geoconvention
Calgary • Canada • May 12-14 2025

Using deterministic and probability methods in uncertainty quantification in time-lapse FWI

Jinji Li and Kristopher A. Innanen
CREWES, University of Calgary

Summary

Time-lapse full waveform inversion (TL-FWI) is an important tool for applications in the energy transition, especially for monitoring CO_2 storage sites. The complex nature of FWI, combined with uncertainties arising from differences between baseline and monitor surveys and the potential limitations in data coverage, necessitates the integration of advanced uncertainty quantification techniques. Among these, Hamiltonian Monte Carlo (HMC) and Stein Variational Gradient Descent (SVGD) offer promising solutions. HMC leverages Hamiltonian dynamics to navigate the model space efficiently, overcoming challenges associated with traditional Markov Chain Monte Carlo (MCMC) methods. On the other hand, SVGD employs a particle-based approach to minimize the Kullback–Leibler (KL) divergence, providing an effective approximation of the posterior distribution. This research investigates the application of HMC and SVGD to a 2D TL-FWI problem, focusing on their performance under similar computational constraints. By comparing the two methods, we reveal their distinct capabilities and limitations in addressing time-lapse inversion challenges, which can further pave the way for more cost-efficient and reliable CO_2 storage monitoring strategies.

Theory

In the Bayesian framework, solving FWI problems involves estimating a probability density function that represents a range of plausible models consistent with the observed data. This function, referred to as the posterior distribution and denoted as $\rho(\boldsymbol{m}|\boldsymbol{d})$, encapsulates the uncertainty of the solution, and can be expressed as:

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

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

$$\rho(\boldsymbol{m}) \propto \exp\left[-\frac{1}{2}\boldsymbol{m}^T \boldsymbol{C}_M^{-1}\boldsymbol{m}\right]. \tag{3}$$

In this context, $\boldsymbol{d}_{syn}$ represents the synthetic data generated from a model based on physical principles, such as the acoustic or elastic wave equations commonly used in FWI problems. The term $\boldsymbol{C}_D^{-1}$ in equation (2) denotes the inverse covariance matrix of the observed data, while in equation (3) $\boldsymbol{C}_M^{-1}$ refers to the inverse covariance matrix of the model.

Hamiltonian dynamics (Hamilton, 1834) can be intuitively understood by imagining a frictionless particle gliding over a 2D surface with varying elevations. The state of the system is defined by the particle's position $\boldsymbol{x}$, and its generalized momentum $\boldsymbol{p}$. The potential energy of the particle is represented as $U(\boldsymbol{x})$, while its kinetic energy is expressed as $\frac{1}{2}\boldsymbol{p}\boldsymbol{M}^{-1}\boldsymbol{p}$, where $\boldsymbol{M}$ is the particle's

mass matrix. Throughout this motion, the particle is influenced by a "gravitational" force aligned with $-U(\boldsymbol{x})$, which corresponds to the steepest descent direction. These dynamics are captured by the Hamiltonian function, $H(\boldsymbol{p},\boldsymbol{x})$, described as:

$$H(\boldsymbol{p},\boldsymbol{x}) = U(\boldsymbol{x}) + K(\boldsymbol{p}), \tag{4}$$

The partial derivatives of the Hamiltonian determine how the position and momentum change over time, t, according to Hamilton's equations

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

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

where i=1,2,…,n denotes the index in the n-dimensional vector.
Variational inference, however, offers an effective alternative to traditional sampling methods for approximating complex posterior distributions by reformulating the problem as an optimization task. This approach aims to minimize the Kullback-Leibler (KL) divergence (Kullback and Leibler, 1951), which, also known as relative entropy, measures the statistical difference between two probability distributions: the approximate posterior generated by the variational inference process and the true posterior distribution (Blei et al., 2017). Mathematically, the KL divergence is defined as:

$$KL[q(\boldsymbol{m})||\rho(\boldsymbol{d}|\boldsymbol{m})] = E_q[logq(\boldsymbol{m})] - E_q[log\rho(\boldsymbol{d}|\boldsymbol{m})], \tag{7}$$

where $q \in Q$ denotes that $q(\boldsymbol{m})$ is within a predefined and known family of computationally tractable probability distributions Q.
The Stein Variational Gradient Descent (SVGD) method is a gradient-based optimization approach designed to minimize the KL divergence while transforming a set of particles that represent the initial parameter distribution (Liu and Wang, 2016). This iterative process enables the gradual construction of the target posterior distribution starting from an arbitrary initial guess (Blei et al., 2017). The smooth transformation employed by SVGD can be expressed as:

$$T(\boldsymbol{m}) = \boldsymbol{m} + \alpha\phi(\boldsymbol{m}), \tag{8}$$

where $\boldsymbol{m}$ is a multi-dimensional vector consisting of all the inversion variables, α is a quantity that controls the magnitude of update with a smooth transform direction. Accordingly, the gradient of the KL divergence with respect to the perturbation's magnitude can be computed using the Stein's operator (Liu and Wang, 2016):

$$\nabla_\alpha KL[q_T(\boldsymbol{m})||\rho(\boldsymbol{d}|\boldsymbol{m})]|_{\alpha=0} = -E_q\big[A_\rho\phi(\boldsymbol{m})\big], \tag{9}$$

where A_ρ is the Stein's operator (Stein, 1972).

Results

In this section, we describe a series of frequency-domain acoustic TL-FWI experiments conducted using both HMC and SVGD methods, ensuring the computational efforts are comparable. The goal of these experiments is to assess the variations in P-wave velocity in a synthetic model before and after CO_2 injection (refer to Figure 1). The grid for the model is configured with 300 points horizontally and 100 points vertically, each spaced by 10 meters. For all simulations, 30 permanent Vertical Seismic Profiling (VSP) sensors are located at the center of the model to simulate monitoring of CO_2 storage. The baseline inversion includes 76 explosive sources placed at four-grid-point intervals along the surface, providing a dense distribution. The monitor inversions for both HMC and SVGD consist of half the number of surface sources. The synthetic dataset is constructed using 10 distinct frequencies, ranging from 5 Hz to 25 Hz. The inversion focuses on estimating the P-wave velocity, which is assumed to follow a Gaussian distribution. Furthermore, the covariance matrix for the data remains constant throughout the analysis. A parallel time-lapse inversion strategy is applied in this study, with initial models for both HMC-FWI and SVGD-FWI tests chosen from a uniform distribution. These starting models are derived from a smoothed version of the actual baseline model.

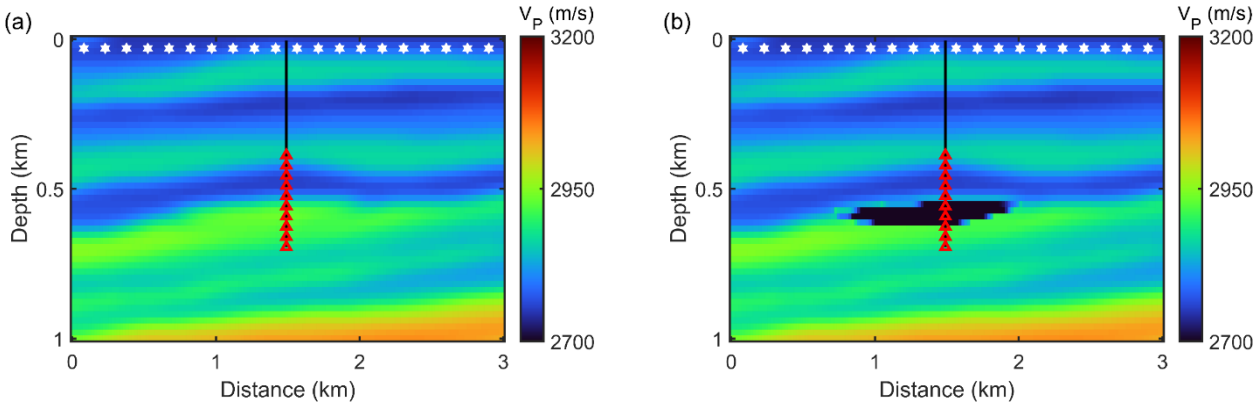

Figure 1. P-wave velocity models. (a) is baseline true model, and (b) is the monitor model.

Figures 2 and 3 illustrate the outcomes of the monitor HMC-FWI and SVGD-FWI methods. In both figures, panel (a) provides a representation of the baseline model, with the model's mean serving as an indicator of its overall structure. In panel (b), the variation in uncertainty levels is displayed across different regions, reflecting the intensity of ray path coverage. The ray paths are categorized into three levels: dense, moderate, and sparse, which are directly related to the quality and distribution of the available seismic data. The histograms shown in panel (c) depict the posterior distributions of the model parameters at three reference points. In the case of the SVGD approach, the uncertainty distribution is notably higher in areas where ray path coverage is sparse. This suggests that SVGD is better at depicting the sensitivity that is defined by the acquisition geometry. Additionally, one of the most interesting observations in the SVGD case is the relatively low uncertainty in the injection area. This is especially noticeable when comparing it to other regions where uncertainty is more pronounced. The low uncertainty in the injection area could be a result of the deterministic nature of the SVGD method.

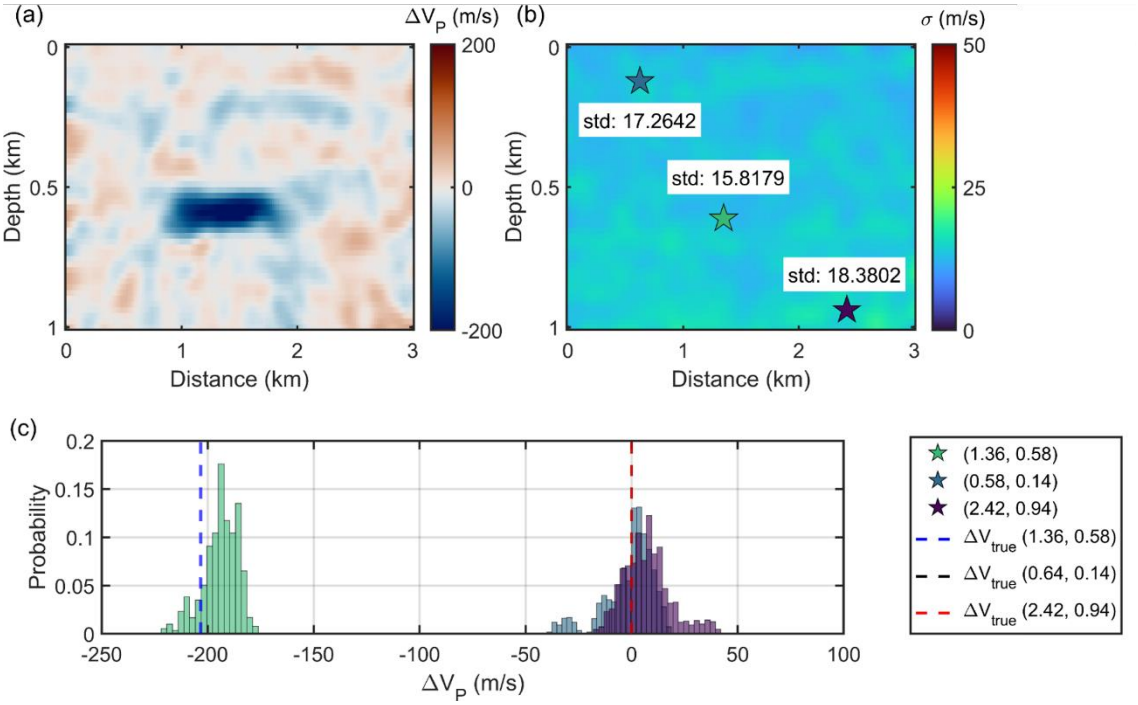

Figure 2. HMC-FWI results. (a) is the mean model, (b) is standard deviation, and (c) is the probability plot for the three reference points.

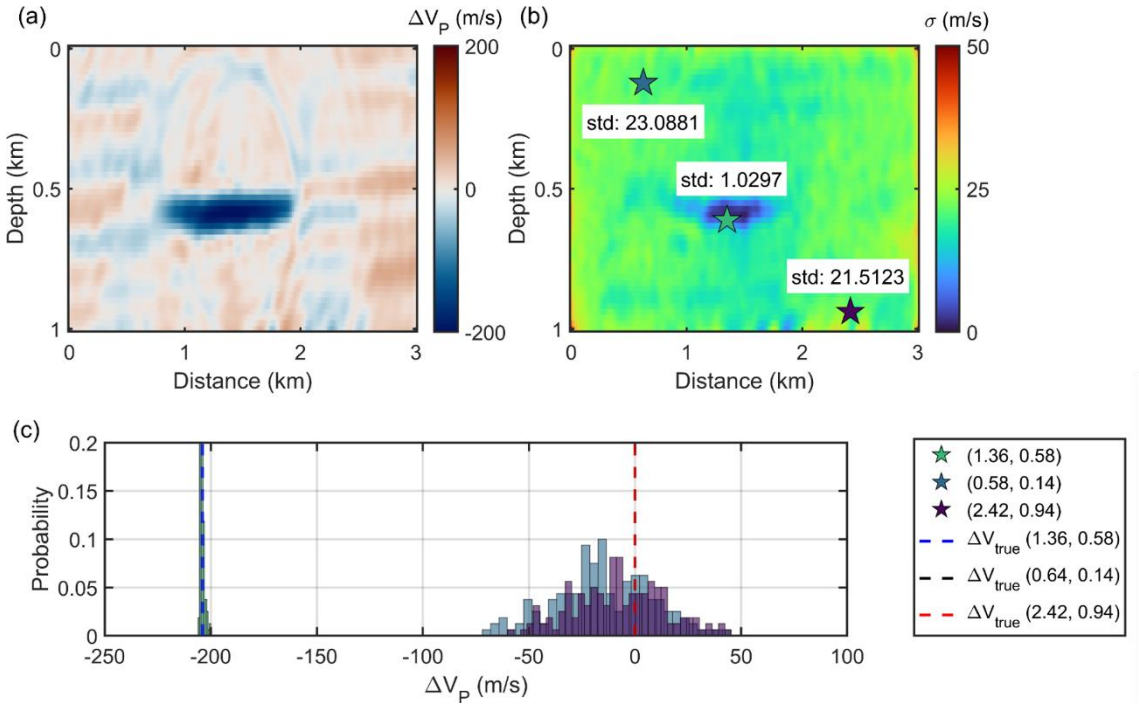

Figure 3. HMC-FWI results. (a) is the mean model, (b) is standard deviation, and (c) is the probability plot for the three reference points.

Acknowledgements

The sponsors of CREWES are gratefully thanked 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. Further support of this work was provided by Emissions Reduction Alberta through the ACT4-SPARSE project.

References

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe, 2017, Variational inference: A review for statisticians: Journal of the American Statistical Association, 112, 859–877.

Gebraad, L., C. Boehm, and A. Fichtner, 2020, Bayesian elastic full-waveform inversion using hamiltonian monte carlo: Journal of Geophysical Research: Solid Earth, 125, e2019JB018428. e2019JB018428 10.1029/2019JB018428.

Hamilton, W. R., 1834, On a general method in dynamics; by which the study of the motions of all free systems of attracting or repelling points is reduced to the search and differentiation of one central relation, or characteristic function: Philosophical Transactions of the Royal Society of London, 124, 247–308.

Kullback, S. and R. A. Leibler, 1951, On Information and Sufficiency: The Annals of Mathematical Statistics, 22, 79 – 86.

Liu, Q. and D. Wang, 2016, Stein variational gradient descent: a general purpose Bayesian inference algorithm: Proceedings of the 30th International Conference on Neural Information Processing Systems, 2378–2386, Curran Associates Inc.