Hamiltonian Monte Carlo based time-lapse seismic FWI and uncertainty quantification in CO_2 monitoring: a VSP feasibility study

Jinji Li and Kristopher A. Innanen*

Summary

Time-lapse full waveform inversion (FWI) stands to play an important role in energy-transition applications, particularly in monitoring CO_2 geo-storage. Because of the nonlinearity of the FWI problem, the existence of various sources of uncertainty between the baseline and monitor surveys, and the likely sparseness (or at least variability) of data coverage in these applications, augmenting FWI with robust uncertainty quantification methods is of critical importance. Within the category of Monte Carlo sampling techniques, Hamiltonian (HMC) methods avoid several limitations of classical Markov Chain Monte Carlo (MCMC) approaches, by combining a simulation of Hamiltonian dynamics within the exploration of model space, with a Metropolis acceptance step. The result is a relatively affordable estimation of the uncertainty in models derived from time-lapse FWI. In time-lapse applications, we argue for the use of posterior information deriving from the baseline inversion as prior information within the monitor inversion. We demonstrate that this integration enhances the efficiency and effectiveness of HMC-based FWI, which are critical features of low-cost CO_2 monitoring. Our conclusions are based on simulations and a synthetic feasibility study, here focusing on VSP acquisition configurations.

Introduction

The global Carbon Capture, Utilization, and Storage (CCUS) sector has undergone significant expansion, as of 2022, with a combined annual capacity exceeding millions of metric tons of CO_2. This growth is evidenced by the proliferation of over 100 projects at various stages of development, including the Shute Creek project in the USA (Parker et al., 2011), the Sleipner CCUS project in Norway (Furre et al., 2017), and the QUEST project in Canada (Duong et al., 2019). An integral component of CO_2 geostorage is the assurance of containment and conformance, for which time-lapse seismic monitoring stands to play an important role (e.g., Lumley, 2001; Arts et al., 2004; Chadwick et al., 2010). In time-lapse seismic, changes in subsurface properties of interest are assessed by comparing models constructed from baseline and subsequent surveys. Driving time-lapse seismic with full waveform inversion (FWI) may turn out to be optimal for monitoring CO_2 storage and injection (Nakata, 2022), because of its inherent use of maximal amounts of data information and its ability in principle to provide high-resolution subsurface images. Various schemes have been proposed for this purpose, including parallel FWI (Plessix et al., 2010), sequential FWI (Asnaashari et al., 2015), double difference FWI (Zhang, 2013), and joint inversion (Rittgers et al., 2016). However, although time-lapse FWI shows promise, non-repeatability between surveys due to localized spatial variations, differences in acquisition geometries, and uncertainties in observations (Kotsi, 2020), are especially troublesome for waveform-based methods. Adding to this interest in minimizing the risk of false positives or negatives in the identification of CO_2 containment issues, the uncertainty quantification in this technology is important.

Hamiltonian Monte Carlo, or HMC (Duane et al., 1987) methods, which are a general class of tools for uncertainty quantification, integrate elements from two distinct methodologies: (1) gradient-based optimization, which efficiently identifies optima but lacks comprehensive uncertainty information, and (2) derivative-free Markov Chain Monte Carlo (MCMC) methods, which may overlook potentially valuable derivative information. HMC facilitates targeted exploration of plausible regions of model space, efficiently sampling high-dimensional parameter spaces as part of complex Bayesian inference problems (Neal, 1993; Brooks et al., 2011). Comprehensive discussions of HMC in FWI can be found in Fichtner et al. (2019) and Gebraad et al. (2020).

In this study, we consider the specific application of HMC to time-lapse seismic FWI. This problem, which involves both baseline and monitoring datasets, and requires estimation of two posterior distributions, poses an increased computational workload. Additionally, baseline and monitor surveys often have distinct acquisition geometries and other factors that can be conflated with variations associated with injection and plume evolution. Incorporating statistical information from the baseline inversion into the monitoring inversion has been suggested (Zhou and Lumley, 2021), and utilizing this kind of statistical information in time-lapse FWI using HMC has also been advocated. Kotsi (2020) employed a 4D-HMC algorithm and illustrated its performance on a simple 4D problem. De Lima et al. (2023) demonstrated the benefits of employing informative prior information in time-lapse studies by comparing the sequential and parallel approaches with analyzing the non-repeatability effects associated with different source locations in the surveys. In this study, we conduct synthetic experiments within a near-parallel time-lapse scheme for time-lapse FWI applied to CO_2 injection with a permanent VSP receiver array and varying surface geometries in baseline and monitor surveys. Additionally, we examine the significance of integrating the baseline posterior into the monitor inversion. Our findings underscore its crucial role in accelerating convergence, thereby reducing computational load; and elucidating temporal changes in the subsurface due to the injection rather than other factors.

Theory

In the Bayesian framework, the solution of FWI problems is a probability density function of plausible models that reasonably represent the observations. Such probability density is called the posterior distribution $\rho(\mathbf{m}|\mathbf{d})$. According to Bayes theorem:

$$\rho(\mathbf{m}|\mathbf{d}) = \frac{\rho(\mathbf{m})\rho(\mathbf{d}|\mathbf{m})}{\rho(\mathbf{d})}, \tag{1}$$

where $\rho(\mathbf{m})$ is the model's prior probability distribution, $\rho(\mathbf{d})$ is the probability distribution of the observed data, which is often regarded a normalization factor. $\rho(\mathbf{d}|\mathbf{m})$ in equation 1 is called the likelihood function that represents the conditional probability of a dataset generated by a model. Typically, under the Gaussian assumption, the likelihood and the prior can be shown by (Tarantola, 2005):

$$\rho(\mathbf{d}|\mathbf{m}) \propto \exp\left[-\frac{1}{2}\left(\mathbf{d}_{\text{syn}} - \mathbf{d}_{\text{obs}}\right)^T \mathrm{C}_\mathrm{D}^{-1}\left(\mathbf{d}_{\text{syn}} - \mathbf{d}_{\text{obs}}\right)\right], \tag{2}$$

$$\rho(\mathbf{m}) \propto \exp\left(-\frac{1}{2}\mathbf{m}^\mathrm{T}\mathrm{C}_\mathrm{M}^{-1}\mathbf{m}\right), \tag{3}$$

where $\mathbf{d}_{\text{syn}}$ is the synthetic data generated from a given model with some physics rules, for example, acoustic/elastic wave equations. $\mathrm{C}_\mathrm{D}^{-1}$ and $\mathrm{C}_\mathrm{M}^{-1}$ in equations 2 and 3 is the inverse covariance matrix of the data and the model.

Hamiltonian Monte Carlo, nested from the Hamiltonian dynamics (Hamilton, 1834), can be conceptually understood by fictitiously visualizing a frictionless particle moving along a U-shaped surface with varying heights. In this scenario, the system's state is described by the position of this particle $\mathbf{q}$ and its generalized momentum $\mathbf{p}$. While moving, this particle's potential energy can be represented by $U(\mathbf{q})$, and its kinetic energy is given by $\mathbf{p}^\mathrm{T}\mathbf{M}^{-1}\mathbf{p}/2$, where $\mathbf{M}$ is the mass matrix of this particle. Such mechanics can be described by a Hamiltonian equation $H(\mathbf{p},\mathbf{q})$:

$$H(\mathbf{p},\mathbf{q}) = U(\mathbf{q}) + K(\mathbf{p}). \tag{4}$$

The partial derivatives of the Hamiltonian determine how $\mathbf{q}$ and $\mathbf{p}$ change over time t, as in Hamilton's equations:

$$\frac{dq_i}{dt} = -\frac{\partial H}{\partial p_i} = [\mathbf{M}^{-1}\mathbf{p}]_i, \tag{5}$$

$$\frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i} = -\frac{\partial U}{\partial q_i}, \tag{6}$$

where $i = 1, 2, \ldots, n$ denotes the index in the n-D vector. The fundamental idea of HMC involves sampling from an auxiliary distribution in a phase space $(\mathbf{p},\mathbf{m})$ with twice the original space's dimensions, where $\mathbf{p}$ denotes the momentum vector, and $\mathbf{m}$, which replaces $\mathbf{q}$ in equation 4, represents the variables of interest (e.g., model parameters in inversion problems). The auxiliary (also called canonical) distribution can be shown as (Davey, 2009):

$$\rho(\mathbf{p},\mathbf{m}) = \rho(\mathbf{p})\rho(\mathbf{m}|\mathbf{d}) \propto \exp(-H(\mathbf{p},\mathbf{m})), \tag{7}$$

where $\rho(\mathbf{p})$ is the momentum distribution, and $H(\mathbf{p},\mathbf{m})$ is the Hamiltonian.

Following past studies (e.g., Tarantola, 2005; Dettmer et al., 2010; Dosso et al., 2014; Fichtner et al., 2018; Fu and Innanen, 2020), we assume the model and data covariance to be known, Gaussian, and uncorrelated. Consequently, the posterior, or the potential energy in HMC can be defined as follows:

$$-\log[\rho(\mathbf{m}|\mathbf{d})] = \frac{1}{2}\left(\mathbf{d}_{\text{syn}} - \mathbf{d}_{\text{obs}}\right)^T \mathrm{C}_\mathrm{D}^{-1}\left(\mathbf{d}_{\text{syn}} - \mathbf{d}_{\text{obs}}\right) + \frac{1}{2}\mathbf{m}^\mathrm{T}\mathrm{C}_\mathrm{M}^{-1}\mathbf{m}. \tag{8}$$

We now can outline a fundamental workflow for implementing HMC in the context of FWI problems. First, the momentum vector, $\mathbf{p}$, is drawn from a multi-dimensional normal distribution with zero means and a covariance matrix $\mathbf{M}$. Second, a transition from the current state $(\mathbf{p}_{\text{cur}},\mathbf{m}_{\text{cur}})$ is performed using the Leapfrog method (ISERLES, 1986) to a new state $(\mathbf{p}_{\text{new}},\mathbf{m}_{\text{new}})$, which is accepted with the probability based on a variant of the Metropolis rule (Metropolis et al., 1953):

$$P_{acceptance} = \min\left[1, \exp\left(H(\mathbf{p}_{\text{cur}},\mathbf{m}_{\text{cur}}) - H(\mathbf{p}_{\text{new}},\mathbf{m}_{\text{new}})\right)\right]. \tag{9}$$

Examples

In this section, we demonstrate acoustic FWI tests using HMC in the frequency domain. We create an artificial structure with varying P-wave velocity, before and after CO_2 injection (see Figure 1). Horizontal and vertical grids are set at 300 and 100, with a 20-meter interval. We place 30 VSP receivers in the middle of the model. Baseline inversion involves 75 explosive sources spaced 80 meters apart on the surface, while monitor inversion employs half the number of sources with intervals doubled. Synthetic data spans 8 frequencies (3-15 Hz). We assume Gaussian distribution for the P-wave velocity model and maintain consistent data covariance. Varying integration lengths and time steps follow Fichtner et al. (2018). The mass matrix in HMC is set to the identity matrix. Initial models for all tests are identical as in Figure 1 (c). To mitigate high-resolution speckles caused by random pixel variations, we apply a Gaussian filter between iterations. This aligns the proposals with the wavelength in our datasets, enhancing compatibility with the data. We generate 30,000 samples across all the HMC tests.

The baseline inversion yields an acceptance rate of 74.15%, with approximately 22,245 accepted samples. To evaluate HMC's performance against naive MCMC, we conduct monitor inversions using both methods. For a comparable computational workload, MCMC attempts 200,000 samples, yielding an acceptance rate of around 51.04%. We present the uncertainty analysis in Figure 2, comparing variations between monitor and baseline model means $\mathbf{m}_m - \bar{\mathbf{m}}_b$ across different sampling stages. It is evident that HMC achieves a significant reduction in uncertainty in the injection area compared to MCMC, which requires more sample iterations for similar outcomes. Additionally, the presence of artifacts in undesired areas is mitigated in HMC because of fewer random-walk behaviors. Although there are minor variations in model means across different sampling stages, the standard deviations indicate that HMC is more efficient than naive MCMC.

Next, we use the diagonal of the inverse posterior covariance matrix from baseline inversion as prior information for monitor inversions. For comparison, we conduct monitor inversions without baseline posterior information with other settings remaining identical. Figure 3 demonstrates that, within the initial 5,000 samples, while the model means appear similar, the injection area's uncertainty is notably lower when incorporating the baseline posterior. This suggests that statistical information from baseline inversions significantly contributes to a more affordable and reliable uncertainty analysis in monitor inversions.

Additionally, we explore the constraining role of the baseline posterior in monitor HMC-FWI in more detail. The statistical information of the initial 5,000 samples is depicted in Figure 4. While the model means in Figure 4 (a) and (c) are comparable, Figure 4 (e) exhibits higher overall uncertainty than Figure 4 (c), where the baseline posterior information is utilized. Reference points in Figure 4 (b) and (d) represent the injection area and regions with varying ray path abundance according to our acquisition geometry. In Figure 4 (c) and (f), clear improvements in uncertainty analysis can be observed, particularly in the injection area. Uncertainty in deeper model regions remains to be larger and less improved because of insufficient ray paths. However, areas with abundant information appear the least influenced by the posterior information.

Conclusions

HMC offers more efficient sampling than naive MCMC, but computational costs can still be a concern for larger-scale time-lapse FWI problems. Our experiments suggest that integrating baseline posterior information into monitor inversions could mitigate this issue. By using this information to guide the sampler, we can achieve better convergence and enhanced sampling efficiency, especially when targeting specific areas of interest like CO_2 injection sites. However, further research is required to optimize the incorporation of baseline posterior information, particularly concerning the design of the artificial mass matrix in HMC. Larger-scale experiments reflecting real-world scenarios could yield valuable insights. In summary, our study lays the groundwork for developing more efficient time-lapse inversion techniques, with significant implications for uncertainty quantification in CO_2 monitoring applications.

Acknowledgements

We thank the sponsors of CREWES for continued support. This work was funded by CREWES industrial sponsors and NSERC (Natural Science and Engineering Research Council of Canada) through the grant CRDPJ 543578-19.

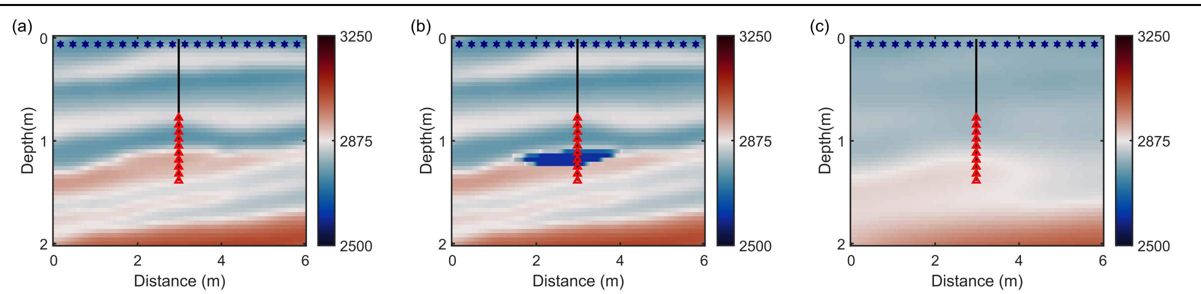

Figure 1: P-wave velocity models and schematic acquisition geometry for HMC-FWI. The source layout is represented by blue stars, and the observation well is shown by the black line with receivers indicated by red triangles. (a) true baseline model. (b) true monitor model. (c) the initial model for baseline and monitor surveys.

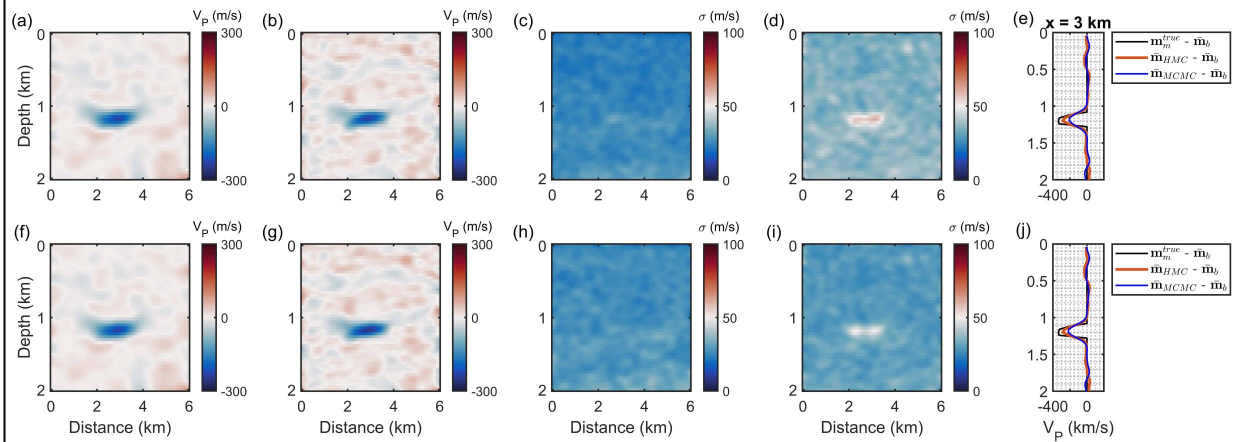

Figure 2: Bayesian time-lapse FWI results. The first row shows results from 2,500 samples in HMC and 15,000 samples in MCMC. (a)-(e), $\mathbf{m}_m - \mathbf{m}_b$ model mean from HMC and MCMC, standard deviations from HMC and MCMC, and a vertical profile in x = 3km with the true variation (black line), variation between baseline model mean and monitor inversion from HMC (red line and MCMC (blue line). (f)-(j), results using 5,000 samples in HMC and 30,000 samples in MCMC.

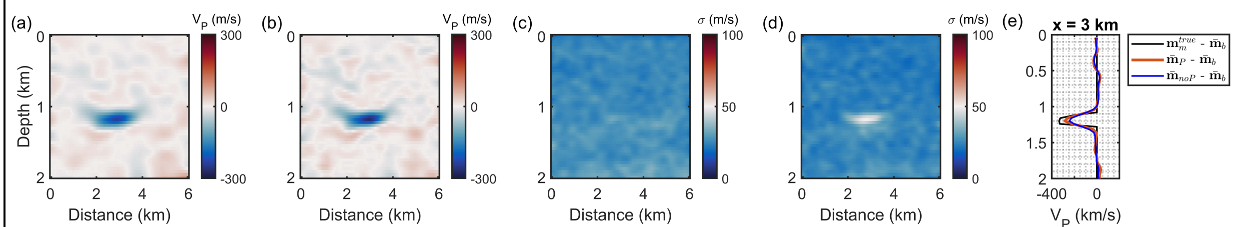

Figure 3: HMC-FWI time-lapse results from 5,000 samples. (a) and (b), $\mathbf{m}_m - \mathbf{m}_b$ model mean with/without baseline posterior. (c) and (d), $\mathbf{m}_m - \mathbf{m}_b$ standard deviations with/without baseline posterior. (e), a vertical profile in x = 3km with the true variation (black line), variation from with (red line), and without (blue line) baseline posterior.

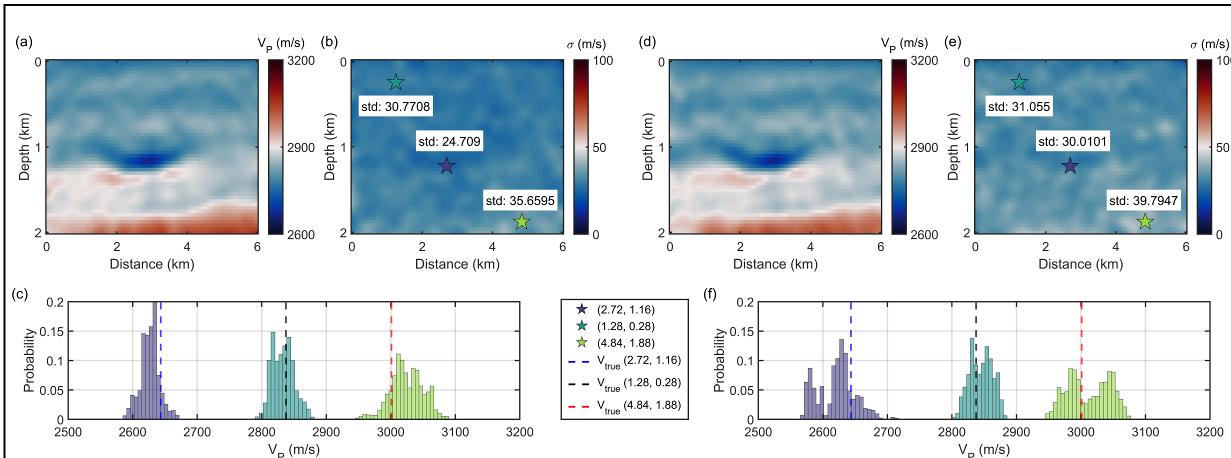

Figure 4: HMC-FWI results with/without baseline posterior. (a)-(c), model mean, standard deviation, and posterior distributions of the reference points in the case baseline posterior is involved. (d)-(f), the same statistical information in the case baseline posterior is not involved. The dashed lines show the true values at the reference positions.