4D Multiparameter Adaptive Metropolis Hastings Inversion

Maria Kotsi†, Alison Malcolm†, & Gregory Ely‡

† Memorial University of Newfoundland
‡ Massachusetts Institute of Technology

MEMORIAL UNIVERSITY

Abstract

In geophysical imaging, uncertainty quantification is crucial for decision making. 4D seismic imaging aims to accurately recover changes that take place within a reservoir. These changes are typically characterized by their magnitude and their extent. We perform a Bayesian inversion using a Metropolis Hastings algorithm to sample our posterior distribution of 4D velocity models given observed data. To model the 4D change we use a discrete cosine transformation, and attempt to recover the lowest frequency coefficients, so that we can model realistic changes with only a few degrees of freedom. Unlike most uncertainty quantification methodologies that use expensive forward solvers, we speed up our computations by using a numerically exact local acoustic solver.

Local Acoustic Solver

The fundamental idea of the local solver is to update a model only in a subdomain of interest while still taking into account all data available. Here, we use the frequency based local solver developed by [5]. In order to use the solver, we first split the full domain into an exterior domain and a local subdomain (Figure 1). The model in the exterior domain together with the initial guess within the local domain make up the background model $\mathbf{m_0}$. In the particular setup of Figure 1 and for a single shot and single frequency one wavefield solve in the local domain takes approximately 0.087 seconds, whereas one wavefield solve in the full domain takes approximately 3 seconds.

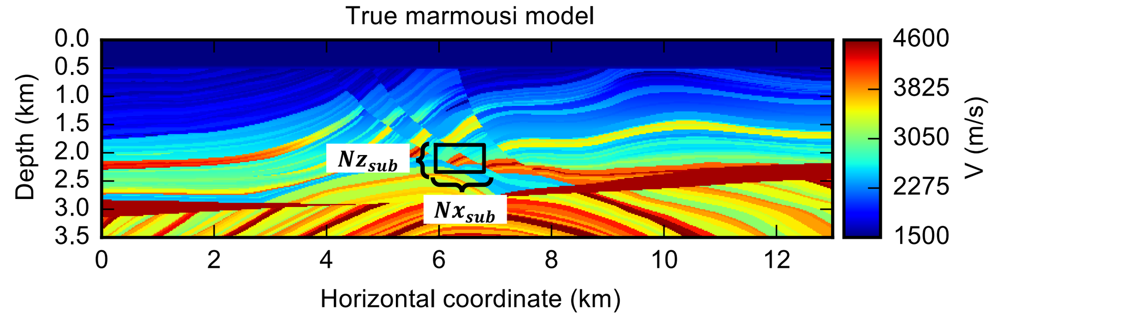

Figure 1: The true baseline velocity model with the black box representing the location of the local domain.

4D Problem Setup and Posterior Calculation

Following [3] we express Bayes' theorem in terms of model residuals $\delta\mathbf{m}$ (4D velocity changes) and observed data residuals $\delta\mathbf{d}$ as

$$p(\delta\mathbf{m}|\delta\mathbf{d}) = \frac{p(\delta\mathbf{d}|\delta\mathbf{m})p(\delta\mathbf{m})}{p(\delta\mathbf{d})}. \quad (1)$$

We compute the likelihood function [4] via

$$L(\delta\mathbf{m}) \equiv p(\delta\mathbf{d}|\delta\mathbf{m}) \propto exp\left[-\frac{1}{2}(F(\delta\mathbf{m}) - \delta\mathbf{d})^T {\Sigma_d}^{-1}(F(\delta\mathbf{m}) - \delta\mathbf{d})\right]. \quad (2)$$

To sample our posterior we use an Adaptive Metropolis Hastings algorithm [2], where the proposal distribution is updated by the history of the process. We first run the Non-Adaptive Metropolis Hastings for a number of iterations (here $N_c = 1000$) with a fixed step-size $C_i = \sigma$. For all remaining iterations we update C_i such that

$$C_i = S_d(Cov[\delta\mathbf{m_0}, ..., \delta\mathbf{m_{i-1}}] + \epsilon\mathbf{I}_d), \quad (3)$$

where $d = length(\delta\mathbf{m_0}, ..., \delta\mathbf{m_{i-1}})$, and $S_d = \frac{2.4^2}{d}$ and $\epsilon \ll 1$.

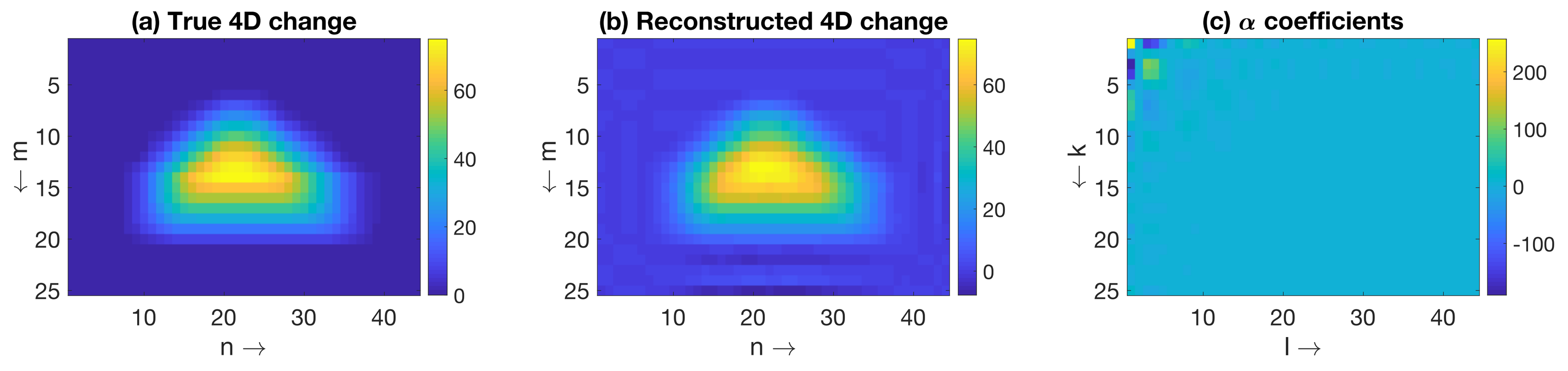

Figure 2: (a) The true time-lapse change with smoothed edges. (b) The reconstructed time-lapse change. (c) $\vec{\alpha}$ coefficients.

Discrete Cosine Transform (DCT)

The DCT is a linear transformation that transforms an n-length vector of amplitudes to an n-length vector containing the coefficients of n different cosine functions [1]. The 2D DCT coefficients for an m-by-n matrix for the k_{th} and l_{th} degree are:

$$DCT_{kl} = \alpha_k\alpha_l \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} \frac{\cos\pi(2m+1)k}{2M}\frac{\cos\pi(2n+1)l}{2N}, \quad (4)$$

where the values of α_k and α_l act as normalizing constants.

Numerical Example with 20 Degrees of Freedom

1. We generate 1100 DCT matrices ($Nz_{sub}*Nx_{sub}$=1100) using Equation 4. Each of these DCT matrices has $m = Nz_{sub}$ rows and $n = Nx_{sub}$ columns. We then generate what we refer to as the Φ matrix by
$$\Phi = [DCT_1(:) \quad DCT_2(:) \quad ... \quad DCT_{1100}(:)]. \quad (5)$$
2. Define the time-lapse change $\delta\mathbf{m}$ such that $\delta\mathbf{m} = \Phi\vec{\alpha}$, where $\vec{\alpha}$ are the coefficients used to generate the DCT transformation.
3. Solve an inverse problem to recover all $\vec{\alpha}$ coefficients (Figure 2).
4. Reduce the dimensionality of the problem by choosing only a subset of the $\vec{\alpha}$ coefficients.
5. Single shot and single frequency of 8 Hz.
6. Run eight different Markov Chains for 100,000 iterations and discard the first half to drop the dependency on the starting model (Figure 3).

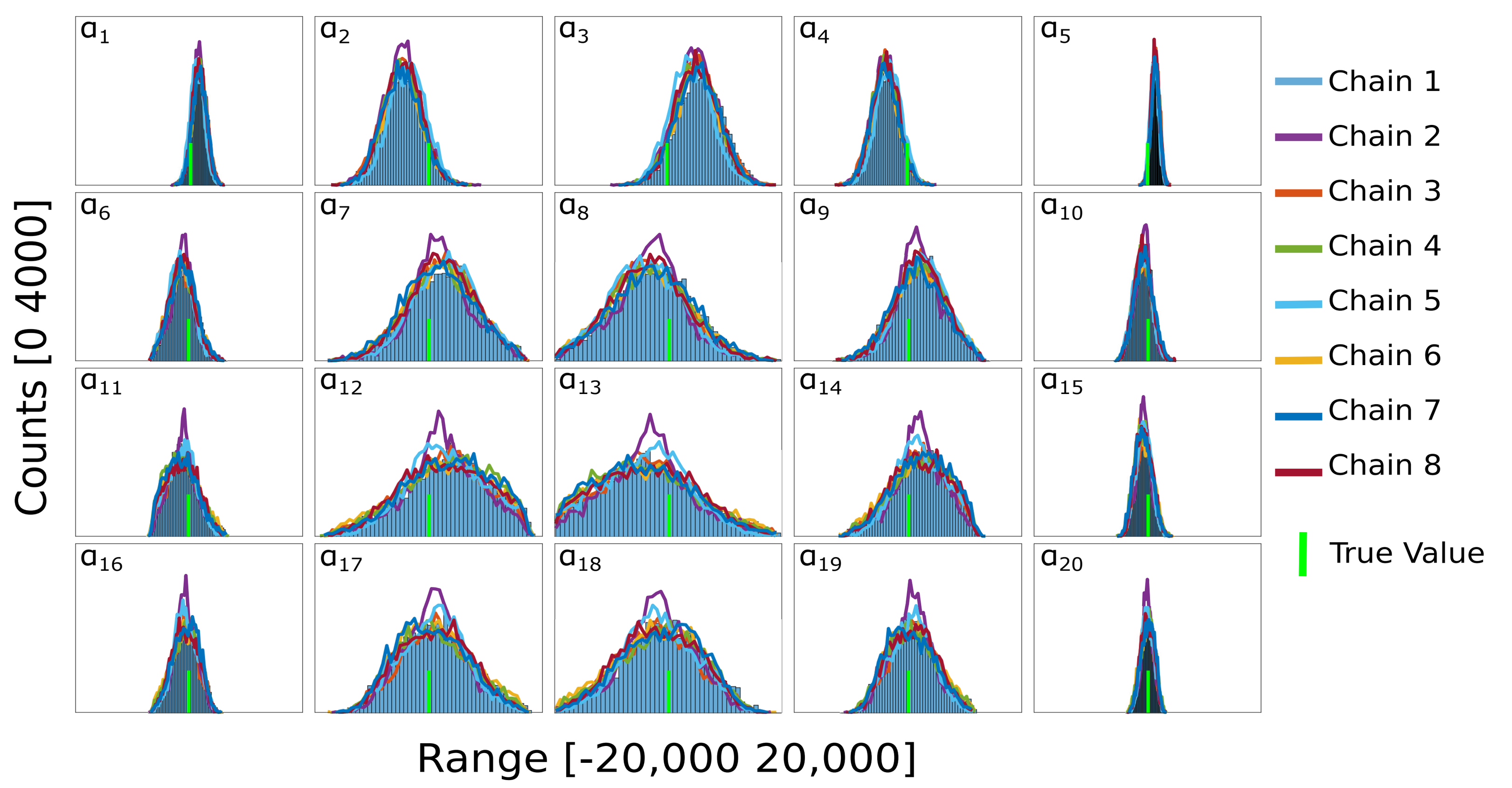

Figure 3: Histograms of the recovered α coefficients for the eight Markov chains. The green line represents the true value for each coefficient.

Quantities of Interest

Typically, 4D changes are characterized by their magnitude and their extent. Thus, we define the following three quantities of interest: (1) vertical extent of the anomaly, (2) horizontal extent of the anomaly, (3) average velocity of the anomaly.

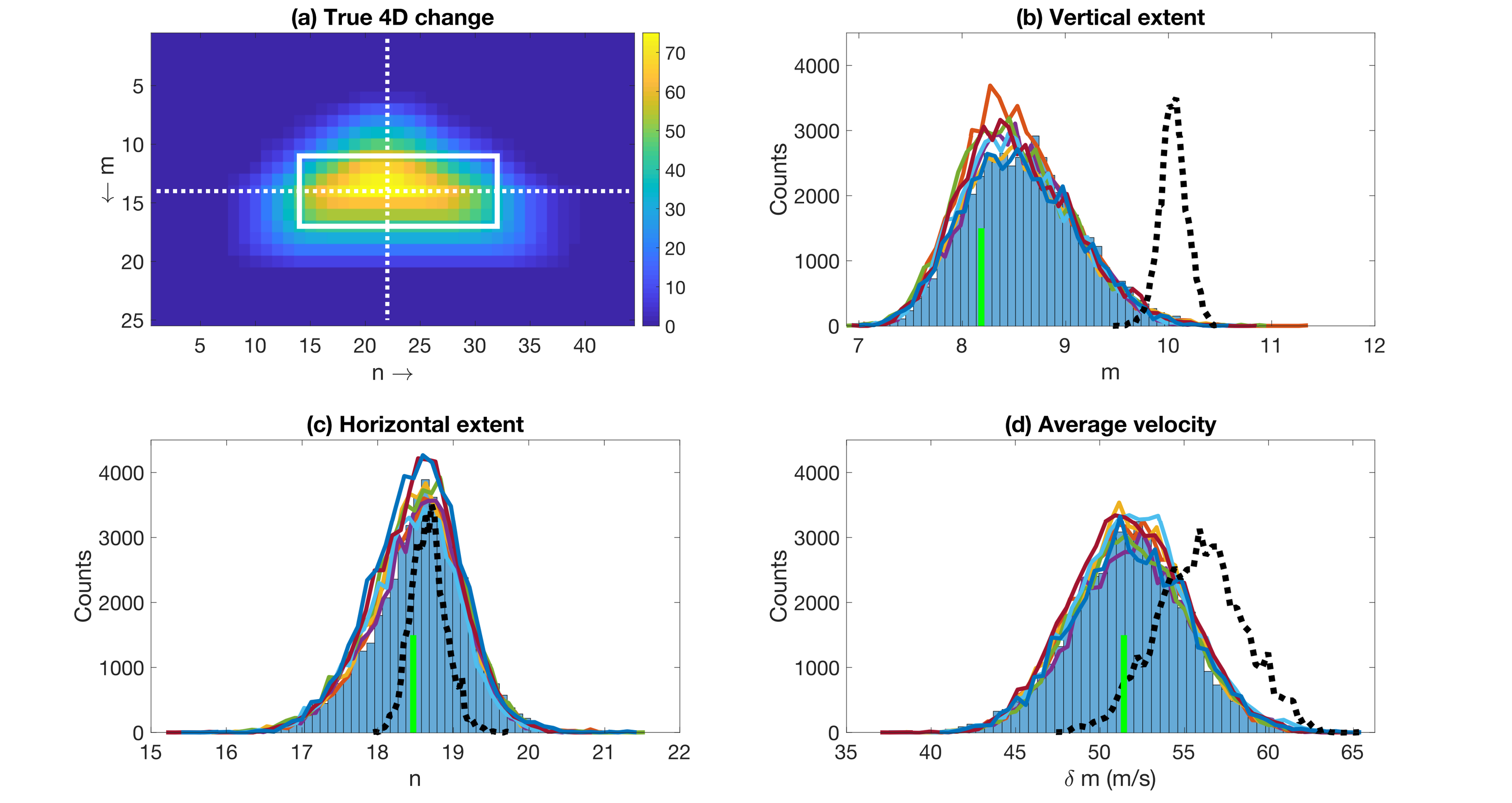

Figure 4: (a) The true time-lapse anomaly with the two dotted white lines indicating the extracted lines used for the vertical and horizontal extent calculations. The white box represents the area in which we computed the average velocity of the time-lapse change. (b) The recovered histograms for the anomaly's vertical extent from the eight Markov Chains. (c) Recovered histograms for the anomaly's horizontal extent from the eight Markov Chains. (d) Recovered histograms for the anomaly's average velocity. In (b), (c), (d) the dotted black line represents the histogram from the 12 DoF while the solid green line is the true average velocity computed from the model in (a).

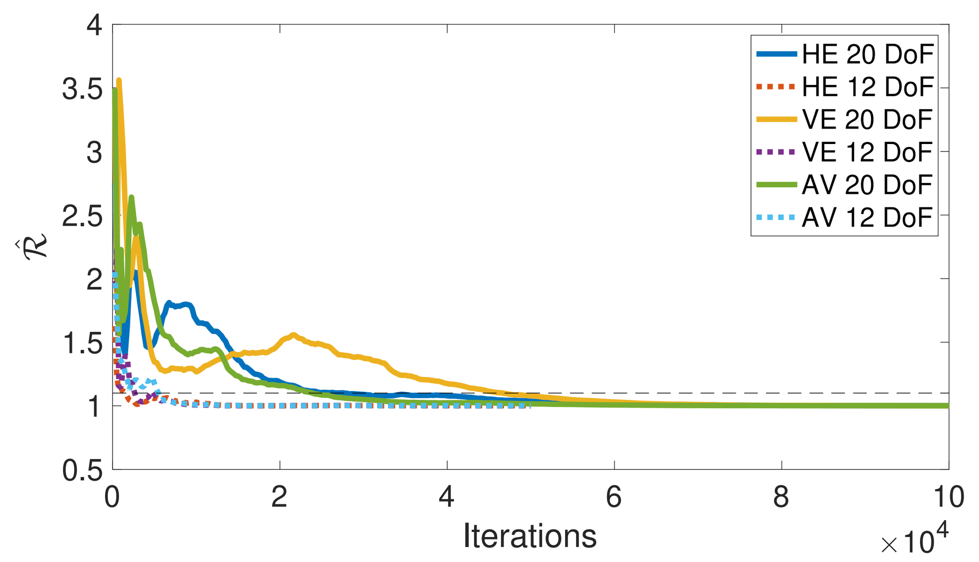

Figure 5: $\hat{\mathcal{R}}$ for the three quantities of interest for both the 20 and 12 DoFs.

To assess convergence we use the $\hat{\mathcal{R}}$ criteria, where we compare the average variance of the quantity of interest within each chain (var_m) to the variance of all of the chains together (var_{mix}) (Figure 5). Generally convergence is declared if $\hat{R}$ is less than 1.1. Note that this convergence does not occur until close to 50,000 models have been sampled meaning that using a method in which you can only sample a few thousand models will not result in a stable, converged solution.

Multiple Frequencies and local minima

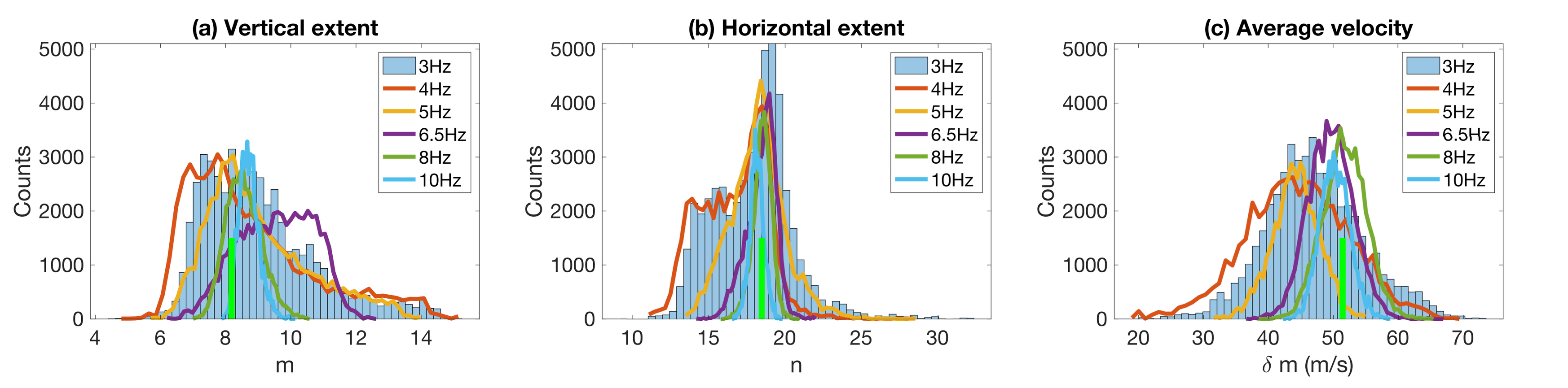

Figure 6: Recovered distributions of the three quantities of interest at different frequencies.

We evaluate the presence of local minima, in terms of the mean of the distribution for the quantities of interest. To do so, we run the MCMC algorithm for one of the chains for the frequencies of 3.0, 4.0, 5.0, 6.5, 8.0, 10.0 Hz. Figure 6 shows the resulted distributions for the three quantities of interest.

Main Contributions

- Propose a local acoustic solver for a fast 4D Bayesian inversion.
- Created a framework that calculates time-lapse uncertainty quantification in a targeted way that is computationally feasible.
- Address the dimensionality issue using image compression.
- Robust framework for multiple degrees of freedom.

References

[1] N. Ahmed, T. Natarajan, and K. R. Rao. Discrete Cosine Transform. *IEEE Transactions on Computers*, C-23(1):90–93, 1974.

[2] H. Haario, E. Saksman, and J. Tamminen. An adaptive metropolis algorithm. *Bernouli*, 7(2):223–242, 2001.

[3] M. Kotsi, A. Malcolm, and G. Ely. 4D Full-Waveform Metropolis Hastings Inversion Using a Local Acoustic Solver. *SEG International Exposition and 88th Annual Meeting, Expanded Abstracts*, pages 5323–5327, 2018.

[4] A. Tarantola. Inverse problem theory and methods for model parameter estimation. *SIAM*, 2005.

[5] B. Willemsen, A. Malcolm, and W. Lewis. A numerically exact local solver applied to salt boundary inversion in seismic full waveform inversion. *Geophysical Journal International*, 204:1703–1720, 2016.

Acknowledgements

This work is supported by Chevron and with grants from the Natural Sciences and Engineering Research Council of Canada Industrial Research Chair Program (IRCPJ 491051-14) and InnovateNL and by the Hibernia Management and Development Corporation.