GEOPHYSICS, VOL. 84, NO. 6 (NOVEMBER-DECEMBER 2019); P. R1015–R1032, 20 FIGS., 1 TABLE.
10.1190/GEO2018-0893.1

A two-stage Markov chain Monte Carlo method for seismic inversion and uncertainty quantification

Georgia K. Stuart[1], Susan E. Minkoff[1], and Felipe Pereira[1]

ABSTRACT

Bayesian methods for full-waveform inversion allow quantification of uncertainty in the solution, including determination of interval estimates and posterior distributions of the model unknowns. Markov chain Monte Carlo (MCMC) methods produce posterior distributions subject to fewer assumptions, such as normality, than deterministic Bayesian methods. However, MCMC is computationally a very expensive process that requires repeated solution of the wave equation for different velocity samples. Ultimately, a large proportion of these samples (often 40%–90%) is rejected. We have evaluated a two-stage MCMC algorithm that uses a coarse-grid filter to quickly reject unacceptable velocity proposals, thereby reducing the computational expense of solving the velocity inversion problem and quantifying uncertainty. Our filter stage uses operator upscaling, which provides near-perfect speedup in parallel with essentially no communication between processes and produces data that are highly correlated with those obtained from the full fine-grid solution. Four numerical experiments demonstrate the efficiency and accuracy of the method. The two-stage MCMC algorithm produce the same results (i.e., posterior distributions and uncertainty information, such as medians and highest posterior density intervals) as the Metropolis-Hastings MCMC. Thus, no information needed for uncertainty quantification is compromised when replacing the one-stage MCMC with the more computationally efficient two-stage MCMC. In four representative experiments, the two-stage method reduces the time spent on rejected models by one-third to one-half, which is important because most of models tried during the course of the MCMC algorithm are rejected. Furthermore, the two-stage MCMC algorithm substantially reduced the overall time-per-trial by as much as 40%, while increasing the acceptance rate from 9% to 90%.

INTRODUCTION

Full-waveform inversion (FWI) is useful in determining events in the earth's subsurface. However, the seismic inverse problem is ill-posed with uncertainty arising from the data and the model. Uncertainty quantification helps us to elucidate the sources of uncertainty and to quantify their effect on the resulting models (Smith, 2013). One of our primary goals is to produce interval estimates for our models that provide a measure of the reliability of the solution.

FWI, a class of techniques based on using the entire waveform as the data for inversion, has been extensively used in seismic modeling. For example, Minkoff and Symes (1997) use FWI to invert for velocities, reflectivities, and the seismic source. Virieux and Operto (2009) provide a thorough review of FWI for the acoustic and elastic cases. Mulder and Plessix (2008) and Symes (2008) discuss the shortcomings of the traditional least-squares techniques for FWI (including the presence of many local minima in the misfit function) and ideas to overcome these shortcomings.

In this work, we apply Bayesian techniques to FWI. A Bayesian, probabilistic approach allows us to quantify and reduce uncertainty. Data and model uncertainties, incorporated in the likelihood function and prior distribution, inform the uncertainty in the posterior distribution (Mosegaard and Tarantola, 2002; Tarantola, 2005). However, Bayesian methods require the use of a prior distribution that may not be obvious or easy to construct (Scales and Snieder, 1997). Scales and Tenorio (2001) show that the commonly used uniform distribution is, in fact, informative and that in higher dimensions incorporating a hard constraint (e.g., the norm of the

Manuscript received by the Editor 28 December 2018; revised manuscript received 24 May 2019; published ahead of production 23 August 2019; published online 1 November 2019.
[1]The University of Texas at Dallas, Department of Mathematical Sciences, 800 W Campbell Rd., Richardson, Texas 75080, USA. E-mail: georgia.stuart@utdallas.edu; sminkoff@utdallas.edu; luisfelipe.pereira@utdallas.edu.

model is less than 1) into the construction of the prior can lead to unintended consequences (e.g., the only likely models have norm of 1). Deterministic algorithms to construct posterior distributions for probabilistic inverse problems are relatively inexpensive computationally but require strict assumptions about the shape of the posterior distributions (Mosegaard and Tarantola, 2002). Gouveia and Scales (1998) apply deterministic Bayesian techniques to the seismic inversion problem, but they assume the posterior distribution is Gaussian to perform uncertainty quantification. This assumption on the shape of the posterior fails to accurately describe the uncertainty if the posterior is, for example, skewed or multimodal.

To avoid assumptions on the posterior distribution, Markov chain Monte Carlo (MCMC) techniques are often used (Robert and Casella, 1999; Gamerman and Lopes, 2006). Koren et al. (1991) apply MCMC techniques to the velocity inversion problem using traveltime data, rather than FWI. Sambridge and Mosegaard (2002) provide an overview of Monte Carlo and MCMC techniques in geophysical inverse problems. MCMC provides better information for uncertainty quantification by not assuming the shape of the posterior distribution, but the need to solve the forward problem many times and the large proportion of rejected models make the technique very computationally expensive. In particular, random-walk MCMC algorithms that rely on partial differential equations or have many parameters often have exceptionally low acceptance rates (Ginting et al., 2011). For example, Efendiev et al. (2006) find acceptance rates as low as 0.001 when inverting for a permeability field characterized by 20 unknowns; thus, 99.9% of all tested models are rejected and not used to characterize the posterior distribution.

Frangos et al. (2011) classify methods for reducing computational cost in the MCMC process into three categories:

1) reducing the computational cost of the forward simulation necessary for a posterior evaluation
2) reducing the number of parameters that define models
3) more efficient sampling of the posterior distribution, thereby requiring fewer forward model simulations to estimate the posterior distribution parameters.

The geophysics literature on MCMC often incorporates one or more of these strategies to reduce computational cost. Ray et al. (2016) use wavenumber integration (strategy 1) to produce acoustic wave data in the frequency domain as the forward simulation for a transdimensional algorithm in which the number of unknowns that define the model is itself an unknown. Further, they reduce the number of parameters that define the models (strategy 2) by assuming that the velocity field is a laterally homogeneous 3D field with approximately 38 layers described by 200 unknowns. Sen and Biswas (2017) apply a reversible jump Hamiltonian Monte Carlo, which uses Hamiltonian dynamics and gradient information to produce samples that are less correlated than in random-walk MCMC for seismic inversion (strategy 3). They use strategies 1 and 2 for the forward problem by using the convolutional model and a flat-layered medium to produce prestack and poststack data. In a subsequent paper, Biswas and Sen (2017) extend their reversible jump Hamiltonian Monte Carlo technique to a 2D velocity field parameterized with Voronoi cells, where the nuclei of the cells are the velocity field unknowns. Ely et al. (2018) simulate recorded wavefields using the fast field expansion method for 2D velocity inversion (strategy 1), and following Datta and Sen (2016), they represent their velocity fields with a limited number of interfaces described by two to seven unknowns with a gradient between the layers (strategy 2). Hong and Sen (2009) develop a multiscale genetic algorithm MCMC method for seismic waveform inversion, which uses multiple Markov chains at various coarse scales. The fast-converging, but less accurate, coarse chains inform intelligent proposals for the slower converging and more expensive fine-grid chains (strategy 3).

Another way to reduce the number of fine-grid simulations required is to use a multistage algorithm that uses a filter to quickly reject unacceptable model proposals. Christen and Fox (2005) lay the statistical groundwork for the multistage MCMC method. Efendiev et al. (2005, 2006) develop a two-stage MCMC framework for flow in porous media problems (see also Ginting et al., 2011, 2015). Laloy et al. (2013) apply the two-stage MCMC technique combined with DiffeRential Evolution Adaptive Metropolis, DREAM(ZS) to groundwater modeling. Kalligiannaki et al. (2012) apply a form of multistage MCMC to stochastic lattice problems for use in simulating physical systems in micromagnetics. However, Akbarabadi et al. (2015) find that the two-stage MCMC algorithm may not be effective when the correlation between the residuals using the filter and the full fine grid is weak. In Stuart et al. (2016), the precursor to this paper, we apply two-stage MCMC to the velocity inversion problem. In that work, we draw inspiration from the reservoir simulation community and reduce the number of unknowns that describe a velocity model with a truncated Karhunen-Loéve expansion (KLE) (see, for example, Efendiev et al., 2005; Ginting et al., 2011, 2015; Akbarabadi et al., 2015). Unfortunately, the Gaussian KLE perturbations provide "blobby" inclusions that are not particularly realistic on the seismic scale.

In this work, we combine all three strategies to produce an efficient and accurate method for quantifying uncertainty in velocity field inversion. To reduce the number of full fine-grid simulations required, we apply the two-stage MCMC algorithm to seismic inversion. We reduce the cost of a forward simulation by using operator upscaling applied to the acoustic wave equation (Vdovina et al., 2005) as the inexpensive coarse-grid MCMC filter. Operator upscaling is a good choice for the filter because it is fast, displaying near-perfect speedup and minimal communication between processes. Other available filters for the acoustic wave equation include a multiscale finite-element method that incorporates fine-scale information by precomputing basis functions (Gibson et al., 2014) and a homogenization technique that avoids the usual homogenization assumption of two scales via a change of coordinates (Owhadi and Zhang, 2008). For operator upscaling, we show that data obtained from the coarse-grid solution are well-correlated with the data obtained from the full fine-grid solution (see Stuart et al., 2016). Finally, we reduce the number of parameters that define our model by constructing 2D velocity fields from laterally varying interfaces with a total of 17–40 unknowns. We choose this number of unknowns to see convergence in 50,000 models or fewer per chain for six chains (300,000 models in total) after burn-in (Ely et al., 2018).

In the remainder of this paper, we begin by discussing the Bayesian inversion framework. We then present the algorithms for a traditional Metropolis-Hastings MCMC algorithm and the two-stage MCMC algorithm. We also briefly describe the operator upscaling algorithm for the constant-density acoustic wave equation. We demonstrate that the upscaled solver is an accurate surrogate for the full fine-grid solution by examining the correlation between residuals using the fine-grid and upscaled solver. In this paper, new velocity

proposals are generated using a random walk sampler. Unfortunately, random walk samplers produce velocity models that are highly correlated, thus limiting our ability to efficiently investigate the posterior space (Neal, 2011). The Hamiltonian Monte Carlo method (HMC) uses gradient information to propose states that are less correlated with the current state but have a high probability of acceptance (Neal, 2011). Sen and Biswas (2017) and Biswas and Sen (2017) demonstrate that HMC is effective for 1D and 2D seismic inversion. However, generating and testing a sample is significantly more expensive in HMC due to the computation of the gradient. To run problems with more unknowns, we will need to combine our two-stage MCMC with a different sampling strategy such as HMC. However, nothing about our method precludes trying other methods of sampling.

We compare results from a traditional MCMC simulation with the two-stage MCMC simulation and show that the two-stage MCMC, with our velocity field parameterization, converges using six chains with fewer than 50,000 samples per chain. We demonstrate the use of two-stage MCMC on models in which we vary the interface positions and velocity values of the materials. These numerical experiments demonstrate the efficacy of our two-stage MCMC algorithm for constructing the posterior distributions of our model parameters. In fact, the two-stage MCMC algorithm gives inversion results (posterior distributions and interval estimates of the unknowns), which are nearly identical to the one-stage MCMC, indicating that the two-stage MCMC can be used in place of the traditional Metropolis-Hastings one-stage MCMC without compromising the information needed for uncertainty quantification. Because most of the proposed models in MCMC will be rejected, reducing the computational cost per rejected model will greatly speed up the algorithm. We demonstrate a reduction in time per rejected model of up to 50%. Furthermore, the two-stage MCMC exhibits a significant increase in the acceptance ratio of samples over the one-stage MCMC algorithm (from 9% to approximately 90% for one experiment) with the required time per sample tried reduced by approximately 40% for the same experiment.

BAYESIAN INVERSION

It is well-known that the seismic inverse problem for velocity is ill-posed with different Earth models fitting the same receiver data equally well (Symes and Carazzone, 1991; Symes, 2008). Therefore, we consider the inverse problem from the point of view of uncertainty quantification. In uncertainty quantification, Bayes' rule provides a mechanism for combining independent prior information about unknowns with information gained from receiver data and direct, predictive numerical simulations.

Bayes' Rule

Bayes' rule states

$$P(\theta|d) \propto P(d|\theta)P(\theta), \tag{1}$$

where $P(\theta|d)$ is the posterior distribution of the model θ given the data d, $P(d|\theta)$ is the likelihood function, and $P(\theta)$ is the prior distribution of the model (Gelman et al., 2013). Essentially, the posterior distribution is proportional to the product of the likelihood function, a measure of how well data predicted by our current model fit the observed data, and the prior distribution, what we believe about the model before incorporating the data.

MARKOV CHAIN MONTE CARLO FOR BAYESIAN INVERSION

In inverse problems in which the forward problem is nonlinear, as is the case in the velocity inversion problem, it is necessary to use Monte Carlo methods to investigate the often-complicated posterior distribution (Mosegaard and Tarantola, 1995). MCMC is a refinement of the general Monte Carlo method in which a Markov chain is constructed with the posterior distribution as the limiting distribution.

The Metropolis-Hastings MCMC algorithm (Hastings, 1970)

1) Generate a new proposal for the parameter θ from a transitional probability distribution $q(\theta|\theta_n)$, where θ_n is the previous step in the Markov chain. Here, we take the transitional probability distribution — the probability of selecting a next step given a previous step — to be a random walk sampler through θ space (Gamerman and Lopes, 2006), which will be discussed in the "Numerical experiments" section.
2) Solve the forward problem and compute the ratio

$$r(\theta_n, \theta) = \frac{q(\theta_n|\theta)P(\theta|d)}{q(\theta|\theta_n)P(\theta_n|d)}, \tag{2}$$

where $P(\theta|d)$ and $P(\theta_n|d)$ are the posterior values for θ and θ_n, respectively. In this work, we take q to be a symmetric random walk sampler; thus, equation 2 simplifies to

$$r(\theta_n, \theta) = \frac{P(\theta|d)}{P(\theta_n|d)}. \tag{3}$$

3) Update the next step in the chain (θ_{n+1}) by the rule

$$\theta_{n+1} = \begin{cases} \theta & \text{with probability } \rho(\theta_n, \theta) \\ \theta_n & \text{with probability } 1 - \rho(\theta_n, \theta) \end{cases}, \tag{4}$$

where

$$\rho(\theta_n, \theta) = \min\{1, r(\theta_n, \theta)\}. \tag{5}$$

The two-stage Markov Chain Monte Carlo algorithm

Unfortunately, the likelihood calculation in Metropolis-Hastings MCMC methods requires solving the full forward problem for every new proposed velocity model. Often, 40%–90% of these models are rejected in an MCMC algorithm with efficient chain mixing (Rosenthal, 2011). These rejected proposals are not used to characterize the posterior distribution. Most of the computation time is, therefore, spent on velocity models that will ultimately be rejected (Akbarabadi et al., 2015). One method to reduce this expense is the *multistage MCMC method* (Christen and Fox, 2005; Efendiev et al., 2006). In the multistage, or preconditioned, MCMC, proposed samples are first evaluated using an inexpensive filter, thus eliminating unacceptable proposals quickly and efficiently. Only samples that pass the filter test are then evaluated via a full fine-grid forward solve.

The two-stage MCMC algorithm (Christen and Fox, 2005; Efendiev et al., 2006)

1) **Generation step:** Generate a model proposal θ' from the transitional probability distribution $q(\theta'|\theta_n)$.
2) **Filter step:** The proposal θ is

$$\theta = \begin{cases} \theta' & \text{with probability } g(\theta_n, \theta') \\ \theta_n & \text{with probability } 1 - g(\theta_n, \theta') \end{cases}, \tag{6}$$

where

$$g(\theta_n, \theta') = \min\left\{1, \frac{P^\star(\theta'|d)}{P^\star(\theta_n|d)}\right\}. \tag{7}$$

Here, $P^\star(\theta|d)$ is the posterior distribution of the model θ:

$$P^\star(\theta|d) = P^\star(d|\theta)P(\theta), \tag{8}$$

where $P^\star(d|\theta)$ is the likelihood function that measures the difference between observed and coarse grid simulation data. We again eliminate the transitional probability distribution from equation 7 by selecting a symmetric random walk sampler.
3) **Fine-grid step:** The next step of the chain (θ_{n+1}) is updated

$$\theta_{n+1} = \begin{cases} \theta & \text{with probability } \rho(\theta_n, \theta) \\ \theta_n & \text{with probability } 1 - \rho(\theta_n, \theta) \end{cases}, \tag{9}$$

where

$$\rho(\theta_n, \theta) = \min\left\{1, \frac{P(\theta|d)P^\star(\theta_n|d)}{P(\theta_n|d)P^\star(\theta|d)}\right\}. \tag{10}$$

Notice that if $\theta = \theta_n$ in step 2, then we know $\theta_{n+1} = \theta_n$, eliminating the need to evaluate θ on the full fine grid. In other words, a rejection in the filter stage means that we need not continue to step 3 of the algorithm for that model. This step results in a significant reduction in computational cost.

OPERATOR UPSCALING FOR THE ACOUSTIC WAVE EQUATION

The two-stage MCMC algorithm uses a filter to quickly reject unacceptable proposals. This filter must be significantly less computationally expensive than solving the full fine-grid problem, while still generating coarse grid data whose residual is well-correlated with the fine-grid residual (Akbarabadi et al., 2015). In this work, we use operator upscaling (Vdovina et al., 2005, 2009; Korostyshevskaya and Minkoff, 2006; Vdovina and Minkoff, 2008) for the filter stage. This method exhibits near-perfect speedup during parallelization and can detect fine-scale heterogeneities (features below the size of a coarse grid block). Figure 1 shows that the upscaling algorithm produces perfect speedup for a representative test using a grid of size 1344×1344 run for 10,000 time steps. Below, we briefly describe the upscaling method. Further details about convergence, accuracy, and implementation of the method can be found in the references given above.

We model wave propagation using the 2D constant-density acoustic wave equation:

$$\frac{1}{c^2(x,z)}\frac{\partial^2 p}{\partial t^2} - \Delta p = f, \tag{11}$$

where c is the wave speed, f is the source of seismic energy, and p is the acoustic pressure. By introducing acceleration, $\mathbf{v} = -\nabla p$, we can rewrite the acoustic wave equation as a first-order system in space:

$$\begin{aligned} \mathbf{v} &= -\nabla p, \\ \tfrac{1}{c^2}\tfrac{\partial^2 p}{\partial t^2} &= -\nabla \cdot \mathbf{v} + f. \end{aligned} \tag{12}$$

Rather than using the strong form of the wave equation, we consider the mixed finite element variational form. Let $H_0(\mathrm{div}; \Omega)$ be the space of vector functions $\mathbf{V}$, where

$$\{\mathbf{V} \in (L^2(\Omega))^2 : \nabla \cdot \mathbf{v} \in L^2(\Omega), \mathbf{v} \cdot \nu = 0 \quad \text{on } \partial\Omega\}, \tag{13}$$

where ν is the unit normal to the boundary ($\partial\Omega$) and Ω is the interior of the 2D domain (Vdovina et al., 2005). The system in equation 12 can be written in weak form: find $\mathbf{v} \in H_0(\mathrm{div}; \Omega)$ and $p \in L^2(\Omega)$ such that

$$\begin{aligned} &\langle \mathbf{v}, \mathbf{u} \rangle = \langle p, \nabla \cdot \mathbf{u} \rangle, \\ &\left\langle \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}, w \right\rangle = -\langle \nabla \cdot \mathbf{u}, w \rangle + \langle f, w \rangle, \end{aligned} \tag{14}$$

for all $\mathbf{u} \in H_0(\mathrm{div}; \Omega)$ and $w \in L^2(\Omega)$. Here, $\langle \cdot, \cdot \rangle$ is defined by

$$\langle \mathbf{u}, \mathbf{v} \rangle = \int_\Omega \mathbf{u} \cdot \mathbf{v} dx. \tag{15}$$

For the upscaling, we decompose acceleration $\mathbf{v}$ into coarse components $\mathbf{v}^c \in V^C$, which live on the edges of the coarse blocks and fine or subgrid components $\delta\mathbf{v} \in \delta V$, which live on the edges of the fine cells. Thus,

$$\mathbf{v} = \mathbf{v}^c + \delta\mathbf{v}. \tag{16}$$

Whereas the pressure could also be decomposed in this manner (see Arbogast et al., 1998), in this work we assume pressure lives only on the fine grid. Figure 2 shows the locations of the coarse-grid acceleration, fine-grid acceleration, and pressure unknowns.

The upscaling algorithm, therefore, consists of two steps:

1) Solve the subgrid problem for internal unknowns, pressure and fine-grid acceleration, on each coarse cell.
2) Solve for acceleration on the coarse grid using the subgrid solutions found in step 1.

In step 1 above, we solve the subgrid problems on each coarse block to determine $\delta\mathbf{v}$ and p by restricting to test functions $\delta\mathbf{u}$ and w, which live on the fine grid. In other words, we solve the system

$$\begin{aligned} &\langle \mathbf{v}^c + \delta\mathbf{v}, \delta\mathbf{u} \rangle = \langle p, \nabla \cdot \delta\mathbf{u} \rangle, \\ &\left\langle \frac{1}{c^2}\frac{\partial^2 p}{\partial t^2}, w \right\rangle = -\langle \nabla \cdot (\mathbf{v}^c + \delta\mathbf{v}), w \rangle + \langle f, w \rangle. \end{aligned} \tag{17}$$

These subgrid problems completely determine pressure p. In this implementation, we assume piecewise constant basis functions

for pressure and piecewise linear functions for acceleration. Other choices of interpolating polynomials are also possible. In a traditional domain decomposition approach, one needs to pass the pressure from each cell living on the edge of the subdomain to adjacent subdomains. However, we decouple the subgrid problems by applying homogeneous Neumann boundary conditions on the edge of each coarse block E_c (our primary simplifying assumption for the upscaling algorithm). In other words, we assume that

$$\delta\mathbf{u} \cdot \nu = 0 \quad \text{on } \partial E_c. \tag{18}$$

This condition allows us to solve for the subgrid unknowns on each coarse block in parallel, achieving near-perfect speedup with minimal communication between processes (Vdovina et al., 2005).

Whereas operator upscaling was developed in the context of mixed finite elements (Oden and Reddy, 1976), due to the equivalence between the lowest-order Raviart-Thomas mixed finite elements and the cell-centered finite difference (Russell and Wheeler, 1983), our implementation at step 1 of the algorithm (the expensive step) is in practice an explicit finite difference scheme in space and time. Furthermore, other discretization schemes are also possible.

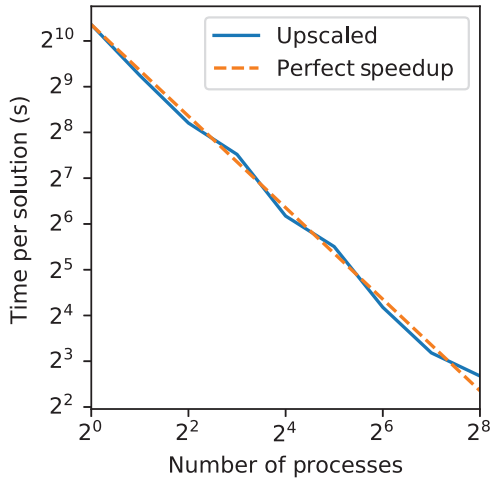


Figure 1. Timing plot for solution of the acoustic wave equation using operator upscaling with one velocity sample run for 10,000 time steps versus number of MPI processes (the solid blue line) and theoretical perfect speedup (the dashed orange line).

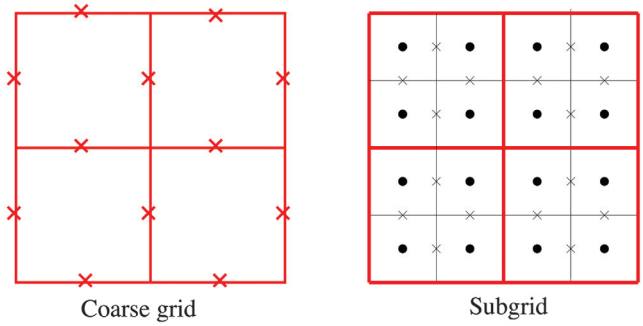


Figure 2. The coarse grid (red) and subgrid (black) unknowns and decomposition for a domain with four coarse blocks. Acceleration, denoted by crosses, lives on the boundaries of the coarse and subgrid cells. Pressure, denoted by dots, lives at the center of each fine cell.

For example, Vdovina et al. (2009) apply operator upscaling to the 3D elastic wave equation using mass lumping and the spectral finite-element method.

In step 2, we solve for the coarse acceleration $\mathbf{v}^c$ using the results from the subgrid problems,

$$\left\langle \mathbf{v}^c + \delta\mathbf{v}, \mathbf{u}^c \right\rangle = \left\langle p, \nabla \cdot \mathbf{u}^c \right\rangle, \quad \text{for all } \mathbf{u}^c \in V^c, \tag{19}$$

where $\mathbf{u}^c$ are the coarse-grid test functions. We note that because the coarse operator incorporates the fine-scale solution, we can capture the fine-scale heterogeneities in the solution (see Vdovina et al. [2005] for several examples in which features much smaller than a coarse block are recovered in the upscaled solution).

UNCERTAINTY QUANTIFICATION

We choose to use the MCMC algorithm for uncertainty quantification. Specifically, we examine the uncertainty of our inverted parameters by constructing kernel density estimates (KDEs) of the posterior distribution (Scott, 1992) and calculating credible intervals using highest posterior density (HPD) intervals (Gelman et al., 2013). KDE produces a probability density function of a random variable from a collection of samples. In this case, the collection of samples corresponds to the entries of our Markov chain. We combine multiple chains to form a sample $(\theta_1, \theta_2, \ldots, \theta_{m\cdot n})$, where m is the number of chains and n is the number of entries per chain. To compute the KDE, we sum over a series of nonnegative kernels constructed about each data point in the sample (Scott, 1992):

$$\hat{f}(\theta) = \frac{1}{n \cdot m} \sum_{i=1}^{n\cdot m} K_h(\theta - \theta_i). \tag{20}$$

For the kernels K_h, we use Gaussians with standard deviation set by the bandwidth h. The bandwidth is computed following the normal reference rule from Scott (1992):

$$h \approx 1.06\hat{\sigma}(n \cdot m)^{-1/5}, \tag{21}$$

where $\hat{\sigma}$ is the standard deviation of the sample.

For interval estimation of our inverted parameters, we construct credible intervals using HPD intervals. Although HPD regions are not necessarily a single contiguous interval, we make the simplifying assumption that our posterior distributions are unimodal to compute contiguous HPD intervals. We find that this is a good assumption because most of the KDEs we observe are unimodal. Although there are infinitely many intervals that cover $(1-\alpha)\times 100\%$ of the distribution, we define the $(1-\alpha)\times 100\%$ HPD interval as the narrowest interval that covers $(1-\alpha)\times 100\%$ of the posterior distribution (Gelman et al., 2013).

CONVERGENCE

Because uncertainty quantification relies on estimates of the posterior distributions by the samples in the chain, to have confidence in the results we require that the estimated probability distributions converge to a stable distribution. We show convergence numerically in two ways: with a visual demonstration showing the stability of the KDE for a parameter at different points in the Markov chain and

by using the multivariate potential scale reduction factor (MPSRF) and max PSRF from Brooks and Gelman (1998). In this section, we will discuss the derivation of the MPSRF, which we then apply to experiment 1 in the "Numerical experiments" section.

The MPSRF approach to demonstrate convergence relies on the pooled covariance matrix $\mathbf{V}$ and the within-chain covariance matrix $\mathbf{W}$, which are calculated from multiple chains (Brooks and Gelman, 1998). The pooled covariance matrix is estimated from

$$\mathbf{V} = \frac{n-1}{n}\mathbf{W} + \left(1+\frac{1}{m}\right)B/n, \tag{22}$$

where B/n is the between-chain covariance matrix, n is the number of samples per chain, and m is the number of chains. The within-chain covariance matrix is

$$\mathbf{W} = \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{i=1}^{n}(\theta_{ji}-\bar{\theta}_j)(\theta_{ji}-\bar{\theta}_j)^T, \tag{23}$$

where θ_{ji} is the ith entry of the jth chain and $\bar{\theta}_j$ is the mean of the jth chain. The between-chain covariance matrix is

$$B/n = \frac{1}{m-1}\sum_{j=1}^{m}(\bar{\theta}_j-\bar{\Theta})(\bar{\theta}_j-\bar{\Theta})^T, \tag{24}$$

where $\bar{\Theta}$ is the mean of all the chains combined.

As the chains evolve, $\mathbf{V}$ and $\mathbf{W}$ should converge. The MPSRF gives a scalar measure of the difference between $\mathbf{V}$ and $\mathbf{W}$. As $\mathbf{V}$ and $\mathbf{W}$ converge, we expect the MPSRF to approach 1. The MPSRF (Brooks and Gelman, 1998) $\hat{R}^p$ is computed by

$$\hat{R}^p = \max_a \frac{a^T\mathbf{V}a}{a^T\mathbf{W}a}, \tag{25}$$

$$= \frac{n-1}{n} + \frac{m+1}{m}\lambda_1, \tag{26}$$

where λ_1 is the largest eigenvalue of $\mathbf{W}^{-1}B/n$. In addition, we look at the maximum of the individual potential scale reduction factor (PSRF). The PSRF is computed in the same manner as the MPSRF, but here we only consider one parameter at a time. Again, we expect each PSRF, and thus the maximum of all PSRFs, to approach 1.

NUMERICAL EXPERIMENTS

We will now describe four synthetic numerical experiments to illustrate the performance, cost savings, and convergence of the two-stage MCMC method. Each experiment is intended to determine characteristics of the velocity field, specifically, either interface positions, velocity values within features, or both. Crosswell tomography can provide well-resolved velocity values, but more uncertainty exists about the topography of material interfaces depending on the acquisition geometry or nature of the interfaces to be imaged (e.g., faults). Thus, in numerical experiment 1, we assume that the velocity values are known and only invert for interface depths that are allowed to vary laterally for each material (i.e., the interfaces are not flat). Numerical experiment 2 is an extension of the first experiment. In this case, we assume that the velocity values and interface depths are unknown which allows us to quantify the uncertainty on the interface depth when the velocity values are also incorrect initially. In contrast, mature fields such as parts of the Midland Basin in West Texas often have dense well coverage with sufficient log data (e.g., gamma ray and resistivity) to allow us to build a robust structural/stratigraphic model of the interfaces. However, many of the wells do not measure acoustic impedance. Hence, for numerical experiment 3, we assume that we know the layer interfaces but must estimate the velocity values. In the fourth experiment, we invert for the interface position and velocity for a flat-layered model. The velocity fields in the third and fourth experiments are constructed from a blocked well log.

For all four experiments, we demonstrate the efficacy of the inversion by comparing the velocity field constructed from the median of the posterior distribution to the initial velocity field and the true velocity field. However, we emphasize that our goal in this work is not the inversion results themselves, but the estimation of uncertainty in our results. We examine the uncertainty in our solution using HPD intervals and KDEs of the posterior distributions. For experiment 1, we show that the two-stage MCMC algorithm and the traditional one-stage MCMC algorithm result in nearly identical posterior distributions, which indicates that the two-stage MCMC process is an appropriate replacement for the more expensive one-stage MCMC process. We examine the correlation between the full fine-grid data and the upscaled data and show an example of the upscaled receiver data compared with receiver data on the full fine grid. Finally, we analyze the convergence of the two-stage MCMC algorithm by showing the stability of the posterior distributions as the chain evolves and by demonstrating the convergence of the MPSRF and max PSRF values to 1. After describing the individual experimental results, we discuss timings for each experiment, namely, the reduction in time needed to reject a model and the time reduction overall per trial.

Construction of the velocity field

The velocity field is determined by the position of material interfaces and the magnitude of the velocity between interfaces. Each interface is parameterized by one or more unknowns. An interface defined by only one unknown is a flat layer. The interfaces are piecewise linear between the unknowns if more than one unknown defines the interface position. Between the interfaces, the velocity is constant within each material (here, one unknown suffices).

In the MCMC process, the velocity field is completely determined by a dimensionless vector θ with each component of the vector corresponding to one unknown (either interface position or velocity value). We transform θ to physical units by adding the initial position or velocity to a rescaled θ using an average step length σ_{step},

$$v = v_0 + \sigma_{\text{step}}^{\text{velocity}}\cdot\theta \quad \text{for velocity, } v, \text{ with initial velocity } v_0 \tag{27}$$

$$d = d_0 + \sigma_{\text{step}}^{\text{interface}}\cdot\theta \quad \text{for interface depth, } d, \text{ with initial depth } d_0. \tag{28}$$

The specific values of $\sigma_{\text{step}}^{\text{velocity}}$, $\sigma_{\text{step}}^{\text{interface}}$, and the initial values will be discussed in the individual experiments.

Each experimental region contains 1344 fine-grid blocks in the x direction and 1344 fine-grid blocks in the z direction with a nearly perfectly matched layer absorbing region (Berenger, 1994) of 160 fine-grid blocks on each edge of the computational domain. Hence, the computational region contains 1024×1024 fine-grid blocks with 1 m grid spacing in x and z. For the upscaled wave solver, each coarse block contains 16 fine-grid blocks in x and 16 fine-grid blocks in z. The computational domain then contains 64×64 coarse-grid blocks in total. We solve the wave equation for 0.5 s total with each time step $\Delta t = 5 \times 10^{-5}$ s. This time step was chosen to satisfy the 2D Courant-Friedrichs-Lewy conditions for velocities up to 10 km/s.

The source and receiver geometry for the four experiments includes 20 equally spaced sources at the top edge of the computational domain ($z = 0$ m), with each source being a 30 Hz Ricker wavelet in time and a point source in space. We note that the Ricker wavelet contains low frequencies that may not be present in real data. We record pressure data along horizontal or vertical receiver lines with receivers spaced 2 m apart. Finally, each experiment was performed using six independent Markov chains containing between 34,000 and 50,000 samples each (204,000 to 300,000 samples total). For numerical experiments 1, 2, and 4, the starting velocity model for each chain is a perturbation of the velocity model produced by the mean of the prior distribution. For numerical experiment 3, the starting velocity model is constant.

The likelihood function, prior distribution, random walk sampler, and acceptance ratios

Likelihood function and prior distribution

We follow Efendiev et al. (2006) (see also Lee, 2005) and assume that the error associated with measurement, modeling, and numerical approximation satisfies a Gaussian distribution. Consequently, the likelihood function is given by

$$P(d|\theta) \propto \exp\left(-\frac{\|u(\theta) - d\|^2}{2\|d\|^2\sigma^2}\right), \tag{29}$$

where $u(\theta)$ is the data simulated assuming the model θ, d is the observed data, $\|\cdot\|$ is the L^2 norm, and σ^2 is the precision parameter that quantifies the noise in the data (Lee, 2005; Akbarabadi et al., 2015). The likelihood function of the coarse-grid solution is constructed similarly:

$$P^\star(d|\theta) \propto \exp\left(-\frac{\|u_C(\theta) - d\|^2}{2\|d\|^2\sigma_C^2}\right), \tag{30}$$

where $u_C(\theta)$ is the data simulated using operator upscaling and σ_C is the coarse-grid precision parameter. For the numerical experiments, we choose the precision parameters σ and σ_C, to allow good chain mixing after burn-in. If the precision parameters are too small, we find that the range of relative residuals that the algorithm will accept is too small and the chains will not mix well.

Choosing a prior distribution for Bayesian inversion is not always straightforward (Scales and Snieder, 1997; Scales and Tenorio, 2001). A prior is meant to represent what we know about the parameters of interest before the data are taken into consideration. It can be used to represent what we think we know from experience about the parameters, or it can be estimated by observations independent from the data (Scales and Tenorio, 2001). For three of our experiments, experiments 1, 2, and 4, the prior distributions are chosen to be Gaussian. For experiment 3, the prior distribution is uniform; i.e., we assume that each velocity model is equally likely in the prior. An appropriate prior distribution for layer interfaces can be estimated from the geologic depositional environment of the area. Another appropriate choice for the prior distribution is the log-normal distribution. However, Gaussian priors are suitable for our numerical experiments, and we find that they do not restrict the posterior to be Gaussian. Here, we will discuss the construction of the Gaussian priors. Because we assume that the unknowns that characterize our velocity fields are independent in the prior, the prior distribution is a product of independent Gaussians (one for each unknown in our problem):

$$\pi(\theta) = \prod_{i=1}^{n_d} \exp\left[-\frac{(\mu_i^{\text{int}} - \theta_i^{\text{int}}\sigma_{\text{step}}^{\text{int}})^2}{2(\sigma_{\text{prior}}^{\text{int}})^2}\right] \times \prod_{i=1}^{n_v} \exp\left[-\frac{(\mu_i^{\text{vel}} - \theta_i^{\text{vel}}\sigma_{\text{step}}^{\text{vel}})^2}{2(\sigma_{\text{prior}}^{\text{vel}})^2}\right], \tag{31}$$

where σ_{step} is as discussed previously in equations 27 and 28, σ_{prior} is the desired standard deviation of the prior, n_d and n_v are the number of unknowns that describe the interface position and velocities, respectively, and μ_i are the means of each independent Gaussian. The superscript labels "vel" and "int" correspond to whether the variable describes a velocity unknown or an interface unknown. The means of the Gaussians μ_i can be estimated from well-log data (see Stuart et al., 2016). Intuitively, each independent Gaussian in the product represents a distribution around a single unknown that indicates our uncertainty about the value of the unknown before we begin the MCMC algorithm. We choose $\sigma_{\text{prior}}^{\text{vel}} = 250$ m/s for velocity and $\sigma_{\text{prior}}^{\text{int}} = 50$ m for interface depth.

Random walk sampler

The progression of the θ vector gives the Markov chain. We construct the next step in the chain θ^{n+1}, by adding a θ_{step} vector to the previous θ vector:

$$\theta^{n+1} = \theta^n + \theta_{\text{step}}^{n+1}. \tag{32}$$

The entries of the vector $\theta_{\text{step}}^{n+1}$ are assumed independent and identically distributed $\mathcal{N}(0,1)$ and are influenced by the previous step direction:

$$\theta_{\text{step}}^{n+1} = \beta\theta_{\text{step}}^n + \sqrt{1-\beta^2}\hat{\theta}, \tag{33}$$

where $\beta \in [0,1]$ is the tuning parameter that determines how much influence the previous step direction has on the new step direction and $\hat{\theta} \sim \mathcal{N}(0,1)$ is a random vector. For all of our experiments, we find that $\beta = 0.3$ gives the best acceptance rate while still enabling good exploration of the posterior distribution.

Numerical experiment 1: Determination of the variable interface depth

In the first numerical experiment, we invert for the laterally varying depth of interfaces assuming known material velocities. We will

compare the two-stage MCMC results to the results from the traditional one-stage MCMC and demonstrate experimentally that the posterior distribution converges for the two-stage MCMC algorithm.

Our velocity field consists of four interfaces with nine equally spaced unknowns per interface for a total of 36 unknowns (see Figure 3, true velocity). The unknowns are allowed to move independently in the z-direction. The initial velocity field, shown in Figure 3 (left), is flat-layered. The velocities between the interfaces range from 2500 to 7000 m/s. We record acoustic pressure on two vertical arrays and one horizontal array of 512 receivers each, placed at the left, right, and top of the computational domain (the green triangles in Figure 3, true velocity). Because the two-stage MCMC algorithm uses residuals from the upscaled forward problem to reject unacceptable samples, we must show that these filter residuals are correlated with the fine-grid residuals. In other words, if a certain velocity field results in a larger fine-grid residual than another velocity field, this relationship should also be reflected in the filter residuals. We demonstrate the correlation between the filter residuals and the fine-grid residuals by first selecting a reference field (the true field in numerical experiment 1) and generating 125 perturbations of that reference field. Then, we compute the filter data and the fine-grid data and plot the relative residuals, $\|d_{\text{ref}} - d_{\text{fine}}\|/\|d_{\text{ref}}\|$ versus $\|d_{\text{ref}} - d_{\text{filter}}\|/\|d_{\text{ref}}\|$, for each perturbed field. Figure 4 demonstrates that the relative residuals for the fine-grid data and the filter data are well-correlated for this acquisition geometry and velocity field. In Figure 5, we show one example of the agreement between receiver data generated by the full fine-grid wave solver and the upscaled wave solver for the line of vertical receivers on the left side of the domain at $x = 0$ m.

In this experiment, we set the fine-grid and coarse-grid likelihood precision parameters to $\sigma = \sigma_C = 0.05$ and the average step length to $\sigma_{\text{step}}^{\text{interface}} = 6$ m (see equation 28). For each of the six independent Markov chains in this study, we test 34,000 samples for a total of 204,000 samples. The relative residual plots for each chain, shown in Figure 6, indicate that the relative residuals stabilize at approximately 0.3, and we remove the first 3000 samples (the burn-in period) from our posterior analysis to reduce the influence of the initial velocity field (Gelman et al., 2013).

In Figure 3, we compare the velocity field generated by the median of the two-stage MCMC posterior distribution to the median field from the one-stage MCMC, the true velocity, and the initial

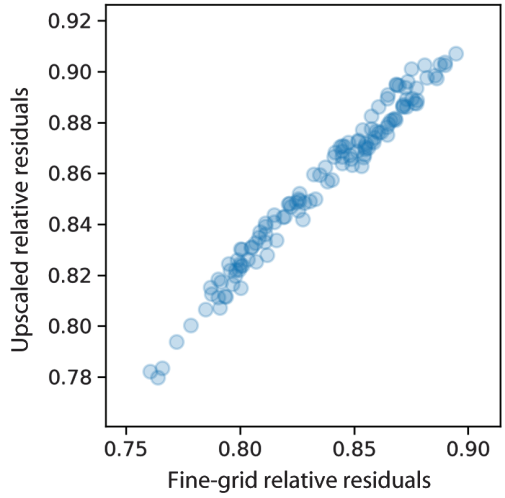


Figure 4. Comparison of fine-grid relative residuals, $\|d_{\text{ref}} - d_{\text{fine}}\|/\|d_{\text{ref}}\|$, versus the upscaled relative residuals, $\|d_{\text{ref}} - d_{\text{filter}}\|/\|d_{\text{ref}}\|$, for 125 proposed velocity models from numerical experiment 1.

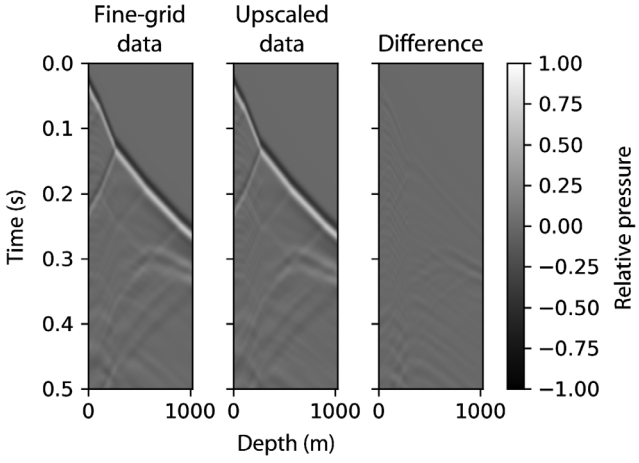


Figure 5. Receiver data on the full fine grid (left), the upscaled receiver data (center), and the difference (right) using the true velocity field from numerical experiment 1 (see Figure 3).

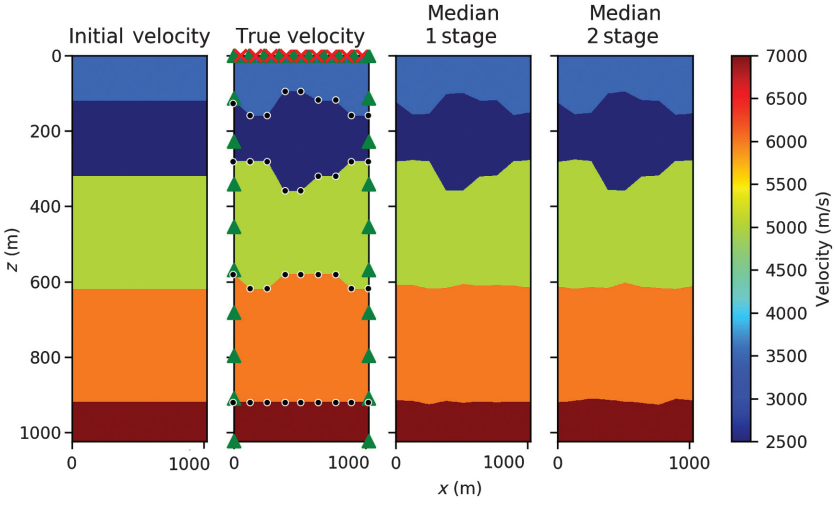


Figure 3. A comparison of the velocity fields constructed from the median of the posterior distributions for the one-stage and two-stage MCMC for numerical experiment 1. Also shown are the initial and true velocity fields. In the true velocity plot, the black dots indicate the locations of the interface unknowns. The green triangles delineate the three arrays of receivers (two vertical and one horizontal array). Each array contains 512 receivers. The red Xs indicate the location of the line of 20 sources at the top of the domain.

velocity. We see that the median of the two-stage MCMC posterior distribution produces a velocity field that is very close to the true one and to the field generated from the one-stage MCMC posterior distribution. Furthermore, Figure 7 shows that the data simulated on the full fine grid using the median velocity field from the two-stage MCMC matches the observed data very well.

One of our main interests in multistage MCMC is analyzing its usefulness for uncertainty quantification. In Figure 8, we plot the posterior distributions of the depth deviation from the initial position generated by KDE for three unknowns at $x = 128$, 512, and 896 m in each layer. We choose not to show all 36 posteriors as these 12 are representative. The blue and orange shaded distributions correspond to the one-stage MCMC and the two-stage MCMC posterior distributions, respectively. Because the two-stage MCMC successfully constructs posterior distributions nearly identical to the one-stage MCMC, the two curves often overlap. The green curve represents the prior distribution, and the red line marks the true unknown deviation from the initial position. In rows 1 and 2 of Figure 8, which correspond to the top two interfaces, we see the tendency of the posterior distributions to move toward the true unknown deviation, with a narrower distribution (less variance) compared to the prior distribution. This indicates we have improved our uncertainty about the positions of the unknowns. For the bottom two interfaces, however, the posterior distribution remains broad, indicating that the uncertainty about these two deeper interfaces is large. For five of the six posterior distributions shown in the bottom two interfaces, the prior distribution begins centered at the correct location and remains centered at the correct location after MCMC. For all interfaces, the posterior distributions for the one-stage MCMC and the two-stage MCMC are very similar, which indicates that the two-stage MCMC produces the same uncertainty information as the one-stage MCMC.

From the posterior distributions, we construct 90% HPD intervals. Figure 9 shows the 90% HPD interval and median position for the one-stage MCMC and the two-stage MCMC (the blue shaded region and solid line, respectively), along with the initial position (dotted green line) and true interface position (dashed red line). For every layer, we capture the true interface position within our 90% HPD intervals. For the top two layers, the 90% HPD interval is quite narrow, which corresponds to the much narrower posterior distributions we see in Figure 8.

We will analyze convergence in two ways for the two-stage MCMC only. First, in Figure 10, we show the posterior distribution for a single unknown at different points in the chain, from 25,000 samples per chain to 50,000 samples per chain. These plots show only small differences in the posterior distribution as the chain evolves, indicating that we have converged to a stable posterior distribution. The posterior distributions for all the other unknowns demonstrate similar behavior. Gelman and Shirley (2011) note that it is practical to consider that the parameters have converged when the PSRF values fall below 1.1. In fact, the parameters considered individually can converge and be well-estimated before the MPSRF value has converged. In Figure 11, we observe that the maximum PSRF value falls below the 1.1 threshold at approximately 50,000 iterations. Furthermore, the MPSRF value $\hat{R}^p$ appears to be trending to 1, which indicates that we do have convergence of the chains.

Numerical experiment 2: Determination of variable interface depths and velocity values

Numerical experiment 2 is the same as numerical experiment 1 except that we now demonstrate the impact on the inversion if the

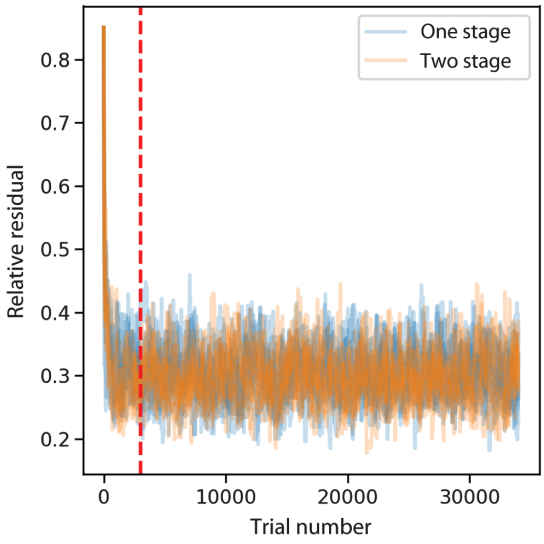


Figure 6. The relative residual plots for each of the six chains of the two-stage MCMC (shown in orange) and the one-stage MCMC (shown in blue) for numerical experiment 1. The red dashed line marks the end of the burn-in period.

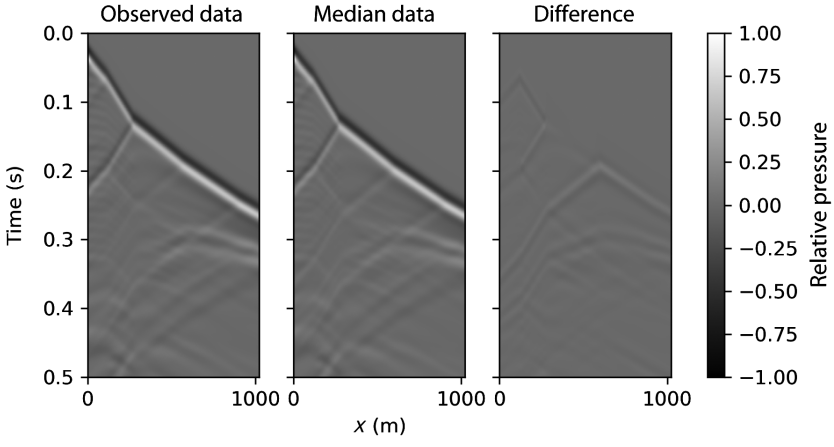


Figure 7. The observed receiver data (left), the simulated receiver data on the full fine grid using the median velocity field from the two-stage MCMC (center), and the difference (right) from the vertical array on the left side of the field at $x = 0$ m from experiment 1.

velocity values are also unknown. Figure 12 shows the initial model with flat interfaces and perturbed velocity (left), the true model (center), and the median posterior model for depth and velocity. We recover the interface positions and the velocity values to within reasonable accuracy. Figure 13 shows the HPD intervals for velocity (left) and interface position (right). The true values fall within the 90% HPD values for velocity and interface position. However, the uncertainty in the velocity impacts the confidence we have in our interface positions. We see that the HPD intervals for interface depth (see Figure 13b) are wider than the equivalent HPD intervals in Figure 9.

Numerical experiment 3: Variable velocity within an angular unconformity model

In our third numerical experiment, we invert for the velocity within interfaces while assuming that we know the interface boundaries. As shown in Figure 14, true velocity, the model contains an angular unconformity and 24 different material regions. To create the model, we used a blocked well log and applied a logistic transformation to the lower 19 interfaces. The blocked well log (Figure 14a) was provided by Pioneer Natural Resources and acquired from a well in the Midland Basin in West Texas with a logging interval that covers portions of the Spraberry, Wolfcamp, and

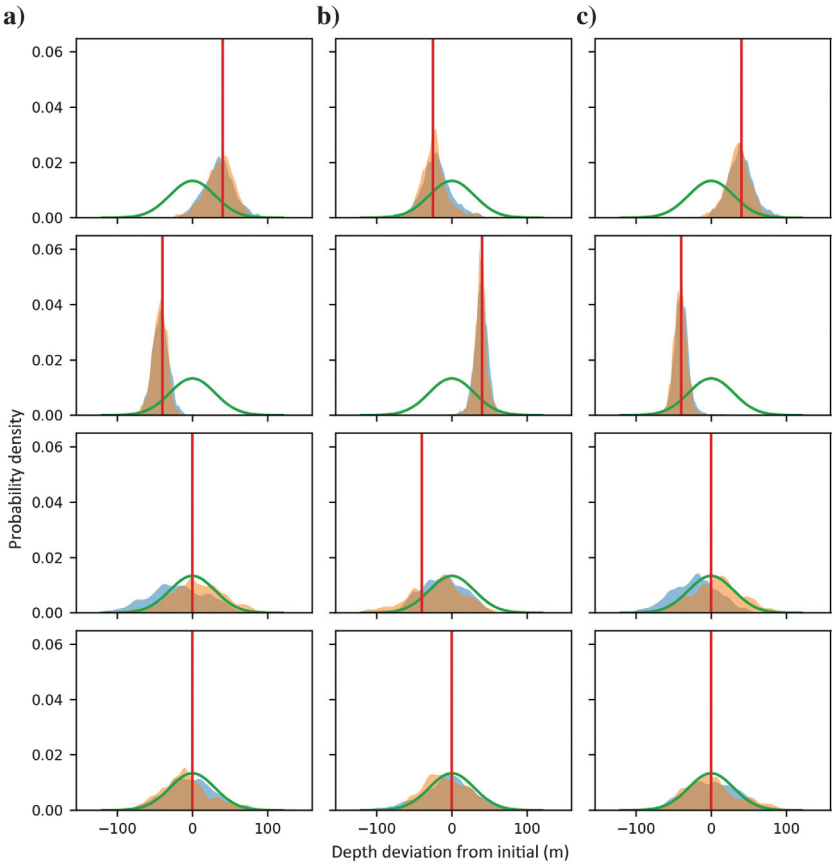


Figure 8. Kernel density estimation of the posterior distributions for 12 unknowns (three in each of the four interfaces) for numerical experiment 1. The blue KDE is the one-stage MCMC posterior distribution. The orange KDE is the two-stage MCMC posterior distribution. The green curve represents the prior distribution for that unknown, and the red line represents the true position. The columns correspond to an x position of 128 m (a), 512 m (b), and 896 m (c). Each row corresponds to one of the four interfaces.

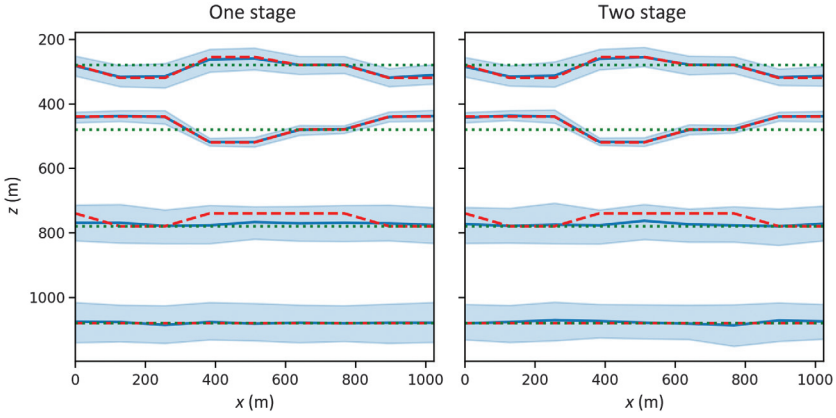


Figure 9. A comparison of the one-stage MCMC and two-stage MCMC HPD intervals for numerical experiment 1. The solid blue lines represent the median depth position. The blue-shaded region represents the HPD interval. The dotted green lines represent the initial depth position. The dashed red lines indicate the true interface position.

Strawn Formations. The velocity values in the true field range from 3559 to 5909 m/s. The model contains materials that vary down to a width of 1 m. We collect pressure data at one vertical array of 512 receivers, located on the left edge of the domain (at $x = 0$ m; see Figure 14, true velocity). We assume a constant velocity of 4447 m/s for our initial velocity model for inversion. In addition, we assume the interface positions are known, as illustrated by the black dotted lines in Figure 14, initial velocity. We have omitted the plot of the correlation between the fine-grid relative residuals and the filter relative residuals because it is very similar to Figure 4 from numerical experiment 1.

The precision parameters for this experiment were chosen to be $\sigma = 0.04$ and $\sigma_C = 0.04$ for fine-grid and coarse-grid precision, respectively. The average step length $\sigma_{\text{step}}^{\text{velocity}}$ was chosen to be 50 m/s. The prior distribution for this experiment is uniform. Each of the six chains contains 35,000 samples for a total of 210,000 samples tried. Again, we remove the first 3000 samples from each chain to reduce the impact of the initial velocity field after the relative residuals stabilize at 0.2 (the burn-in; figure is omitted because it is similar to Figure 6 in numerical experiment 1).

The median velocity of the posterior distribution, as shown in Figure 14, recovers the true velocity field for this experiment. In Figure 15, we illustrate the uncertainty of the velocities corresponding to the well location by taking a 1D vertical slice of the velocity field on the left edge of the domain ($x = 0$ m). The green-shaded region represents the 90% HPD interval for the velocity values. For all velocities, we capture the true velocity value, the orange line in Figure 15, within the 90% HPD interval. However, for the thin regions at approximately 400, 550, and 810 m depth, the 90% HPD intervals are quite large. These thin regions are between 1 and 11 m wide, which are hard to capture given the resolution of our 30 Hz source (Yilmaz, 2001). As such, the uncertainty about the velocity values of these thin regions is high. In Figure 16, we show three KDE posterior distributions, which correspond to the velocity regions marked 1 through 3 in Figure 15. Velocity regions 1 and 3 demonstrate relatively narrow posterior distributions around the true velocity value (marked with a red line). Velocity region 2, however, has an extremely broad KDE posterior distribution. This region is only 11 m wide, demonstrating again that we have large uncertainty for the thin regions below resolution.

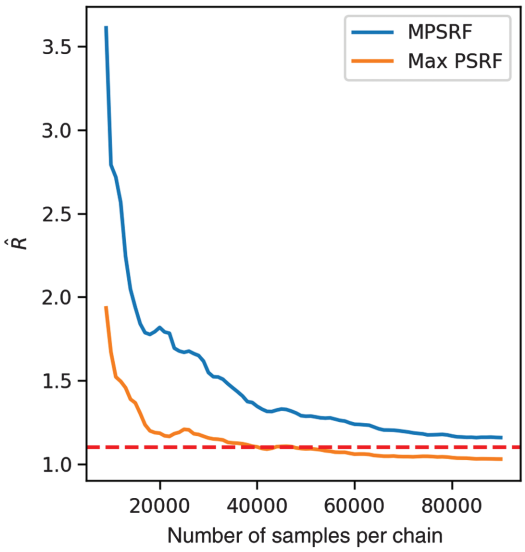


Figure 11. MPSRF (blue) and maximum PSRF (orange) curves for numerical experiment 1. The threshold indicating convergence (1.1) is denoted by the dashed red line.

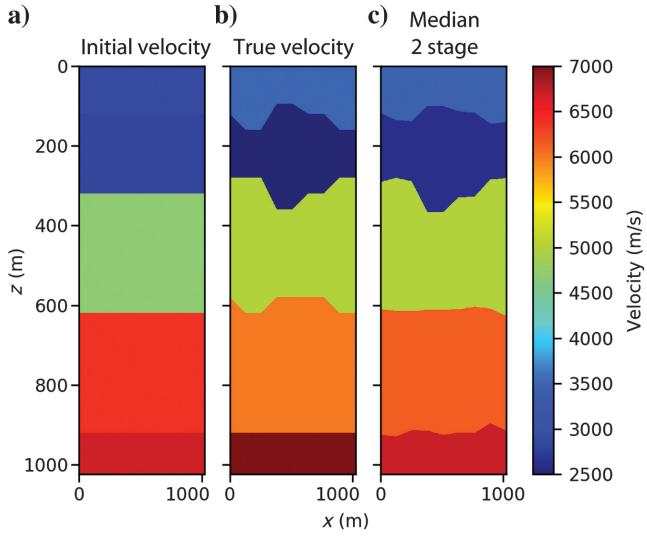


Figure 12. The initial model (a), true model (b), and median model (c) for numerical experiment 2. In this experiment, the velocity values and interfaces are determined via inversion.

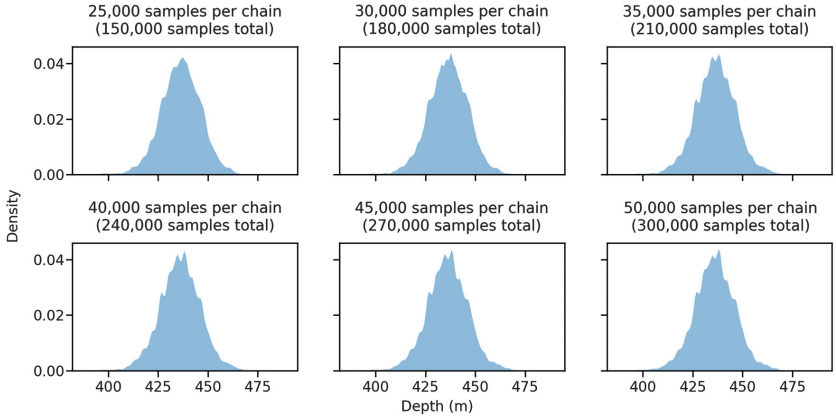


Figure 10. The posterior distributions for one unknown from numerical experiment 1. Each distribution is constructed at a different point in the chains, from 25,000 samples per chain to 50,000 samples per chain. The total number of samples from the combined chains used in the KDE is given in parentheses.

Numerical experiment 4: A flat-layered velocity field with unknown interface depth and velocity

In our fourth and final experiment, we construct a flat-layered velocity field and invert for the depth of the layer interfaces and the velocity values within each layer (similar to experiment 2). This model has a total of 17 dimensions: eight interface depths and nine velocities. The velocity field was created by further blocking the well log (see Figure 14a) into nine layers and extending the log horizontally to create a 2D field, as shown in Figure 17b. As in experiment 3, the velocity values range from 3559 to 5909 m/s. The initial velocity field, shown in Figure 17a, was created by perturbing interface depths by 100 m and velocity values by 500 m/s. The 20 sources are equally spaced along the top of the computational domain, and the pressure wave is recorded at 512 receivers placed 2 m apart on the left side of the computational domain, as indicated in Figure 17b. We again omit the correlation picture because the relationship between the fine-grid relative residuals and

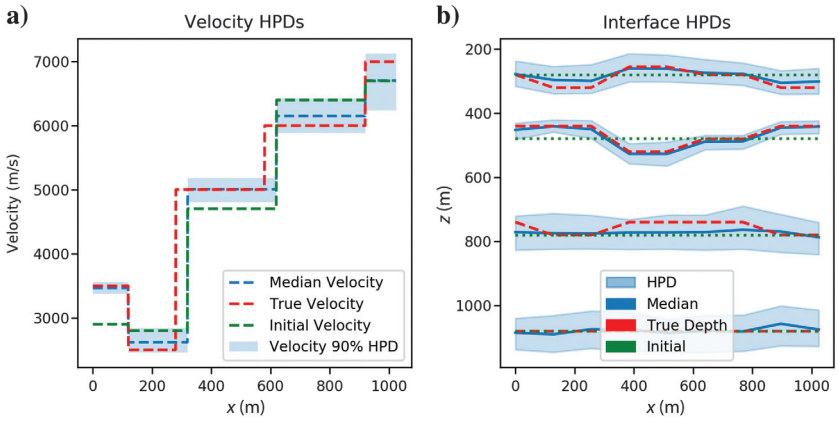


Figure 13. (a) A 1D slice of the median velocity field at $x = 0$ m (the dashed blue lines), true velocity field (the dashed red lines), and initial velocity field (the dashed green lines) with 90% HPD intervals on the velocity values only (blue-shaded region) for numerical experiment 2. (b) The initial interfaces (green dotted lines), true interfaces (the dashed red lines), median posterior interfaces (blue lines), and 90% HPD intervals for the interfaces (blue-shaded regions).

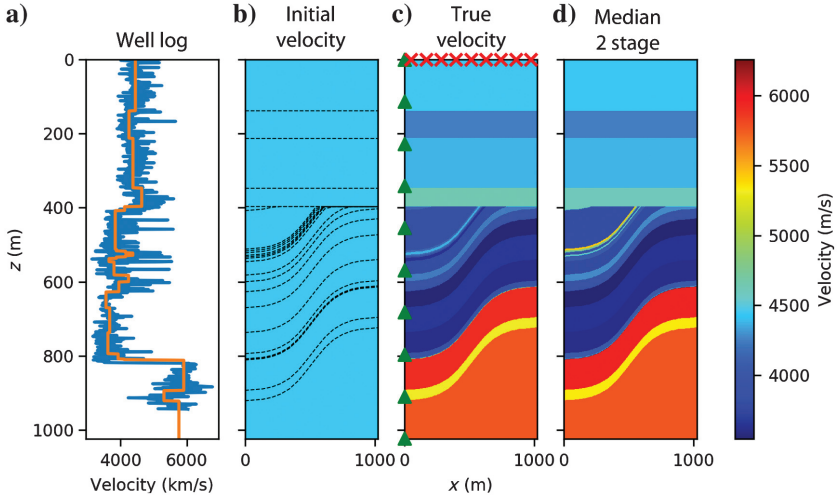


Figure 14. The velocity field constructed from the median of the two-stage posterior distribution for numerical experiment 3 (d), compared with the true velocity field (c), the initial velocity field (b) with known interfaces marked with black dashed lines, and the median blocked well log (a), courtesy of Pioneer Natural Resources. On the true velocity plot, the green triangles show the vertical array of 512 receivers. We mark the line of 20 sources at the top with red Xs.

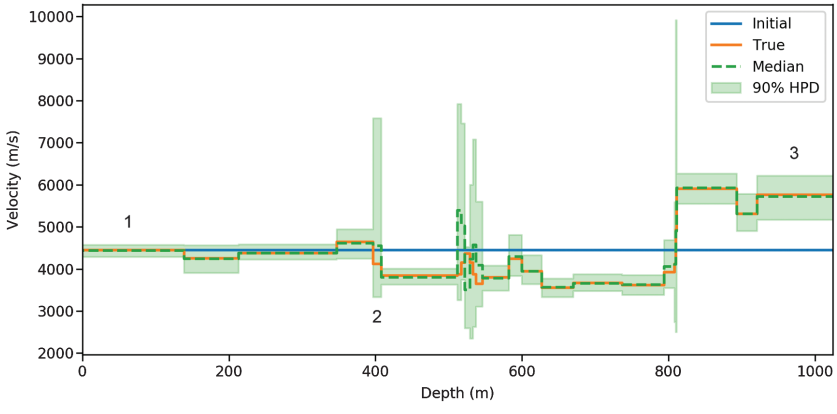


Figure 15. A 1D vertical slice of the velocity field for numerical experiment 3 at the left side of the domain ($x = 0$ m). The dashed green line represents the velocity generated by the median of the posterior distribution. The green-shaded region represents the 90% HPD interval of the velocity values. The orange line represents the true velocity value, and the blue line marks the initial velocity value.

the filter relative residuals is similar to that of Figure 4 from numerical experiment 1.

For this experiment, the precision parameters are $\sigma = \sigma_C = 0.04$ for fine-grid and coarse-grid precision. We define the two step lengths as $\sigma_{\text{step}}^{\text{velocity}} = 50$ m/s and $\sigma_{\text{step}}^{\text{depth}} = 6$ m. The six chains tested 50,000 samples each for a total of 300,000 samples tried. For this experiment, we remove a burn-in period containing the first 11,000 samples from each chain because many of the chains appear to take longer to stabilize than in numerical experiments 1, 2, and 3.

Figure 17 compares the velocity field constructed from the median of the posterior distribution with the initial velocity and true velocity. We can view the true velocity field as three large velocity regions containing several smaller layers. The three regions are from approximately 0 to 400 m depth, 400 to 800 m depth, and 800 to 1024 m depth, respectively. We see in Figure 17 that we can recover the large velocity regions, but we have larger uncertainty when determining the smaller layers within each region. Figure 18 shows that data simulated using the median velocity field is a good match for the observed data. In particular, we recover the direct wave and the larger reflections including those that cross $x = 0$ m at approximately 0.2, 0.275, and 0.41 s. Figure 19a displays 1D vertical slices of the initial, true, and median velocity fields. The median velocity field slice is constructed from the median of the depth and velocity posterior distributions. The shaded region in Figure 19a illustrates the 90% HPD interval for velocity only (the depth uncertainty is shown in Figure 19b). The true velocity for each layer is captured by the 90% HPD interval. In Figure 19b, we see that for the interfaces that mark divisions with large velocity contrasts (interfaces 4 and 7 in Figure 19a), the 90% HPD intervals for the interface depth are narrow. For interfaces with small velocity contrasts (e.g., interfaces 2, 3, and 5 in Figure 19), we see wider 90% HPD intervals for the interface depth, which indicates a greater uncertainty about the position of the layer interface. The selected posterior distributions in Figure 20 correspond to the numbered interfaces 4 and 6 in Figure 19 and the velocities of the materials directly above them. Interface 4, which lies between materials with a large velocity difference, has a narrow posterior distribution, whereas interface 6, which is between two similar velocity materials, has a much broader posterior distribution.

Timing analysis

To determine how much more computationally efficient the two-stage MCMC algorithm is than the one-stage MCMC algorithm for seismic inversion, we look at three metrics:

1) The acceptance rate, defined as the proportion of velocity fields accepted on the fine grid to the number of velocity fields tested on the fine grid (Efendiev et al., 2006). In the two-stage MCMC algorithm, the number of fine-grid simulations is the same as the number of velocity fields accepted on the filter.
2) Reduction in time per sample tried, defined as

$$\text{Reduction per sample tried} = \frac{\text{time per sample(one stage)} - \text{time per sample(two stage)}}{\text{time per sample(one stage)}}, \tag{34}$$

where

$$\text{Time per sample} = \frac{\text{total elapsed time}}{\text{total number of samples tried}}. \tag{35}$$

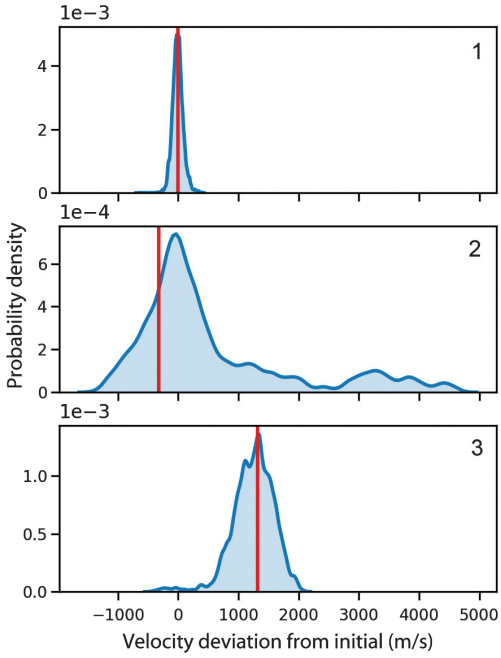


Figure 16. A selection of three KDE posterior distributions for numerical experiment 3 from the velocity regions labeled 1 through 3 in Figure 15. The red line marks the true velocity.

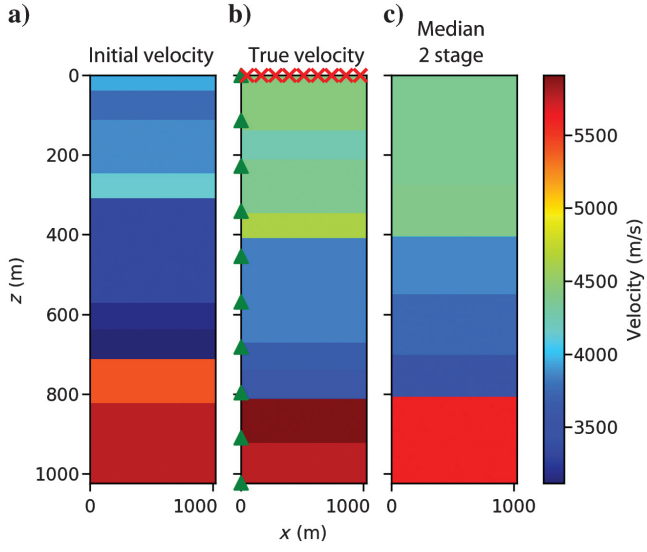


Figure 17. The initial velocity field (a), true velocity field (b), and velocity field constructed from the median of the posterior distribution (c) for numerical experiment 4. On the true velocity plot, the vertical array of 512 receivers is marked with green triangles. The line of 20 sources is marked with red Xs.

Figure 18. The observed receiver data from a vertical array at $x = 0$ m (a), simulated receiver data on the full fine grid using the median velocity field (b) from the two-stage MCMC algorithm, and the difference (c) for numerical experiment 4.

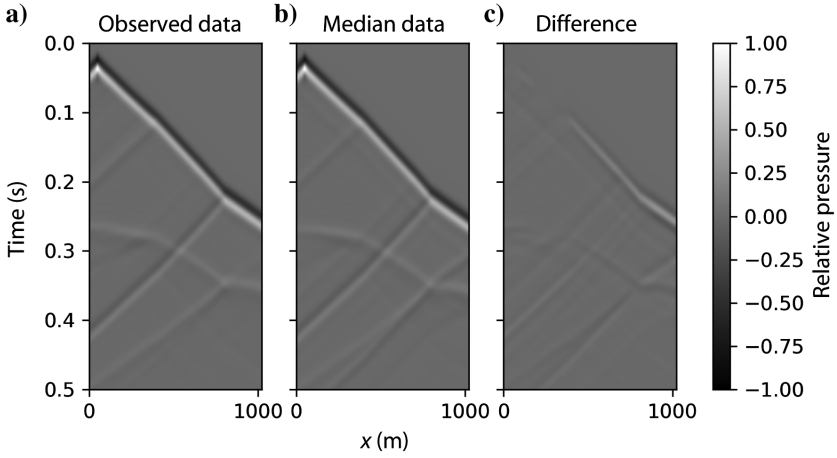


Figure 19. (a) One-dimensional vertical slices of the initial velocity field (blue), true velocity field (orange), and median velocity field (green) for numerical experiment 4. The green shaded region is the 90% HPD interval for velocity only. (b) The median (green X), initial (blue X), and true (orange X) interface depths for numerical experiment 4. The interface number corresponds to the interfaces marked in (a). The green bar shows the 90% HPD interval of the interface depth for each interface.

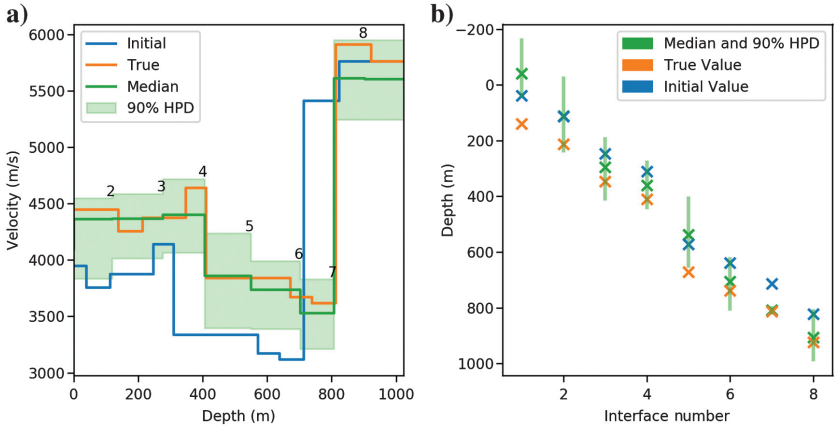
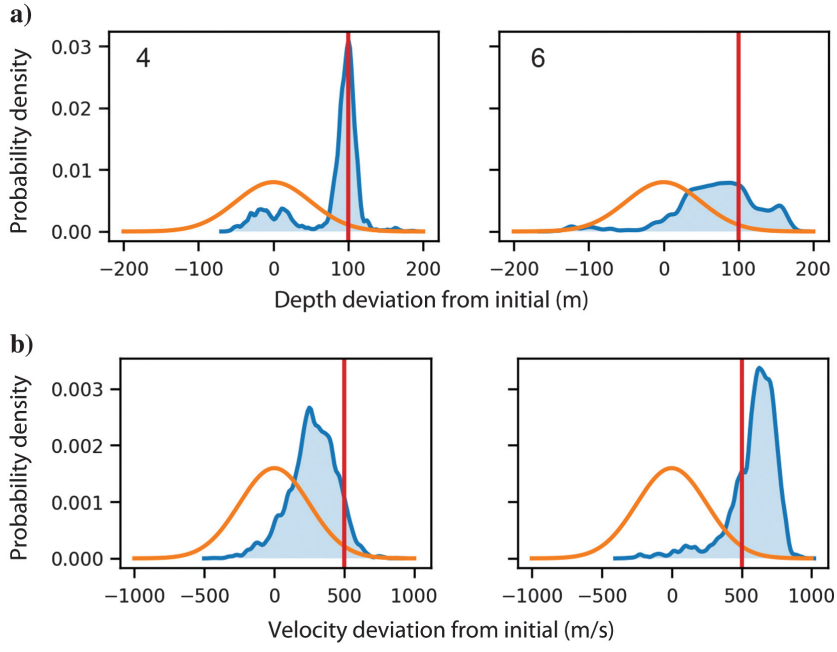


Figure 20. (a) Two posterior distributions (blue-shaded curves) for interfaces 4 and 6 marked in Figure 19. (b) The posterior distributions of the velocities of materials directly above the selected interfaces for numerical experiment 4. The red line marks the true value, and the orange curve indicates the prior distribution.



3) Reduction in time per sample rejected, defined as above with

$$\text{Time per rejected model} = \frac{\text{time spent on filter rejections} + \text{time spent on fine} - \text{grid rejections}}{\text{total number of rejections}}. \quad (36)$$

This metric is particularly important because most of the time spent in the MCMC algorithm is on rejected models.

The following timings were obtained using 144 message passing interface (MPI) processes on 144 cores per chain on the Ganymede cluster at UT Dallas. Changing the precision parameters, tuning parameter, and step length can impact the acceptance rate and reduction in time per trial. We find that a single solution using the upscaling algorithm — i.e., the time needed to produce data for a single model — requires one-third to one-half as long as a single solution using the full fine-grid wave solver for the given configuration.

In Table 1, we see an increase in the acceptance rate from 9% to almost 90% for experiment 4, from 17% to 32% for experiment 3, from 10% to 40% for experiment 1, and from 8% to 33% for experiment 2. We see decreases in the time per rejected model ranging from 33% for experiment 1 to 47% for experiment 4. Experiment 4 achieves a near-perfect reduction in time per rejection (47%) due to its two-stage acceptance rate of approximately 90%. We see that the reduction in time per trial is 39% for experiment 4 and approximately 22% for experiments 1, 2, and 3.

DISCUSSION

Our experiments demonstrate that the two-stage MCMC algorithm is an effective and computationally efficient algorithm for producing posterior distributions of the velocity field unknowns. In all experiments, we see narrow posterior distributions and HPD intervals where we expect to have greater certainty about the values of the velocity field unknowns — e.g., shallower regions or interfaces, broader regions, or interfaces that separate materials with large contrasts in velocity values. Similarly, we see wide posterior distributions and HPD intervals for unknowns that are typically more difficult to recover and have greater uncertainty — e.g., deeper regions and interfaces, regions too thin for the resolution limits of our data, and interfaces that separate materials with similar velocity values. This indicates that the two-stage MCMC algorithm is effective at providing uncertainty information and is consistent with the findings of those who have used two-stage MCMC for flow problems (Efendiev et al., 2005, 2006; Mondal et al., 2010; Ginting et al., 2011; Laloy et al., 2013). This assertion is further supported by the comparison of the two-stage MCMC and the traditional Metropolis-Hastings MCMC for experiment 1. We see that the two-stage MCMC method effectively recreates the solution of the traditional single-stage MCMC.

In terms of computational efficiency, all four experiments lead to substantial improvements in acceptance rate (up to a ninefold increase), reduction in time per rejection (up to 47%), and reduction in time per trial (up to 39%). Therefore, the two-stage MCMC significantly reduces the computational cost of probabilistic velocity inversion without impacting the solution. This demonstration provides a path for future research in reducing the computational cost for seismic inversion via multistage MCMC methods, including investigating other filtering methods to use in combination with operator upscaling in a multiple filter algorithm.

We believe that the significant decrease in computation time makes the two-stage algorithm suitable for 3D data sets and for elastic FWI. In fact, Vdovina et al. (2009) describe operator upscaling for forward modeling of the 3D elastic wave equation. However, our limitation in terms of solving inverse problems with more unknowns lies with the random walk sampler, not the multilevel MCMC method. The random walk sampler is inefficient at sampling from the full posterior space in higher dimensions (Neal, 2011). In future work, we will attempt to remove this restriction by combining the computationally efficient two-stage MCMC method with Hamiltonian Monte Carlo methods that use gradient information to propose samples that are less correlated with the previous sample and are also more likely to be accepted (Neal, 2011).

Table 1. The acceptance rates and reductions in time per sample or rejection for each numerical experiment.

Stage	Acceptance rate	Total CPU time for 50,000 models (days)	Time per rejection (s)	Percent reduction in time per rejection	Time per sample (s)	Percent reduction in time per sample
			Experiment 1: Variable interface			
Two stage	0.401	6.90	10.2	33.3	11.9	21.7
One stage	0.104	8.83	15.2	N/A	15.2	N/A
			Experiment 2: Variable interface with uncertain velocity			
Two stage	0.336	6.75	10.3	32.4	11.7	23.5
One stage	0.082	8.83	15.2	N/A	15.2	N/A
			Experiment 3: Angular unconformity			
Two stage	0.322	7.42	10.2	38.4	12.8	22.3
One stage	0.168	9.56	16.5	N/A	16.5	N/A
			Experiment 4: Flat layer			
Two stage	0.894	6.32	9.5	46.6	10.9	38.7
One stage	0.091	10.31	17.8	N/A	17.8	N/A

We believe that the two-stage MCMC method could be extended to address the cycle-skipping problem in gradient-based FWI using methods such as the source-receiver extension (Huang et al., 2017) and optimal transport (Métivier et al., 2018). Because the two-stage MCMC algorithm is computationally efficient and produces the same uncertainty information as the traditional Metropolis-Hastings MCMC, it is an ideal framework to study the impact of changes to the misfit function on uncertainty quantification for seismic inversion.

However, we note a well-known shortcoming of the Metropolis-Hastings MCMC algorithm in general. This shortcoming is true of traditional MCMC as well as multistage MCMC. The MCMC algorithm relies on user-specified parameters: the tuning parameter in the random walk sampler β and the precision parameters in the likelihood and upscaled likelihood functions σ and σ_C. Finding appropriate values for these parameters often requires trial and error, which adds to the setup time for a problem and can be difficult to determine in the absence of a known solution. One proposed solution to this problem is the class of adaptive Metropolis algorithms that tune the parameters of the proposal distribution adaptively as the chain progresses (Haario et al., 2001).

CONCLUSION

In this work, we demonstrate the effectiveness of a new method for solving seismic inverse problems and quantifying uncertainty — a two-stage MCMC method. A filter stage is used to quickly reject unacceptable velocity field proposals before the proposals are tested using a full fine-grid wave solver. We use operator upscaling for the acoustic wave equation to create a computationally inexpensive filter because it produces data that are well-correlated with the full fine-grid data. We have found that two-stage MCMC provides the same information needed for uncertainty quantification as traditional MCMC, but at a greatly reduced computational cost.

We presented four synthetic numerical experiments. For the first experiment, we inverted for unknowns that characterize interfaces between materials and assumed that the material velocities were known. A comparison of the results from the two-stage MCMC algorithm and the results from the traditional MCMC algorithm shows that the two-stage algorithm has a significantly higher acceptance ratio while obtaining near-identical results to the one-stage algorithm (i.e., the posterior distributions constructed by the one-stage and two-stage algorithms are equivalent). Furthermore, we demonstrated convergence using two metrics: a visual assessment of the evolution of the posterior distributions over the course of the algorithm and a numerical assessment using the maximum PSRF value. In the second experiment, we extended numerical experiment 1 to include unknown velocity values to demonstrate the impact of uncertain velocities on the layer uncertainty information. In the third and fourth experiments, we inverted for material velocities while assuming known interface positions in a model with an angular unconformity and for velocity and layer depth in a flat-layered model. For all four experiments, we recover uncertainty information consistent with what is expected from the acquisition geometry and velocity field structure and capture the true value of the velocity or interface position within 90% HPD intervals for most unknowns. We demonstrated that this algorithm achieves a reduction in time per rejected model of up to 47%, which results in a large overall cost savings because most models (approximately 90%) are rejected. Furthermore, the two-stage MCMC algorithm results in a substantial reduction in time per velocity field tried, up to 40%, as well as a significant increase in acceptance rate, from 10% to a maximum of 90%.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation's Enriched Doctoral Training Program, DMS grant no. 1514808. We thank R. Meek of Pioneer Natural Resources and M. McChesney of Guidon Energy for their insightful guidance on this project. This work has used the Cyber-Infrastructure Research Services computing resources at the University of Texas at Dallas Office of Information Technology and the Seismology Group HPC cluster. F. Pereira was also funded in part by Science without Borders/CNPq-Brazil grant no. 400169/2014-2. G.K. Stuart was supported by the UT Dallas 3D+4D Seismic Imaging and Inversion Research Consortium.

DATA AND MATERIALS AVAILABILITY

All synthetic data associated with this research are available and can be obtained by contacting Georgia Stuart. We are not able to distribute the well log itself.

APPENDIX A

DETAILED BALANCE EQUATION IN TWO-STAGE MCMC

The two-stage MCMC process satisfies the detailed balance equation (Efendiev et al., 2006). For a one-stage Metropolis-Hastings MCMC algorithm, the transition kernel of the Markov chain is

$$K(\theta_n,\theta) = \rho(\theta_n,\theta)q(\theta_n|\theta) + \left(1 - \int \rho(\theta_n,\theta)q(\theta_n|\theta)d\theta_n\right)\delta_{\theta_n}(\theta), \quad \text{(A-1)}$$

where $\rho(\theta_n,\theta)$, as in equation 5, $q(\theta_n|\theta)$ is the transitional probability of the random walk sampler and $\delta_{\theta_n}(\theta)$ is the Dirac mass in x (Robert and Casella, 1999). In two-stage MCMC, this transition kernel is the effective proposal distribution for the fine-grid stage

$$Q(\theta|\theta_n) = g(\theta_n,\theta)q(\theta_n|\theta) + \left(1 - \int g(\theta_n,\theta)q(\theta_n|\theta)d\theta_n\right)\delta_{\theta_n(\theta)}, \quad \text{(A-2)}$$

where $g(\theta_n,\theta)$ is as in equation 7 (Efendiev et al., 2006). The transition kernel for the fine-grid stage of two-stage MCMC is then

$$K(\theta_n,\theta) = \begin{cases} \rho(\theta_n,\theta)Q(\theta|\theta_n) & \theta \neq \theta_n \\ 1 - \int_{\theta\neq\theta_n} \rho(\theta_n,\theta)Q(\theta|\theta_n)d\theta & \theta = \theta_n \end{cases}, \quad \text{(A-3)}$$

where $\rho(\theta_n,\theta)$ is from equation 10 (Efendiev et al., 2006).

The detailed balance equation is

$$P(\theta_n)K(\theta_n,\theta) = P(\theta)K(\theta,\theta_n), \quad \text{(A-4)}$$

where $P(\theta)$ is the prior distribution (Robert and Casella, 1999). For $\theta = \theta_n$, clearly the detailed balance equation is satisfied. For $\theta \neq \theta_n$, we have

$$P(\theta_n)K(\theta_n,\theta) = P(\theta_n)\rho(\theta_n,\theta)Q(\theta|\theta_n) \tag{A-5}$$

$$= P(\theta_n)Q(\theta|\theta_n)\min\left\{1, \frac{P(\theta|d)Q(\theta_n|\theta)}{P(\theta_n|d)Q(\theta|\theta_n)}\right\} \tag{A-6}$$

$$= \min\{P(\theta_n)Q(\theta|\theta_n), P(\theta)Q(\theta_n|\theta)\} \tag{A-7}$$

$$= \min\left\{\frac{P(\theta_n)Q(\theta|\theta_n)}{P(\theta)Q(\theta_n|\theta)}, 1\right\}P(\theta)Q(\theta_n|\theta) \tag{A-8}$$

$$= \rho(\theta,\theta_n)P(\theta)Q(\theta_n|\theta) \tag{A-9}$$

$$= P(\theta)K(\theta,\theta_n). \tag{A-10}$$

Thus, the two-stage transition kernel satisfies the detailed balance equation.

REFERENCES

Akbarabadi, M., M. Borges, A. Jan, F. Pereira, and M. Piri, 2015, A Bayesian framework for the validation of models for subsurface flows: Synthetic experiments: Computational Geosciences, **19**, 1231–1250, doi: 10.1007/s10596-015-9538-z.

Arbogast, T., S. Minkoff, and P. Keenan, 1998, An operator-based approach to upscaling the pressure equation, *in* V. N. Burganos, ed., Computational methods in water resources XII: Computational Mechanics Publications, 405–412.

Berenger, J.-P., 1994, A perfectly matched layer for the absorption of electromagnetic waves: Journal of Computational Physics, **114**, 185–200, doi: 10.1006/jcph.1994.1159.

Biswas, R., and M. Sen, 2017, 2D full-waveform inversion and uncertainty estimation using the reversible jump Hamiltonian Monte Carlo: 87th Annual International Meeting, SEG, Expanded Abstracts, 1280–1285, doi: 10.1190/segam2017-17680416.1.

Brooks, S. P., and A. Gelman, 1998, General methods for monitoring convergence of iterative simulations: Journal of Computational and Graphical Statistics, **7**, 434–455.

Christen, J. A., and C. Fox, 2005, Markov chain Monte Carlo using an approximation: Journal of Computational and Graphical Statistics, **14**, 795–810, doi: 10.1198/106186005X76983.

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

Efendiev, Y., A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick, 2005, An efficient two-stage Markov chain Monte Carlo method for dynamic data integration: Water Resources, **41**, W12423.

Efendiev, Y., T. Hou, and W. Luo, 2006, Preconditioning Markov chain Monte Carlo simulations using coarse-scale models: SIAM Journal on Scientific Computing, **28**, 776–803, doi: 10.1137/050628568.

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**, no. 2, R63–R75, doi: 10.1190/geo2017-0321.1.

Frangos, M., Y. Marzouk, K. Willcox, and B. van Bloemen Waanders, 2011, Surrogate and reduced-order modeling: A comparison of approaches for large-scale statistical inverse problems, *in* L. Biegler, G. Biros, O. Ghattas, M. Heinkenschloss, D. Keyes, B. Mallick, Y. Marzouk, L. Tenorio, B. van Bloemen Waanders, and K. Willcox, eds., Large-scale inverse problems and quantification of uncertainty: John Wiley and Sons, 123–150.

Gamerman, D., and H. F. Lopes, 2006, Markov chain Monte Carlo: Stochastic simulation for Bayesian inference: CRC Press.

Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, 2013, Bayesian data analysis, 3rd ed.: Chapman and Hall/CRC.

Gelman, A., and K. Shirley, 2011, Inference from simulations and monitoring convergence, *in* S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, eds., Handbook of Markov chain Monte Carlo: CRC Press 6, 163–174.

Gibson, R. L., Jr., K. Gao, E. Chung, and Y. Efendiev, 2014, Multiscale modeling of acoustic wave propagation in 2D media: Geophysics, **79**, no. 2, T61–T75, doi: 10.1190/geo2012-0208.1.

Ginting, V., F. Pereira, M. Presho, and S. Wo, 2011, Application of the two-stage Markov chain Monte Carlo method for characterization of fractured reservoirs using a surrogate flow model: Computational Geosciences, **15**, 691–707, doi: 10.1007/s10596-011-9236-4.

Ginting, V., F. Pereira, and A. Rahunanthan, 2015, Multi-physics Markov chain Monte Carlo methods for subsurface flows: Mathematics and Computers in Simulation, **118**, 224–238, doi: 10.1016/j.matcom.2014.11.023.

Gouveia, W. P., and J. A. Scales, 1998, Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis: Journal of Geophysical Research, **103**, 2759–2779, doi: 10.1029/97JB02933.

Haario, H., E. Saksman, and J. Tamminen, 2001, An adaptive Metropolis algorithm: Bernoulli, **7**, 223–242, doi: 10.2307/3318737.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, **57**, 97–109, doi: 10.1093/biomet/57.1.97.

Hong, T., and M. K. Sen, 2009, A new MCMC algorithm for seismic waveform inversion and corresponding uncertainty analysis: Geophysical Journal International, **177**, 14–32, doi: 10.1111/j.1365-246X.2008.04052.x.

Huang, G., R. Nammour, and W. Symes, 2017, Full-waveform inversion via source-receiver extension: Geophysics, **82**, no. 3, R153–R171, doi: 10.1190/geo2016-0301.1.

Kalligiannaki, E., M. A. Katsoulakis, P. Plecháč, and D. G. Vlachos, 2012, Multilevel coarse graining and nano-pattern discovery in many particle stochastic systems: Journal of Computational Physics, **231**, 2599–2620, doi: 10.1016/j.jcp.2011.12.011.

Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, 1991, Monte Carlo estimation and resolution analysis of seismic background velocities: Journal of Geophysical Research, **96**, 20289–20299, doi: 10.1029/91JB02278.

Korostyshevskaya, O., and S. E. Minkoff, 2006, A matrix analysis of operator-based upscaling for the wave equation: SIAM Journal on Numerical Analysis, **44**, 586–612, doi: 10.1137/050625369.

Laloy, E., B. Rogiers, J. A. Vrugt, D. Mallants, and D. Jacques, 2013, Efficient posterior exploration of a high-dimensional groundwater model from two-stage Markov chain Monte Carlo simulation and polynomial chaos expansion: Water Resources Research, **49**, 2664–2682, doi: 10.1002/wrcr.20226.

Lee, Y.-H., 2005, Stochastic error analysis of multiscale flow simulations: The two-phase oil reservoir problem: Ph.D. thesis, Stony Brook University.

Métivier, L., A. Allain, R. Brossier, Q. Mérigot, E. Oudet, and J. Virieux, 2018, Optimal transport for mitigating cycle skipping in full-waveform inversion: A graph-space transform approach: Geophysics, **83**, no. 5, R515–R540, doi: 10.1190/geo2017-0807.1.

Minkoff, S. E., and W. W. Symes, 1997, Full waveform inversion of marine reflection data in the plane-wave domain: Geophysics, **62**, 540–553, doi: 10.1190/1.1444164.

Mondal, A., Y. Efendiev, B. Mallick, and A. Datta-Gupta, 2010, Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump Markov chain Monte Carlo methods: Advances in Water Resources, **33**, 241–256, doi: 10.1016/j.advwatres.2009.10.010.

Mosegaard, K., and A. Tarantola, 1995, Monte Carlo sampling of solutions to inverse problems: Journal of Geophysical Research, **100**, 12431–12447, doi: 10.1029/94JB03097.

Mosegaard, K., and A. Tarantola, 2002, Probabilistic approach to inverse problems, *in* W. H. K. Lee, P. Jennings, C. Kisslinger, and H. Kanamori, eds., International handbook of earthquake and engineering seismology, Part 1: Academic Press, 237–265.

Mulder, W. A., and R.-E. Plessix, 2008, Exploring some issues in acoustic full waveform inversion: Geophysical Prospecting, **56**, 827–841, doi: 10.1111/j.1365-2478.2008.00708.x.

Neal, R. M., 2011, MCMC using Hamiltonian dynamics, *in* S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, eds., Handbook of Markov chain Monte Carlo: Chapman and Hall/CRC, Handbooks of Modern Statistical Methods, 113–162.

Oden, J. T., and J. N. Reddy, 1976, An introduction to the mathematical theory of finite elements: Wiley.

Owhadi, H., and L. Zhang, 2008, Numerical homogenization of the acoustic wave equations with a continuum of scales: Computer Methods in Applied Mechanics and Engineering, **198**, 397–406, doi: 10.1016/j.cma.2008.08.012.

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, doi: 10.1093/gji/ggw061.
Robert, C. P., and G. Casella, 1999, Monte Carlo statistical methods: Springer.
Rosenthal, J. S., 2011, Optimal proposal distributions and adaptive MCMC, *in* S. Brooks, A. Gelman, G. L. Jones, and X.-L. Meng, eds., Handbook of Markov chain Monte Carlo: CRC Press, Handbooks of Modern Statistical Methods, 93–111.
Russell, T. F., and M. F. Wheeler, 1983, Finite element and finite difference methods for continuous flows in porous media, *in* R. E. Ewing, ed., The mathematics of reservoir simulation: SIAM, 35–106.
Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: Reviews of Geophysics, **40**, 1–29, doi: 10.1029/2000RG000089.
Scales, J. A., and R. Snieder, 1997, To Bayes or not to Bayes?: Geophysics, **62**, 1045–1046, doi: 10.1190/1.6241045.1.
Scales, J. A., and L. Tenorio, 2001, Prior information and uncertainty in inverse problems: Geophysics, **66**, 389–397, doi: 10.1190/1.1444930.
Scott, D. W., 1992, Multivariate density estimation: Theory, practice, and visualization: John Wiley and Sons.
Sen, M., and R. Biswas, 2017, Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm: Geophysics, **82**, no. 3, R119–R134, doi: 10.1190/geo2016-0010.1.
Smith, R. C., 2013, Uncertainty quantification: Theory, implementation, and applications: SIAM.
Stuart, G. K., W. Yang, S. Minkoff, and F. Pereira, 2016, A two-stage Markov chain Monte Carlo method for velocity estimation and uncertainty quantification: 86th Annual International Meeting, SEG, Expanded Abstracts, 3682–3687, doi: 10.1190/segam2016-13865449.1.
Symes, W. W., 2008, Migration velocity analysis and waveform inversion: Geophysical Prospecting, **56**, 765–790, doi: 10.1111/j.1365-2478.2008.00698.x.
Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663, doi: 10.1190/1.1443082.
Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.
Vdovina, T., and S. Minkoff, 2008, An a priori error analysis of operator upscaling for the acoustic wave equation: International Journal of Numerical Analysis and Modeling, **5**, 543–569.
Vdovina, T., S. Minkoff, and S. Griffith, 2009, A two-scale solution algorithm for the elastic wave equation: SIAM Journal of Scientific Computing, **31**, 3356–3386, doi: 10.1137/080714877.
Vdovina, T., S. E. Minkoff, and O. Korostyshevskaya, 2005, Operator upscaling for the acoustic wave equation: Multiscale Modeling and Simulation, **4**, 1305–1338, doi: 10.1137/050622146.
Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26, doi: 10.1190/1.3238367.
Yilmaz, Ö., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: SEG Books, Investigations in Geophysics.