\{-\chi(\mathbf{m})\}, \tag{7}$$

where

$$\chi(\mathbf{m}) = \frac{1}{2}(\mathbf{m} - \mathbf{m}_0)^T\mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0) + \frac{1}{2}(\mathbf{d}_{syn}(\mathbf{m}) - \mathbf{d}_{obs})^T\mathbf{C}_d^{-1}(\mathbf{d}_{syn}(\mathbf{m}) - \mathbf{d}_{obs}). \tag{8}$$

$\chi(\mathbf{m})$ is a misfit function constrained by the prior information.

The Metropolis-Hastings MCMC algorithm

Since the seismic acquisition geometry is limited on space distribution, the inverse problem of FWI is strongly nonlinear. In the case of lacking a good starting model, the DO methods with a local searching capability are often incompetent. Therefore, introducing Monte Carlo methods (Mosegaard and Tarantola, 1995) that have the global searching capability is natural and necessary. MCMC achieves by constructing a Markov chain using an expected distribution as the distribution of equilibrium states that are treated as the samples of the expected distribution. The Metropolis-Hastings (MH) MCMC algorithm (Hastings, 1970) as a classical algorithm can be summarized as follows:

1) Use transitional probabilities $T(\mathbf{m}^*|\mathbf{m})$ to produce the proposal model $\mathbf{m}^*$, in which $\mathbf{m}$ is the present model and T is the proposal distribution.

2) Calculate the acceptance probability

$$a(\mathbf{m}^*, \mathbf{m}) = min(1, \frac{T(\mathbf{m}|\mathbf{m}^*)p(\mathbf{m}^*|\mathbf{d}_{obs})}{T(\mathbf{m}^*|\mathbf{m})p(\mathbf{m}|\mathbf{d}_{obs})}). \tag{9}$$

Since we take T as a uniform distribution that is a symmetric random walk sampler, which means $T(\mathbf{m}|\mathbf{m}^*) = T(\mathbf{m}^*|\mathbf{m})$, then the the acceptance probability $a(\mathbf{m}^*, \mathbf{m})$ can be rewriten as

$$a(\mathbf{m}^*, \mathbf{m}) = min(1, \frac{p(\mathbf{m}^*|\mathbf{d}_{obs})}{p(\mathbf{m}|\mathbf{d}_{obs})}). \tag{10}$$

3) Generate a random number u according to the uniform distribution $U(0, 1)$.

4) If $a > u$, accept $\mathbf{m}^*$ as the present model $\mathbf{m}$; if $a <= u$, reject $\mathbf{m}^*$ and keep $\mathbf{m}$ as the present model.

5) Iterate 1) - 4) until the set number of iterations is reached, then abstract the equilibrium states as the samples of the expected distribution $p(\mathbf{m}|\mathbf{d}_{obs})$.

Note that we are using the MH algorithm in a special case that the proposal distribution is symmetric, which is also called the Metropolis algorithm in such case, that is first presented by (Metropolis et al., 1953).

INVERSION STRATEGIES

Double-difference time-lapse FWI

Thus in this study, we use the DDFWI that can focus on the time-lapse target and reduce coherent errors. In DDFWI, the first inversion is a baseline inversion for the baseline model, in which the input elements are the baseline data and a reasonable starting model. And in the second monitoring inversion for the monitoring model, DDFWI uses the inverted baseline model as the starting model and a composited data as an alternative to the monitoring data. The composited data is given by

$$\mathbf{d}'_{obs2} = \mathbf{b}_{syn}(\mathbf{m}'_1) + (\mathbf{d}_{obs2} - \mathbf{d}_{obs1}), \tag{11}$$

where $\mathbf{d}_{syn}(\mathbf{m}_1')$ is synthetic data of the inverted baseline model $\mathbf{m}_1'$, $\mathbf{d}_{obs2}$ and $\mathbf{d}_{obs1}$ are observed monitoring data and baseline data, respectively. Then during the monitoring inversion, the mistfit function of DDFWI can be expressed as

$$E_{DDFWI}(\delta\mathbf{m}') = \frac{1}{2}||\mathbf{d}_{syn}(\mathbf{m}_1' + \delta\mathbf{m}') - \mathbf{d}_{obs2}'||^2, \tag{12}$$

where $||\ ||^2$ denots L2 norm, $\delta\mathbf{m}'$ is the model perturbation under the background model $\mathbf{m}_1'$. Defining

$$\delta\mathbf{d}_{syn}(\delta\mathbf{m}') = \mathbf{d}_{syn}(\mathbf{m}_1' + \delta\mathbf{m}') - \mathbf{d}_{syn}(\mathbf{m}_1'), \tag{13}$$

then

$$\mathbf{d}_{syn}(\mathbf{m}_1' + \delta\mathbf{m}') = \delta\mathbf{d}_{syn}(\delta\mathbf{m}') + \mathbf{d}_{syn}(\mathbf{m}_1'), \tag{14}$$

where $\delta\mathbf{d}_{syn}(\delta\mathbf{m}')$ is the synthetic difference data. Putting equation 11 and 14 into equation 12, we have

$$E_{DDFWI}(\delta\mathbf{m}') = \frac{1}{2}||\delta\mathbf{d}_{syn}(\delta\mathbf{m}') - \delta\mathbf{d}||^2, \tag{15}$$

where $\delta\mathbf{d} = \mathbf{d}_{obs2} - \mathbf{d}_{obs1}$ is the observed difference data. Therefore, when we are minimizing equation 12, actually, we are minimizing the residual of two difference data, that makes DDFWI focus on the target area. And finally, the inverted time-lapse model is $\delta\mathbf{m}'$.

But we have to note that the inverted time-lapse model $\delta\mathbf{m}'$ is different from the true time-lapse model $\delta\mathbf{m}$. $\delta\mathbf{m}'$ corresponds to the difference wavefield $\delta\mathbf{d}_{syn}(\delta\mathbf{m}')$ under the background wavefield $\mathbf{d}_{syn}(\mathbf{m}_1')$, but $\delta\mathbf{m}$ corresponds to the difference wavefield $\delta\mathbf{d}$ under the background wavefield $\mathbf{d}_{obs1}$. When $\mathbf{d}_{syn}(\mathbf{m}_1')$ is close to $\mathbf{d}_{obs1}$, i.e., $\mathbf{m}'$ is close to the true baseline model, $\delta\mathbf{m}'$ will be close to $\delta\mathbf{m}$, and vice versa. Thus a good inverted baseline model is important for DDFWI (Asnaashari et al., 2011). And how the input starting monitoring model impacts the final inverted time-lapse model will be discussed in this study.

MCMC time-lapse FWI

For a conventional DO DDFWI, both twice inversions are based on the DO FWI, but in this study, we develop a new workflow referred to MCMC DDFWI. In MCMC DDFWI, we use the DO FWI to do the first baseline inversion, then use the Baysian FWI using HM MCMC to perform the second monitoring inversion. Then the difference between the inverted monitoring model and the inverted baseline model is the inverted time-lapse model. In this study, we will have a competition between the conventional DO DDFWI and the MCMC DDFWI.

Local-updating target-oriented time-lapse inversion

Although DDFWI is a target-oriented method focusing on the time-lapse change, it requires high repeatability between baseline and monitoring surveys. The differences, such as the different source wavelets, different acquisition geometries, and different noises between the twice surveys will cause distractions outside the target area (time-lapse change area). An additional local constraint is helpful for DDFWI to further focus on the target

area. Before using a local constraint, we need to detect the target area. Detecting the target area is possible by combining the depth migration of the observed data difference and some prior information of the reservoir. Raknes and Arntsen (2014) use a simple edge detection algorithm to the depth migration section to automatically detect the target area, and keep the model updating only in the target area. Asnaashari et al. (2014) apply the Gaussian function to build a weighting matrix to constraining the target area, after manually and approximately detecting the center of the target area. In this study, we also manually choose the target area and keep the model updating only in the chosen area.

Currently, more advanced techniques are employing the local solvers (Robertsson and Chapman, 2000; Yang et al., 2012; Huang et al., 2018) that can avoid the wavefield calculation for the whole model space and constrain the wavefield calculation in a local space, it is a good time-saving way to perform FWI. However, in local-solver-based FWI, a number of Green's functions need to be calculated and the local solver for multisource seismic data that we will use in this study has not been reported. In our study, we still calculate the wavefield for the whole model space for the reasons of easy operation and using the multisource seismic data.

For the conventional DO DDFWI, a target-oriented inversion is a hopeful tool that can avoid the artifacts outside of the target area, but for the MCMC DDFWI, it could be vital. Since the local-updating target-oriented strategy can dramatically reduce the number of unknowns. For a standard MH MCMC, the cost of producing independent samples raises with $O(n^2)$ when the parameter number (dimension) increases by n (Creutz, 1988), which means the target-oriented inversion strategy can easily save computing time by tens of times. Another critical reason for using this strategy is that most MH MCMC algorithms tend to fail for high-dimension problems (Chib and Greenberg, 1995), whereas the dimension of the inverse problem of FWI is often beyond thousands for a small 2D model or tens of thousands for a regular 2D model.

Multisource waveform inversion

Multisource waveform inversion is developed to save computing time of FWI, it excites multiple sources at the same time rather than generating shot gathers one by one. For DO FWI, we have to use some special treats (e.g., the encoded multisource method of (Krebs et al., 2009)) to each source to decay the crosstalk artifacts in the final inverted model. But for MCMC-based FWI, these treats are not necessary, since we have no middle produce of gradient calculating, through which the crosstalks between shot gathers will be transferred to the model as crosstalk artifacts. MCMC-based FWI decreases data misfit by directly accepting the stochastic samples. Thus in MCMC DDFWI, we can more freely employ multisource seismic data and do not need to worry about the model crosstalk artifacts.

Random Walk sampler

Random walk sampling is initiated by Metropolis et al. (1953), in which the proposal model $\mathbf{m}^*$ in equation 9 is given by

$$\mathbf{m}^* = \mathbf{m} + \mathbf{X}, \tag{16}$$

where $\mathbf{m}$ is the present model, $\mathbf{X}$ is a random number drawn from the proposal distribution that is a uniform distribution with zero mean value and a proper standard deviation in our study.

Normally, the present model $\mathbf{m}$ at the first time is input as a random model. However, for the strongly nonlinear and high-dimensional FWI, the random start-up model means a lot of extra computation and even failure. Thus using the inverted baseline model from DO FWI as the start-up model of the monitoring inversion of MCMC DDFWI is a guarantee for the success of MCMC DDFWI.

In addition, we use a component-wise scheme to update the model, updating only one parameter rather than all at a time and all parameters are updated sequentially, since the model dimension is pretty high. Even though we are using the target-oriented inversion strategy, the dimension can easily be as many as hundreds or even thousands. Generating a perturbation $\mathbf{X}$ that suitable for all parameters is rarely possible under such model dimension, as different parameters have different behaviors. For example, the target area of a time-lapse model contains at least two types of parameters, one has model change and the other has no model change, the former prefers to accept the perturbation when the latter prefers to preserve the start-up model, and the situation will be more complicated in practice. Thus updating all parameters at a time is effectless here.

Model prior information

In the Bayesian inference frame described in equation 7 and 8, it requires both priors for data and model, that is, data error covariance $\mathbf{C}_d$ and the model covariance $\mathbf{C}_m$. In our work, $\mathbf{C}_m$ is assumed as a diagonal matrix corresponding to independently distributed model parameters. $\mathbf{C}_m$ can be obtained by an adaptive Metropolis (AM) algorithm developed by Haario et al. (2001). In which the mean values and model covariance are calculated from all previous models. In our work, we utilize only the calculated model covariance and the mean value is taken as the present model. This is similar to the adaptive proposal algorithm (Haario et al., 1999) in which the proposal distribution is centered on the present model, and the covariance is calculated from a fixed finite number of previous models.

In accordance with the AM algorithm, the covariance for iteration $k+1$ proposal model is given by

$$\mathbf{C}_{k+1} = \frac{k-1}{k}\mathbf{C}_k + \frac{s_m}{k}(k\bar{\mathbf{m}}_{k-1}\bar{\mathbf{m}}_{k-1}^T - (k+1)\bar{\mathbf{m}}_k\bar{\mathbf{m}}_k^T + \mathbf{m}_k\mathbf{m}_k^T + \epsilon\mathbf{I}_m), \tag{17}$$

where $\bar{\mathbf{m}}_k = (1/(k+1)\sum_{i=0}^{k}\mathbf{m}_i)$ is the average model, s_m is a small constant that depends only on model dimension m, $\mathbf{I}_m$ is a m-dimension identity matrix, ϵ is a positive constant. Hereon, $\epsilon\mathbf{I}_m$ is to avoid $\mathbf{C}_{k+1}$ to be singular, and s_m is typically adopted as $s_d = 2.4^2/m$ suggested by Gelman et al. (1996). Equation 17 is a recursion formula that consumes little computation. At the beginning period, we use $\mathbf{C}_{k+1} = \sigma_{m0}\mathbf{I}_m$ as an alternative of equation 17, in which σ_{m0} is a reasonable constant from some prior information or by trial-and-error method, after enough models are obtained, we switch to equation 17. Furthermore, all models are constrained in given bounds.

Data prior information

In this subsection, we propose a new method to calculate the data prior information, that is, data error covariance $\mathbf{C}_d$ in equation 7 and 8. When considering the data error to be identical and independently distributed, data error covariance can be expressed as $\mathbf{C}_d = \sigma_d \mathbf{I}_s$ where standard deviations of all data points are given as one value σ_d and I_d is an identity matrix with the same size as $\mathbf{C}_d$. In most cases, data errors are often entangled with the effective data corresponding to underground media, so it is difficult to separate them directly. Usually, people need to estimate σ_d by trial-and-error method or treating it as an unknown updating during the inversion, such as the hierarchical inversion method (Malinverno and Briggs, 2004; Bodin et al., 2012).

For our case, things become easier, since we can separate data errors and effective data according to the feature of time-lapse data. In DDFWI, the composited data in equation 11 is constructed by the synthetic data and the difference data. All noises in the composited data come from the difference data in which we can easily distinguish where the effective data are. Thus we use the part containing no effective data, called the pure noise part later, to estimate the distribution of data errors. In our work, we assume that data errors conform to zero-mean Gaussian distribution (it may change with cases), so we only need to calculate σ_d from the pure noise part. This method works depending on two requirements are satisfied, the first is that the baseline and monitoring surveys have good repeatability, so the pure noise part can be distinguished, and the second is that the pure noise part contains enough samples, so the estimated distribution is representative. In practice, the first requirement may not easy to meet but the second one is not a problem. Satisfying the first requirement is also key to perform DDFWI successfully. In our study, we will use synthetic baseline and monitoring data with the same acquisition geometry. For practice, if the acquisition geometries for twice surveys are different, some processes Hicks (2002); Yang et al. (2016); Fu et al. (2020) should be taken.

NUMERICAL EXAMPLES

To demonstrate the feasibility of our method on both model inversion and uncertainty qualification, we show some numerical examples using a 2D constant-density acoustic time-lapse model in this section. The models and corresponding multisource shot gather seismic data used are displayed in Figure1a-f. The model size is 50-by-50, model spacing is 10m, a 40m/s identical time-lapse velocity change is centered in the time-lapse model (Figure 1c, the monitoring model minus the baseline model), 10 sources are spread on the surface, receivers are put on the surface and two sides, all sources are excited simultaneously using an identical Ricker wavelet with a 40Hz central frequency. The method used to solve the wave equation is a time-domain finite difference method that is eighth-order in space and second-order in time. The black box in Figure 1c is the target area where model updating is taken, which is of a size of 19-by-22 and contains 418 parameters. So the dimension is 418 that is much less than 2500, the dimension for the whole model. In the monitoring inversion of MCMC DDFWI, we run 8 chains simultaneously with the same starting model to obtain a number of samples quickly, and each chain is sampled 83600 times. And the burn-in period contains the first 21900 models that will not be used.

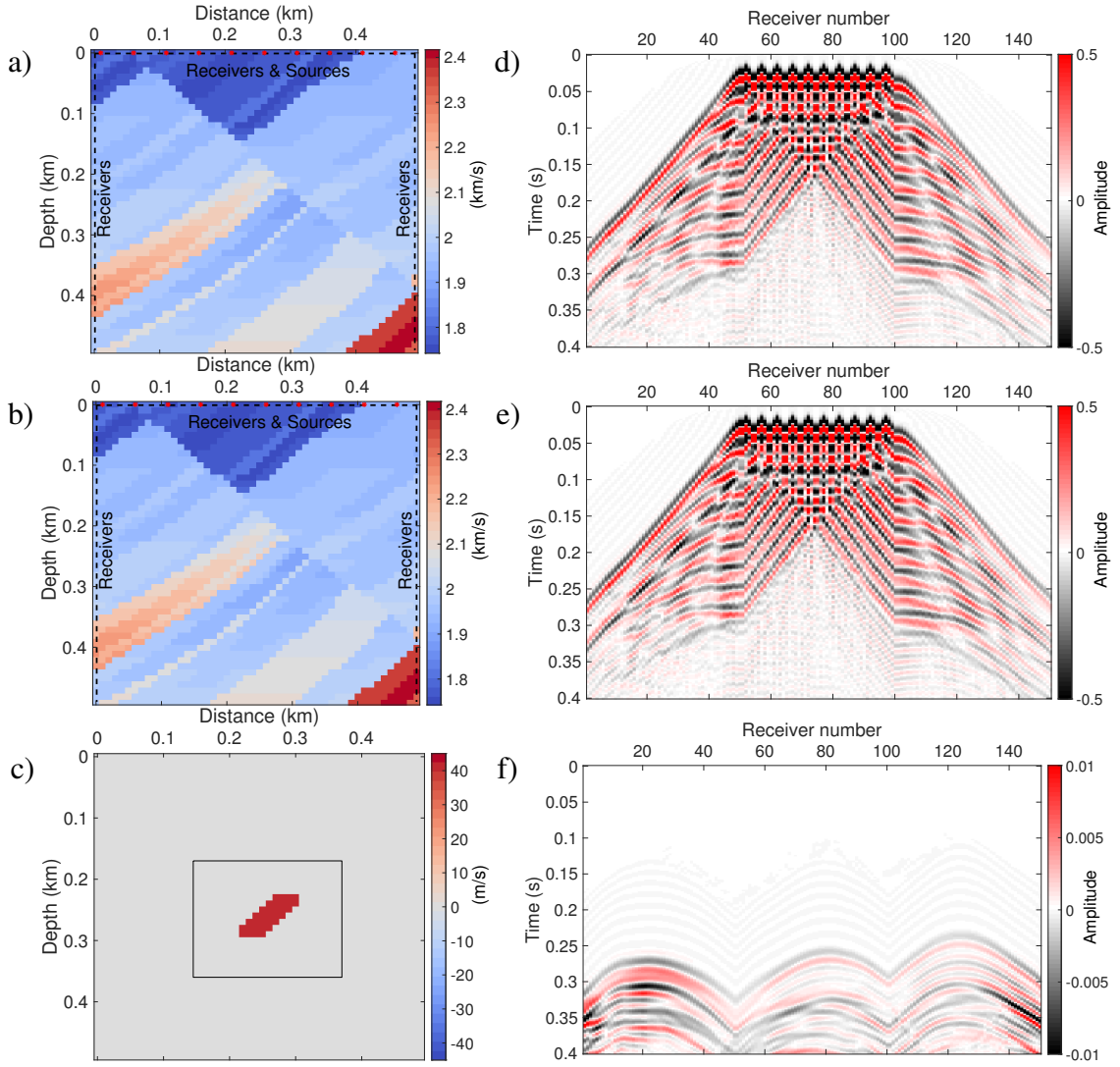

FIG. 1. (a), (b), and (c) are baseline model, monitoring model, and time-lapse model, respectively. Accordingly, (d), (e), and (f) are baseline data, monitoring data, and difference data, respectively. The acquisition geometries in (a) and (b) are identical. The black dash lines are the receivers, and the red stars are sources. The black box in (c) is the target area where model updating is taken.

DO DDFWI vs MCMC DDFWI

We use four different starting models (Model 1, 2, 3, and 4) displayed in Figure 2a-d and noise-free monitoring data as the input of the monitoring inversion for both DO DDFWI and MCMC DDFWI. The four starting models include the true baseline model, the inverted baseline model of DO FWI, and two smooth models. The inverted baseline model is obtained from the DO FWI using noise-free baseline data and a smooth starting model in Figure 2c. In Figure 2, the models become worse from left to right. All the results of MCMC DDFWI in Figure 3 are average models, and the corresponding absolute model errors of Figure 3 are displayed in Figure 4. From Figure 3, we can see that, compared with DO DDFWI, the results of MCMC DDFWI contain more random noises that become heavier with the starting model becomes worse. To reduce the random noise, we use a median filter to the results of MCMC DDFWI and also use the same filter to the results

of DO DDFWI for fair comparisons. From Figure 3 and 4, we can see that as the starting model becomes worse the results also become worse, and compared with the results of DO DDFWI, that of MCMCM DDFWI have clearer edges of nonzero model change, less coherent model errors. And in Figure 5, we display the curves of L2 norm of the monitoring data misfit of MCMC DDFWI versus sampling number for different starting models, which shows that all chains converge well and a better starting model gives a lower data misfit level.

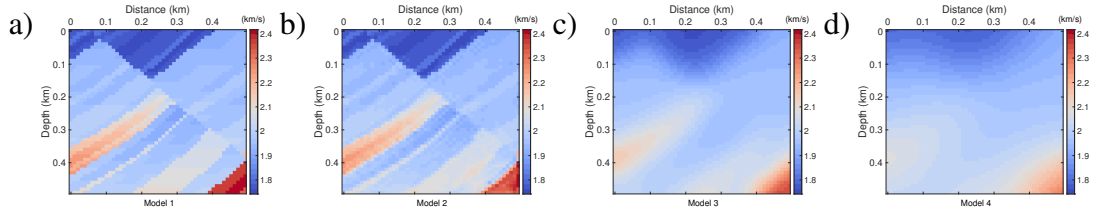

FIG. 2. Different starting models (Model 1, 2, 3, 4) for monitoring model inversion. (a) is the true baseline model. (b) is the inverted baseline model of the DO FWI using (d) as the starting model, (c) and (d) are models from the true baseline model after using a Gaussian filter with window size 50m and 100m, respectively. The models become worse from left to right.

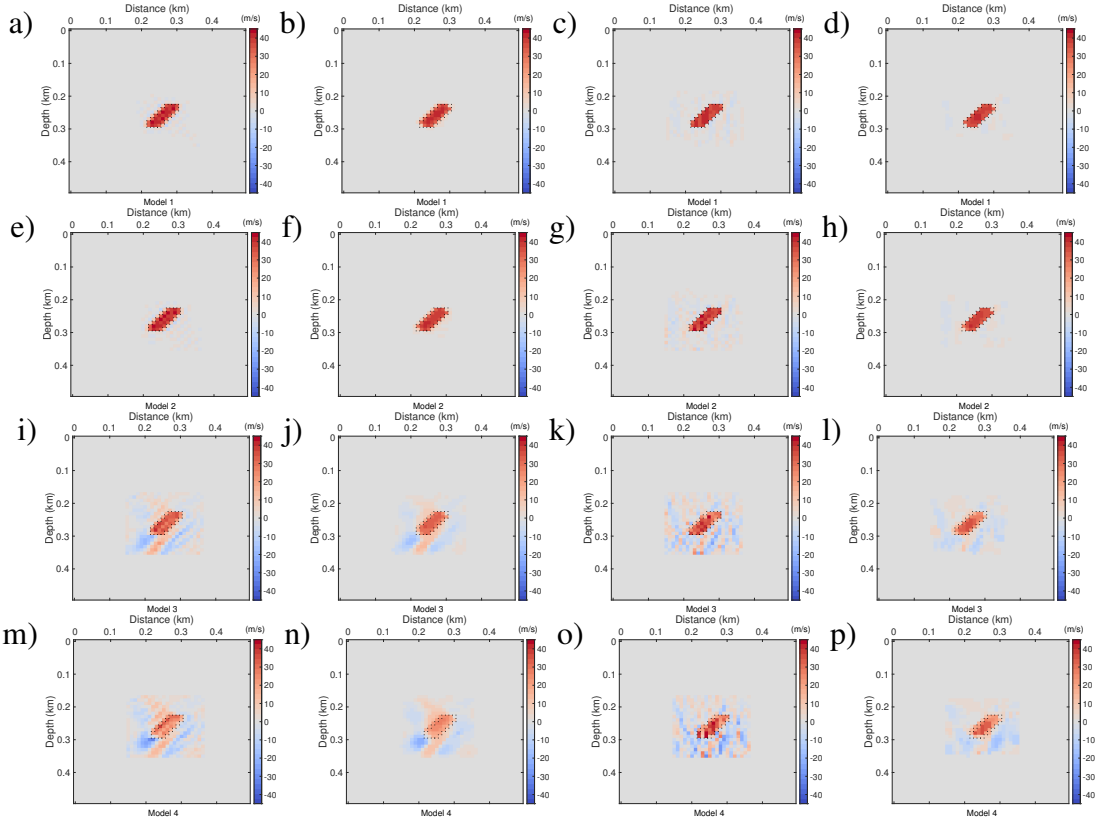

FIG. 3. Inverted time-lapse models of DO DDFWI and MCMC DDFWI from different starting models in Figure 2. Row 1, 2, 3, and 4 are results, respectively, for model 1, 2, 3, and 4. The results in column 1 are inverted from DO DDFWI, and that in column 2 are obtained by using a median filter to results in column 1. The results in column 3 are inverted from MCMC DDFWI, and that in column 4 are obtained by using a median filter to results in column 3. The black dot line box is the edge of the non-zero model change.

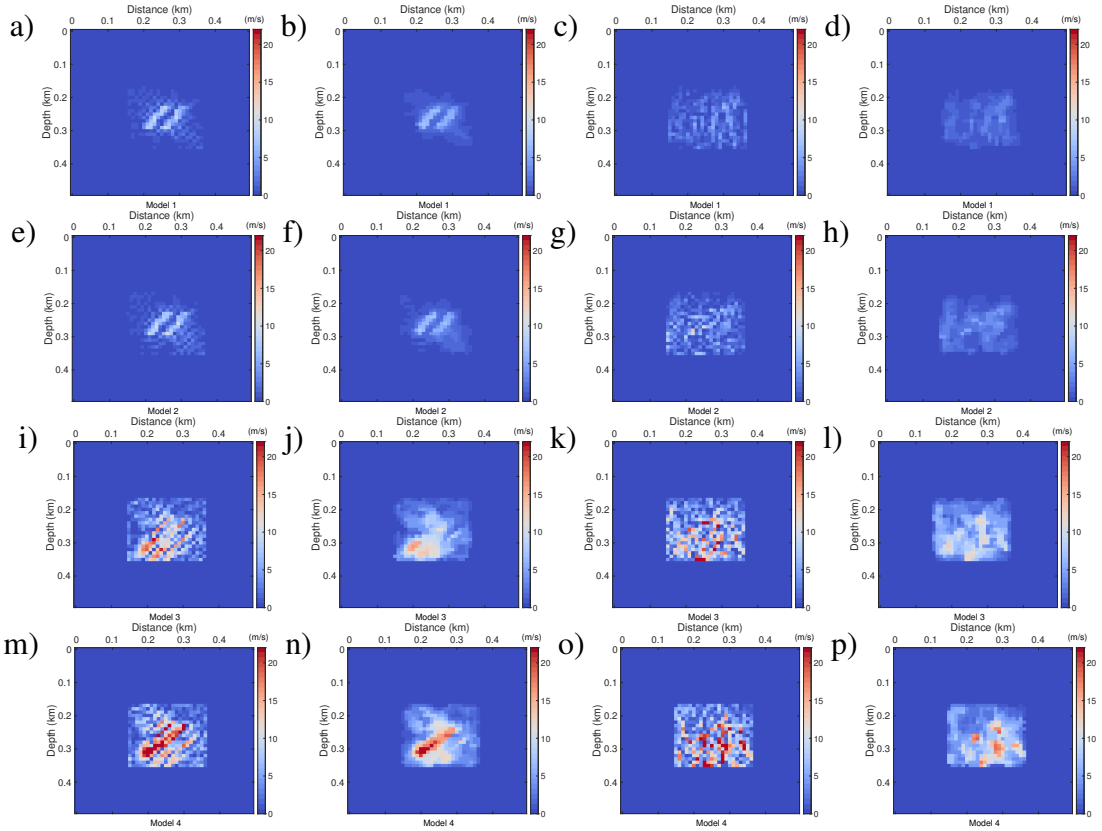

FIG. 4. (a)-(p) are, respectively, absolute values of model errors of results in Figure 3a-p.

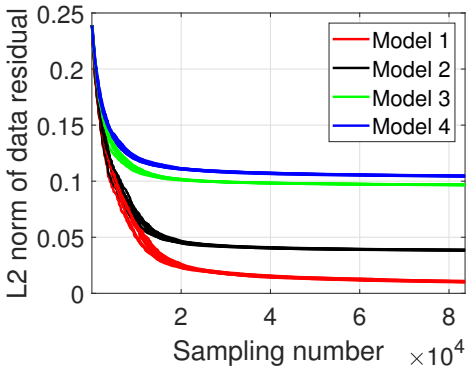

FIG. 5. Curves of L2 norm of the monitoring data misfit of MCMC DDFWI versus sampling number for different starting models.

Noisy data tests for MCMC DDFWI

In this subsection, we use data with different noise levels to test our method. The noises we are adding are zero-mean Gaussian noise with standard deviations $2.5*10^{-4}$, $2.5*10^{-3}$, and $5*10^{-3}$, respectively, and the standard deviation of the noise-free difference data is $1.2*10^{-3}$. The noisy difference data are displayed in Figure 6a-c. The data errors standard deviations are estimated almost accurately from the pure noise parts in Figure 6a-c. The starting model input is Model 2 (Figure 2b). Figure 7 displays curves of the data misfit versus sampling number for eight chains, from which we can see that with the increase of

noise, data misfit converges to higher values and fluctuates more intensely. In Figure 8, we display the inverted time-lapse models including average models and the maximum posterior probability (MAP) models, and the corresponding posterior model standard deviations. With the increase of noise level, average models get worse but can still show the edges of the nonzero area clearly, and they have a better quality than MAP models. The posterior model standard deviations also rise up as the increase of noise level, which depend on not only the value of time-lapse change but also the position in the model. Figure 9 displays the posterior distributions of 6 parameters located at the black dots in Figure 8a, including three with nonzero true time-lapse change and three with zero true time-lapse change. In Figure 9a, since the model errors standard deviation is too small and the rejection rate is too high, we do not obtain enough effective models to characterize a reasonable distribution for some parameters. But in Figure 9b and c, they show that the inverted models can match the standard Gaussian distribution well. In Figure 10 and 11, we display the predicted noise-free difference data and the noises, both are very close to the true values.

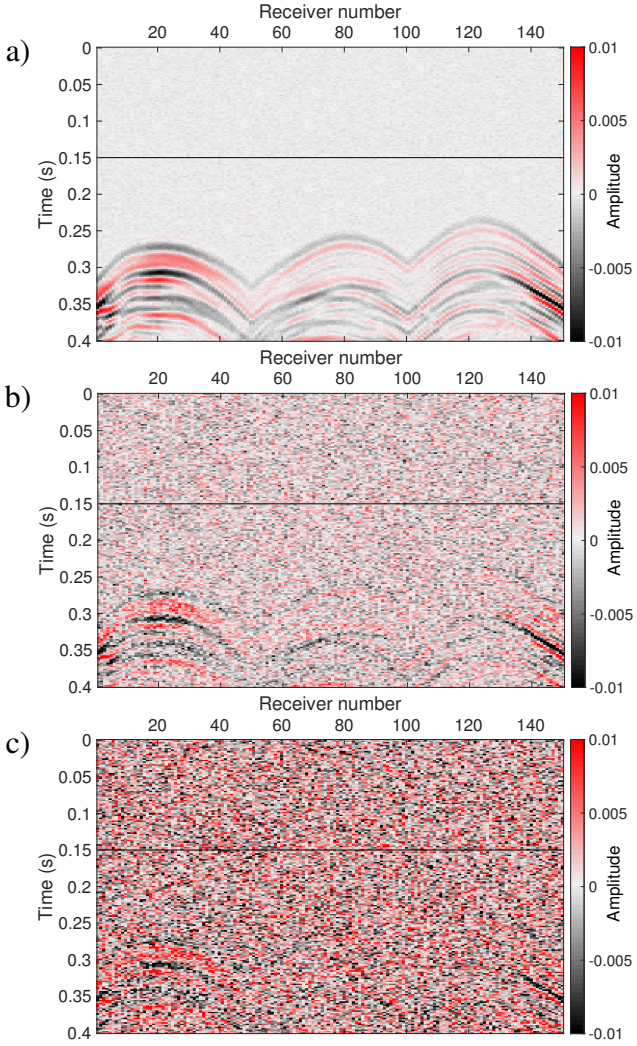

FIG. 6. The difference data with different noise levels. Noises confirm to zero-mean Gaussian distributions with different standard deviations that are, respectively, $2.5 * 10^{-4}$ for (a), $2.5 * 10^{-3}$ for (b), and $5 * 10^{-3}$ for (c). The data above the black line in each panel is the pure noise part used to estimate the data errors standard deviations.

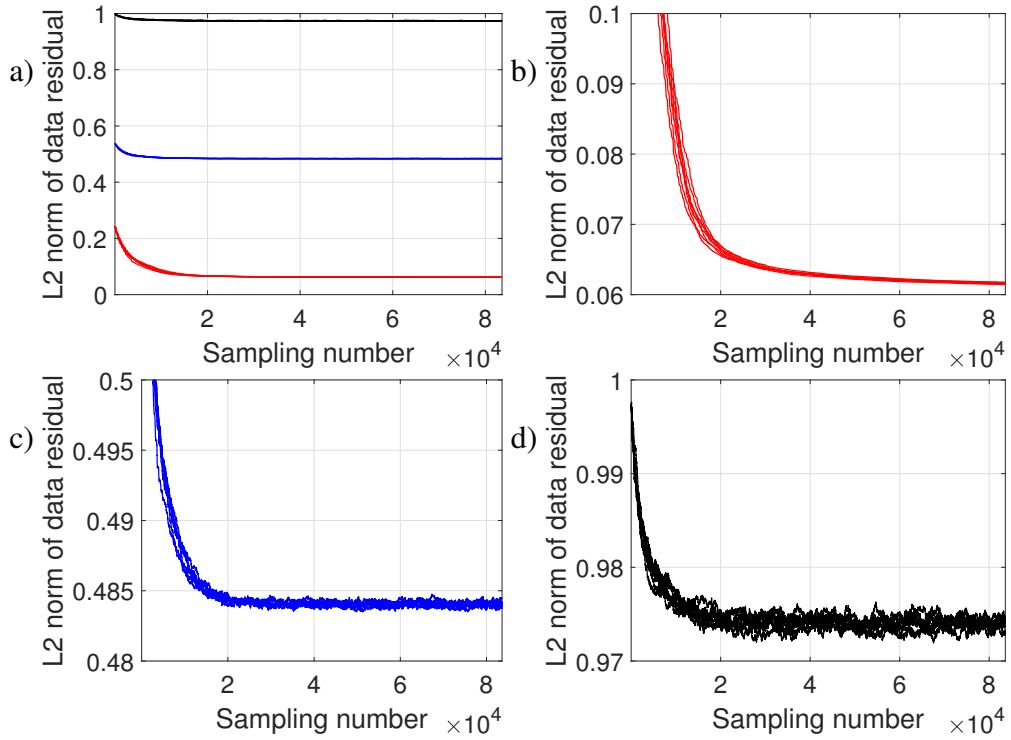

FIG. 7. Curves of L2 norm of the monitoring data misfit of MCMC DDFWI versus sample number for different level noisy data. The red, blue, and black curves, respectively, correspond to the data in Figure 6a, b, and c. All curves are displayed together in panel a and separately displayed in panel b-d.

DISCUSSION

Using global stochastic optimization methods for FWI is still a time-consuming mission. In our work, we are using a small time-lapse acoustic 2D model, so the computing time is acceptable. It takes near three hours on an ordinary desktop. But when considering the whole 2D model, a 2D elastic model, a large 2D model, or even a 3D model, more advanced strategies should be taken, such as more powerful global stochastic algorithms (e.g., Hamiltonian Monte Carlo), high-performance computers (e.g., GPU), faster wave equation solvers, parameter dimension reducing.

CONCLUSIONS

In this study, we have proposed a time-lapse FWI based on an MH MCMC algorithm and a new method to estimate the data error standard deviation for time-lapse data according to the feature of difference data. To achieve the MCMC-based time-lapse FWI, we have employed the inversion strategies including DDFWI, multisource data, local-updating target-oriented inversion, calculating model covariance with the AM algorithm, and the new data error standard deviation estimation. Synthetic data tests using a 2D acoustic model have demonstrated the feasibility of both model inversion and uncertainty quality of our method. We also have compared our method with the conventional DO DDFWI, which shows that MCMC DDFWI can provide the results with clearer edges of the nonzero time-lapse model change and fewer coherent errors.

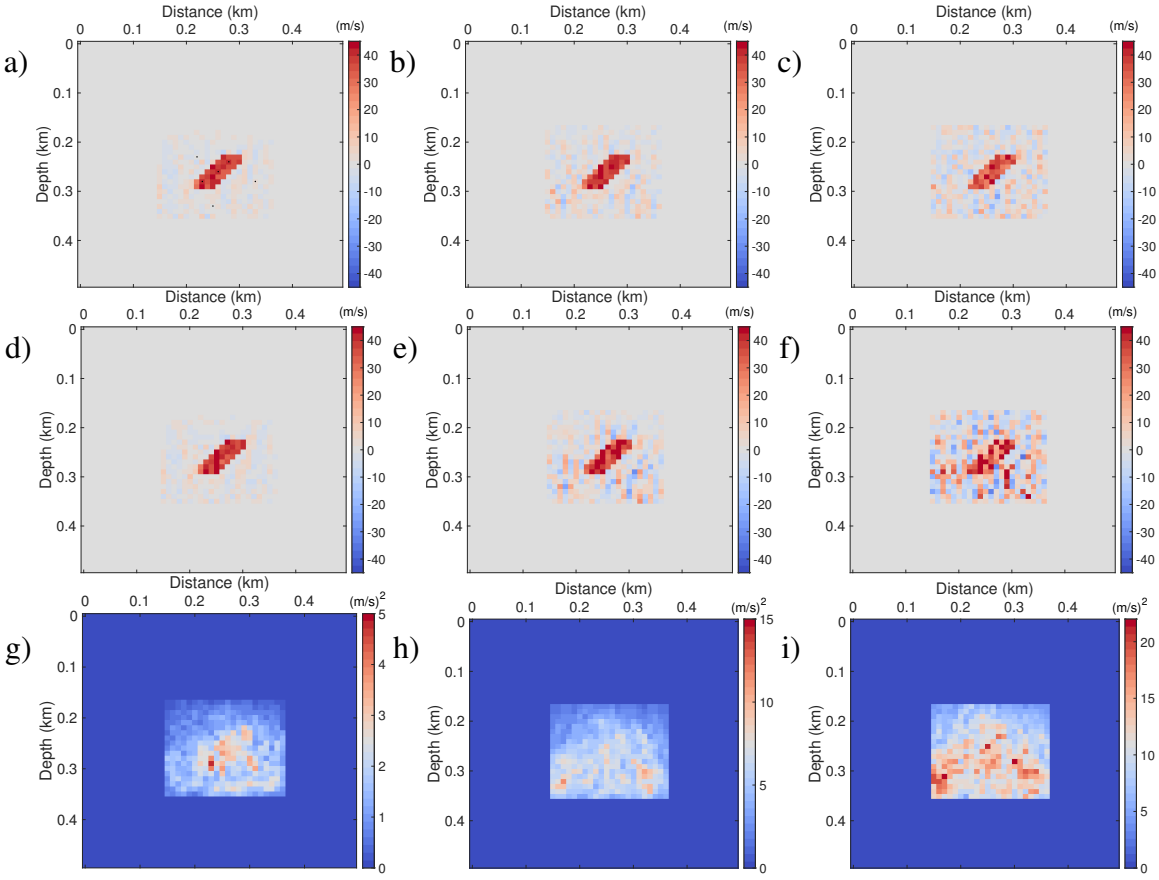

FIG. 8. Inverted time-lapse models and posterior model standard deviations of MCMC DDFWI using Model 2 as the starting model and data with different noise levels. (a), (b), and (c) are average models using data in Figure 6a, b and c, respectively. (d), (e), and (f) are MAP models using data in Figure 6a, b and c, respectively. (g), (h), and (i) are posterior model standard deviations using data in Figure 6a, b and c, respectively. The six black dots in (a) are the positions where we will appraise the posterior distributions in Figure 9.

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.

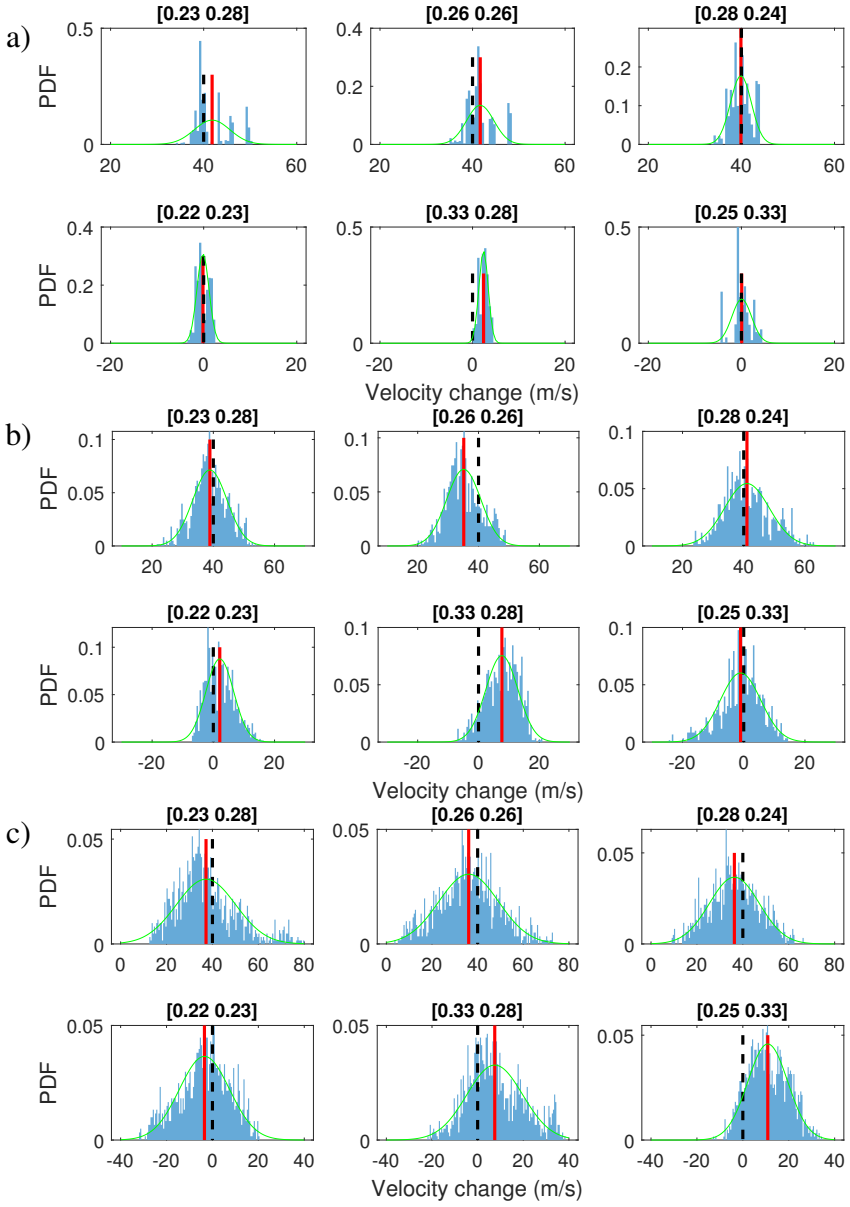

FIG. 9. The posterior distributions of 6 parameters located at the black dots in Figure 8a, including three with nonzero true time-lapse change and three with zero true time-lapse change. (a), (b), and (c), respectively, correspond to the noisy data in Figure 6a, b, and c. Blue histograms denote all samples excepting the burn-in ones abstracted from the eight chains. Green curves are standard Gaussian probability density functions (PDFs) best fitting the histograms. Black dash lines and red solid lines denote true values and means, respectively. The title above each panel denotes the parameter location in the model.

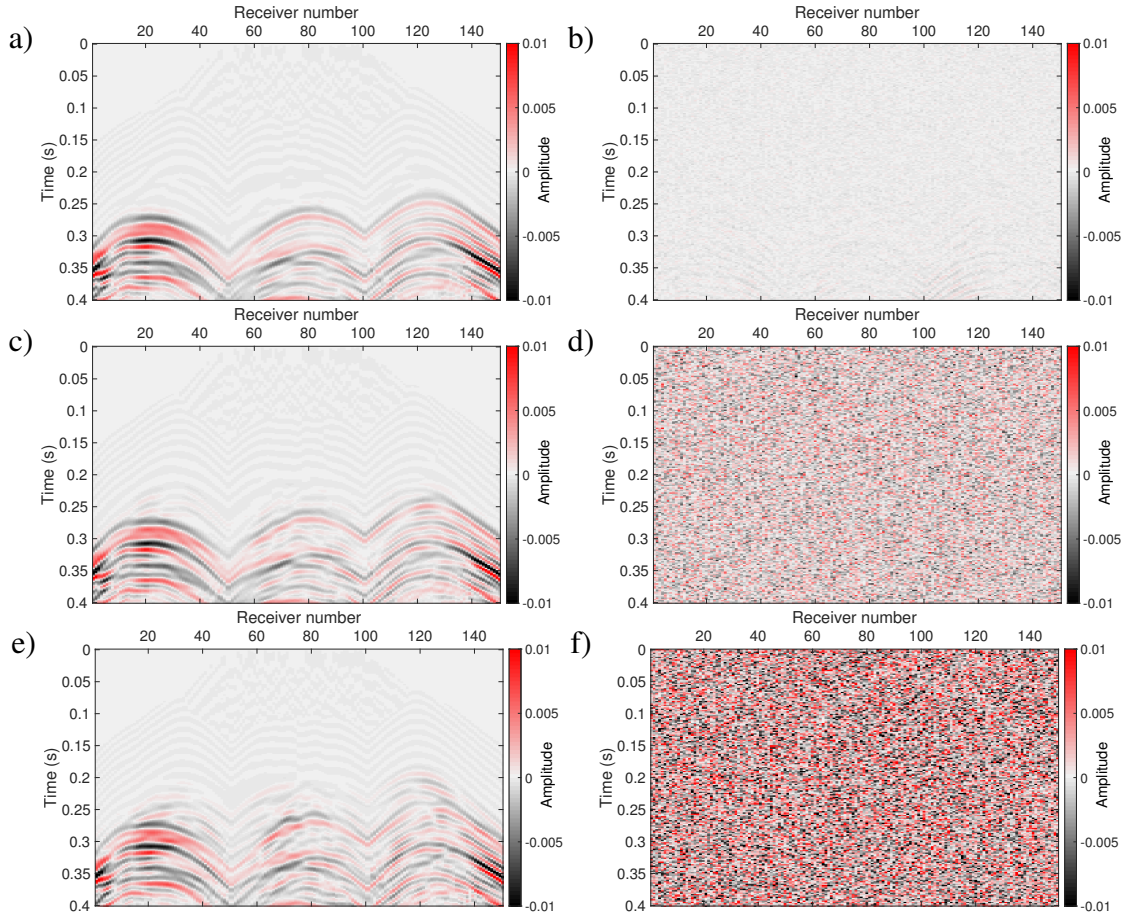

FIG. 10. The predicted noise-free difference data (a, c, e) and noises (b, d, f) using the inverted average models in Figure 8a, b, and c, respectively. The predicted data in (a) and the predicted noise in (b) correspond to the noisy data in Figure 6a. (c) and (d) correspond to the noisy data in Figure 6b. And (e) and (f) correspond to the noisy data in Figure 6c.

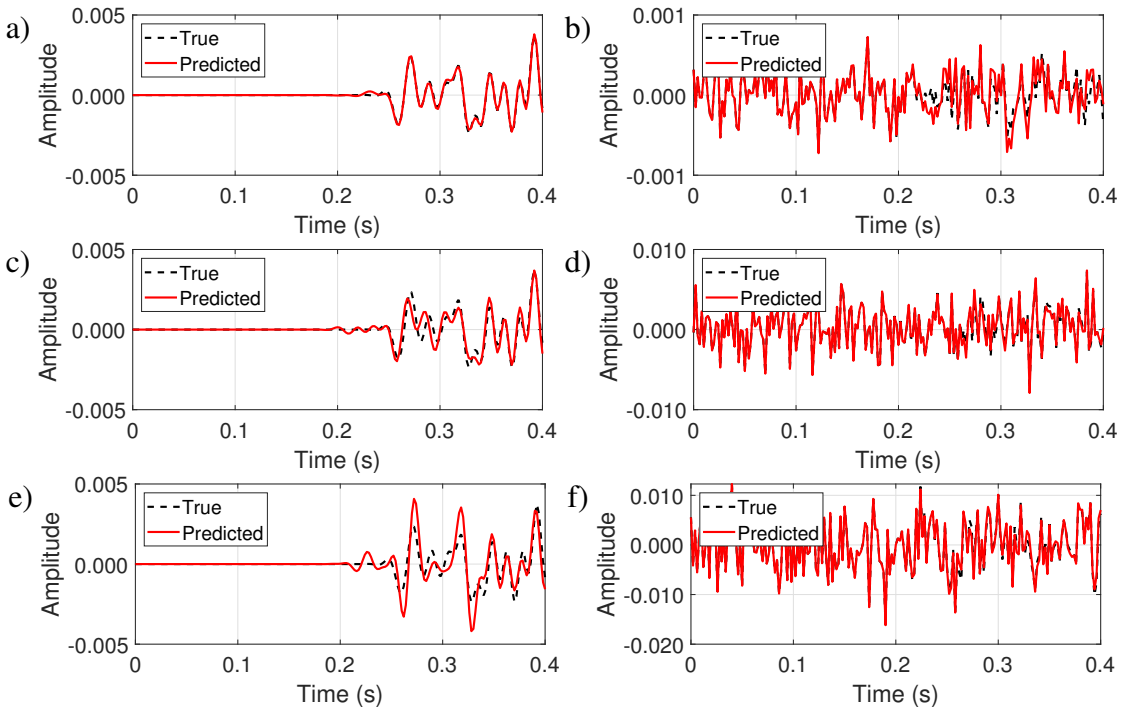

FIG. 11. Red curves in (a)-(f) are central traces of Figure 10a-f,respectively. Black dash curves are the corresponding true values. The panels in the left column display the noise-free data, and that in the right column display the noises.

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.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J., 2011, Sensitivity analysis of time-lapse images obtained by differential waveform inversion with respect to reference model, *in* SEG Technical Program Expanded Abstracts 2011, Society of Exploration Geophysicists, 2482–2486.

Asnaashari, A., Brossier, R., Garambois, S., Audebert, F., Thore, P., and Virieux, J., 2014, Time-lapse seismic imaging using regularized full-waveform inversion with a prior model: which strategy?: Geophysical prospecting, **63**, No. 1, 78–98.

Biswas, R., and Sen, M., 2017, 2d full-waveform inversion and uncertainty estimation using the reversible jump hamiltonian monte carlo, *in* SEG Technical Program Expanded Abstracts 2017, Society of Exploration Geophysicists, 1280–1285.

Bodin, T., Sambridge, M., Rawlinson, N., and Arroucau, P., 2012, Transdimensional tomography with unknown data noise: Geophysical Journal International, **189**, No. 3, 1536–1556.

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

Chib, S., and Greenberg, E., 1995, Understanding the metropolis-hastings algorithm: The american statistician, **49**, No. 4, 327–335.

Creutz, M., 1988, Global monte carlo algorithms for many-fermion systems: Physical Review D, **38**, No. 4, 1228.

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.

Denli, H., and Huang, L., 2009, Double-difference elastic waveform tomography in the time domain, *in* SEG Technical Program Expanded Abstracts 2009, Society of Exploration Geophysicists, 2302–2306.

Ely, G., Malcolm, A., and Poliannikov, O. V., 2018, Assessing uncertainties in velocity models and images with a fast nonlinear uncertainty quantification method: Geophysics, **83**, No. 2, R63–R75.

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.

Fu, X., Romahn, S., and Innanen, K., 2020, Double-wavelet double-difference time-lapse waveform inversion, *in* SEG Technical Program Expanded Abstracts 2020, Society of Exploration Geophysicists, 3764–3767.

Gebraad, L., Boehm, C., and Fichtner, A., 2020, Bayesian elastic full-waveform inversion using hamiltonian monte carlo: Journal of Geophysical Research: Solid Earth, **125**, No. 3, e2019JB018,428.

Gelman, A., Roberts, G., and Gilks, W., 1996, Efficient metropolis jumping hules: Bayesian statistics.

Haario, H., Saksman, E., and Tamminen, J., 1999, Adaptive proposal distribution for random walk metropolis algorithm: Computational Statistics, **14**, No. 3, 375–396.

Haario, H., Saksman, E., Tamminen, J. et al., 2001, An adaptive metropolis algorithm: Bernoulli, **7**, No. 2, 223–242.

Hastings, W. K., 1970, Monte carlo sampling methods using markov chains and their applications.

Hicks, G. J., 2002, Arbitrary source and receiver positioning in finite-difference schemes using kaiser windowed sinc functions: Geophysics, **67**, No. 1, 156–165.

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

Huang, X., Jakobsen, M., Eikrem, K. S., and Nævdal, G., 2018, A target-oriented scheme for efficient inversion of time-lapse seismic waveform data, *in* SEG Technical Program Expanded Abstracts 2018, Society of Exploration Geophysicists, 5352–5356.

Kotsi, M., Malcolm, A., and Ely, G., 2020, Uncertainty quantification in time-lapse seismic imaging: a full-waveform approach: Geophysical Journal International, **222**, No. 2, 1245–1263.

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.

Maharramov, M., and Biondi, B., 2014, Joint full-waveform inversion of time-lapse seismic data sets, *in* SEG Technical Program Expanded Abstracts 2014, Society of Exploration Geophysicists, 954–959.

Malinverno, A., and Briggs, V. A., 2004, Expanded uncertainty quantification in inverse problems: Hierarchical bayes and empirical bayes: Geophysics, **69**, No. 4, 1005–1016.

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.

Mosegaard, K., and Tarantola, A., 1995, Monte carlo sampling of solutions to inverse problems: Journal of Geophysical Research: Solid Earth, **100**, No. B7, 12,431–12,447.

Oldenborger, G. A., Routh, P. S., and Knoll, M. D., 2007, Model reliability for 3d electrical resistivity tomography: Application of the volume of investigation index to a time-lapse monitoring experiment: Geophysics, **72**, No. 4, F167–F175.

Onishi, K., Ueyama, T., Matsuoka, T., Nobuoka, D., Saito, H., Azuma, H., and Xue, Z., 2009, Application of crosswell seismic tomography using difference analysis with data normalization to monitor co2 flooding in an aquifer: International Journal of Greenhouse Gas Control, **3**, No. 3, 311–321.

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.

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.

Raknes, E. B., and Arntsen, B., 2014, Time-lapse full-waveform inversion of limited-offset seismic data using a local migration regularization: Geophysics, **79**, No. 3, WA117–WA128.

Raknes, E. B., Weibull, W., and Arntsen, B., 2013, Time-lapse full waveform inversion: Synthetic and real data examples, *in* SEG Technical Program Expanded Abstracts 2013, Society of Exploration Geophysicists, 944–948.

Ray, A., Sekar, A., Hoversten, G. M., and Albertin, U., 2016, Frequency domain full waveform elastic inversion of marine seismic data from the alba field using a bayesian trans-dimensional algorithm: Geophysical Journal International, **205**, No. 2, 915–937.

Robertsson, J. O., and Chapman, C. H., 2000, An efficient method for calculating finite-difference seismograms after model alterations: Geophysics, **65**, No. 3, 907–918.

Routh, P., Palacharla, G., Chikichev, I., and Lazaratos, S., 2012, Full wavefield inversion of time-lapse data for improved imaging and reservoir characterization, *in* SEG Technical Program Expanded Abstracts 2012, Society of Exploration Geophysicists, 1–6.

Routh, P. S., and Anno, P. D., 2008, Time-lapse noise characterization by inversion, *in* SEG Technical Program Expanded Abstracts 2008, Society of Exploration Geophysicists, 3143–3147.

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.

Stuart, G. K., Minkoff, S. E., and Pereira, F., 2019, A two-stage markov chain monte carlo method for seismic inversion and uncertainty quantification: Geophysics, **84**, No. 6, R1003–R1020.

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

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

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.

Watanabe, T., Shimizu, S., Asakawa, E., and Matsuoka, T., 2004, Differential waveform tomography for time-lapse crosswell seismic data with application to gas hydrate production monitoring, *in* SEG Technical Program Expanded Abstracts 2004, Society of Exploration Geophysicists, 2323–2326.

Yang, D., Liu, F., Morton, S., Malcolm, A., and Fehler, M., 2016, Time-lapse full-waveform inversion with ocean-bottom-cable data: Application on valhall field: Geophysics, **81**, No. 4, R225–R235.

Yang, D., Meadows, M., Inderwiesen, P., Landa, J., Malcolm, A., and Fehler, M., 2015, Double-difference waveform inversion: Feasibility and robustness study with pressure data: Geophysics, **80**, No. 6, M129–M141.

Yang, D., Zheng, Y., Fehler, M., and Malcolm, A., 2012, Target-oriented time-lapse waveform inversion using virtual survey, *in* SEG Technical Program Expanded Abstracts 2012, Society of Exploration Geophysicists, 1–5.

Zheng, Y., Barton, P., and Singh, S., 2011, Strategies for elastic full waveform inversion of time-lapse ocean bottom cable (obc) seismic data, *in* SEG Technical Program Expanded Abstracts 2011, Society of Exploration Geophysicists, 4195–4200.