Bayesian inference based on algorithms: MH, HMC, Mala and Lip-Mala for Prestack Seismic Inversion

Richard Perez-Roa[1], Saba Infante[2,3], Gabriel Barragan[2], Raul Manzanilla[2]

[1]Geology and Applied Geophysics Research Group, School of Earth Sciences, Energy and Environment, Yachay Tech University, Hacienda San José, Urcuquí, 100119, Ecuador
[2]Numerical Analysis and Data Science Research Group, School of Mathematical and Computational Sciences, Hacienda San José, Urcuquí, 100119, Ecuador
[3]Department of Mathematics, FACyT Universidad de Carabobo, Naguanagua, Valencia, 2005, Venezuela

Correspondence to: Richard Perez-Roa (rperez@yachaytech.edu.ec)

Abstract. Seismic inversion for estimating elastic properties is a key technique for reservoir characterization after drilling. The choice of inversion method strongly influences the accuracy, efficiency, and reliability of results. Bayesian inference based on Markov Chain Monte Carlo (MCMC) algorithms provides a robust framework for incorporating data uncertainty and prior geological knowledge. In this study, we compare the performance of four inversion methods—Metropolis-Hastings (MH), Hamiltonian Monte Carlo (HMC), the Metropolis-Adjusted Langevin Algorithm (MALA), and its variant Lip-MALA—in prestack seismic inversion using both synthetic models and real data from an eastern Venezuelan hydrocarbon reservoir. Results indicate that gradient-based methods (HMC, MALA, Lip-MALA) outperform MH in velocity estimation, while density inversion remains more challenging. MH and MALA achieve shorter execution times, whereas HMC and Lip-MALA improve accuracy at higher computational cost. This analysis evaluates mean values and standard deviation (SD) estimates for P-wave velocity, S-wave velocity, and density, with quality assessed through correlation metrics, objective function behavior, seismic traces, and Root Mean Square Error (RMSE). A two-dimensional inversion with real data further demonstrates algorithms performance under complex geological conditions. The findings highlight trade-offs between accuracy and efficiency, providing practical guidelines for selecting inversion method in seismic reservoir characterization.

1 Introduction

The accurate characterization of hydrocarbon reservoirs is fundamental for effective ~~reservoir management.~~exploration and production decisions. This ~~task~~ requires integrating ~~two~~ complementary sources of information: general ~~reservoir~~geological knowledge from analogous reservoirs and rock physics principles, and reservoir-specific observations~~. General knowledge arises from analogous reservoir studies and established principles in seismic and rock physics, whereas specific observations include~~ from well logs, seismic surveys, and production history.

Seismic data play a central role ~~in reservoir characterization because of~~ due to their extensive spatial coverage~~. Unlike well logs, which are restricted to individual boreholes, seismic surveys provide~~, providing a continuous description of the ~~reservoir.~~subsurface beyond individual well locations. To ~~leverage this information,~~translate seismic amplitudes ~~must be~~

transformed into meaningful elastic and petrophysical properties relevant to reservoir description. This transformation is performed through, such as P-wave velocity (V_P), S-wave velocity (V_S), and density (ρ), seismic inversion, a is employed. This geophysical inverse problem that aims to infer subsurface properties (e.g., velocities, density, lithology) from observed seismic data. The forward modelling process that relates model parameters to recorded data can be, typically expressed as (Tarantola, 2005):

$$\boldsymbol{d}_{obs} = g(\boldsymbol{m}) + \xi, \tag{1}$$

where $\boldsymbol{d}_{obs}$ are the observed data, g is the forward operator linking the model parameters $\boldsymbol{m}$ to the data, and ξ represents measurement and modeling errors.

Traditional seismic inversion methods can be broadly categorized into deterministic and stochastic approaches. Deterministic methods, such as least-squares inversion (Tarantola, 2005) and gradient-based optimization (Buland and Omre, 2003), provide single best-fit models but often lack robust uncertainty quantification. Stochastic methods, including simulated annealing (Ma, 2002) and genetic algorithms, explore the model space more broadly but can be computationally intensive and may not guarantee convergence to the posterior distribution.

In prestack amplitude versus offset (AVO) inversion (Helland-Hansen et al., 1997; Ma, 2002; Buland and Omre, 2003), the problem is ill-posed (Landa and Treitel, 2016): solutions are non-unique and extremelyhighly sensitive to measurementdata noise and modeling errors. (Landa and Treitel, 2016). The most important elastic parameters are P-wave velocity (V_P), S-wave velocity (V_S), and density (ρ)(V_P, V_S, ρ), from which Lamé parameters can be derived. These are sensitive to rock fluid content and saturation (Clochard et al., 2009) and can be further related to petrophysical parameters such as porosity, sand/shale ratio, and gas saturation (Goodway, 2001). Accurately estimating these parameters is therefore crucial for identifying hydrocarbon accumulations.

To address these challenges, Bayesian inference provides has emerged as a robustpowerful framework for seismic inversion, as it explicitly accounts for uncertainty and allows the combination of incorporates prior geological knowledge with the likelihood of seismic data. Monte Carlo methods, particularly and quantifies uncertainty through the posterior distribution. Markov Chain Monte Carlo (MCMC) algorithms, are widely used to sample from methods are particularly well-suited for sampling complex, high-dimensional posterior distributions, address nonlinearity, and quantify uncertainty in nonlinear inverse problems (Bosch et al., 2007).

Several MCMC algorithms have been proposed and applied in geophysics.geophysical inversion. The Metropolis–Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) constructsis a stationary Markov chain that converges to the target posterior distribution. A more generalfoundational method is but can suffer from slow convergence in high-dimensional spaces. Hamiltonian Monte Carlo (HMC), originally developed in lattice quantum chromodynamics) (Duane et al., 1987) and later applied to Bayesian neural networks (; Neal, 1996; MacKay, 2003; Bishop, 2006). HMC has since been employed in

fields ranging from molecular simulation (Dubbeldam et al., 2016) to seismic inversion 2012) leverages gradient information to propose more efficient moves, making it suitable for high-dimensional problems, as demonstrated in full-waveform inversion (Gebraad et al., 2020;) and acoustic inversion (de Lima et al., 2023). In geophysics, HMC sampling has been combined with adjoint techniques, mass-matrix optimization, and full waveform inversion to improve efficiency and resolution in noisy and limited-data scenarios.

Other algorithms are derived from Langevin dynamics. The -based methods, such as the Metropolis-aAdjusted Langevin aAlgorithm (MALA) (Roberts and Tweedie, 1996; Welling) and Teh, 2011) exploits gradient information to accelerate convergence. Lip-MALA, a more recent variant, incorporates a locally Lipschitzits adaptive step size to enhance sampling efficiency and stabilityvariant Lip-MALA (Nemeth and Fearnhead, 2020θ1; Izzatullah et al., 2021).2021), use gradient information to guide proposals, improving mixing and convergence in complex posteriors.

Building upon this background, the present study investigates the impact of inversion method choice on Despite these advances, a systematic comparison of these MCMC algorithms in the context of prestack seismic inversion—particularly using both synthetic and real data in 1D and 2D settings—remains limited. Most studies focus on a single algorithm or synthetic cases, leaving a gap in understanding their practical performance under realistic geological complexity and computational constraints.

This study aims to fill this gap by conducting a comprehensive evaluation of four MCMC algorithms—MH, HMC, MALA, and Lip-MALA—for Bayesian prestack seismic inversion. Specifically, we compare theWe assess their performance of MH, HMC, MALA, and Lip-MALA algorithms in estimating elastic parameters from (V_P, V_S, ρ) using both noise-free synthetic data and real seismic data. Thedata from an eastern Venezuelan reservoir. Our analysis evaluates their performanceincludes:

- Quantitative metrics: mean, standard deviation, correlation, and RMSE.
- Computational efficiency: acceptance rates and execution times.
- Convergence diagnostics: multivariate effective sample size (mESS).
- Practical applicability: extension to 2D inversion with real data.

Our contributions are threefold:

1. We provide a benchmark comparison of MCMC algorithms in terms of seismic inversion, highlighting trade-offs between accuracy, uncertainty quantification, geological realism, and computational feasibility. Diagnostic methods are used to assess the quality of generated samplescost.
2. We demonstrate the scalability of these methods from 1D to 2D real-data scenarios, validating their robustness in geologically complex settings.

3. We offer practical guidelines for algorithm selection based on project requirements, data characteristics, and available computational resources.

In Prestack AVO, Bayesian approaches with MCMC have been previously applied to estimate impedance or elastic parameters (e.g., Bosch et al., 2007; Wu et al., 2019), but these studies focus on a single algorithm or mainly on synthetic cases, but we go further and test it on real data.

The ~~rest~~remainder of this paper is organized as follows~~. In~~: Section 2 ~~a review of~~formulates the ~~theory of~~ seismic inversion ~~and amplitude versus offset (AVO) analysis.~~problem within a Bayesian framework and describes the forward modeling approach based on AVO theory. Section 3 ~~outlines~~details the ~~four~~theoretical background of the MCMC algorithms ~~considered~~. Section 4 presents ~~the results for~~result from synthetic and real data experiments. Section 5 ~~the results are analysed. Finally,~~discusses the implications and practical recommendations. Section 6 ~~summarizes~~concludes the ~~main conclusions~~study.

2 The Seismic Inversion Problem

~~Seismic~~ Having established the importance of Bayesian MCMC methods for prestack seismic inversion ~~uses seismic traces to infer subsurface properties model, denote usually by $\boldsymbol{m} \in \mathbb{R}$, such as velocities and densities from observed seismic data $\boldsymbol{d}_{obs} \in \mathbb{R}^{N_D}$, where $N_{\boldsymbol{m}}$ and N_D are the dimensions~~in Section 1, this section provides a comprehensive formulation of the ~~model parameters and the observed seismic data, respectively. This~~inverse problem ~~can be formulated~~ within a ~~Bayesian framework, where inversion is treated as a~~ rigorous statistical framework. We detail the mathematical foundations, model parameterization, forward modeling approach, and inference ~~problem.~~

~~In~~workflow that form the basis of our comparative study. The complete Bayesian inference~~, we begin with a prior probability distribution of the subsurface parameters model, representing existing geological knowledge before considering the seismic observations data. By incorporating seismic data through Bayes' theorem, this prior distribution is updated to obtain the posterior probability distribution, which reflects the degree of confidence in the estimated model parameters. The posterior distribution allows a rigorous~~ flow implemented in this study is visually summarized in Figure 1, which illustrates the steps from input data to subsequent uncertainty quantification ~~of uncertainty, as it integrates contributions from seismic data, prior information, and the forward modelling operator.~~

~~Fully characterizing the posterior distribution generally requires sampling the model space, which is computationally demanding. In this work, we adopt the Bayesian inference framework commonly used in geophysical inversion (Tarantola, 2005; Bosch, 2004; Izzatullah et al., 2021; de Lima et al., 2023), which provides a practical foundation for implementing seismic Bayesian inversion in real-world geophysical applications~~.

2.1 Bayesian ~~Inference Framework for Seismic Data~~Formulation of the Inverse Problem

Within the Bayesian ~~inversion aims~~paradigm, the solution to ~~characterize~~the inverse problem is expressed probabilistically through the posterior distribution ~~of unknown parameters given~~(Tarantola, 2005; Bosch et al., 2007) $p(\boldsymbol{m}|\boldsymbol{d}_{obs})$, which combines prior knowledge with information from the observed ~~seismic~~ data. ~~This requires specifying two key components~~Applying Bayes' theorem, we have:

~~Prior probability density $\rho(\boldsymbol{m})$: which encodes our confidence in~~

$$p(\boldsymbol{m}|\boldsymbol{d}_{obs}) = \frac{p(\boldsymbol{d}_{obs}|\boldsymbol{m})p(\boldsymbol{m})}{p(\boldsymbol{d}_{obs})} \propto cL(\boldsymbol{m})\rho(\boldsymbol{m}), \tag{2}$$

where:

- $\rho(\boldsymbol{m})$ is the ~~subsurface parameters before incorporating~~prior distribution, representing geological knowledge independent of the seismic data.
- $L(\boldsymbol{m})$~~Likelihood~~ is the likelihood function ~~$L(\boldsymbol{m})$: which quantifies~~, quantifying the probability of observing the data given a ~~candidate subsurface model, (Izzatullah et al., 2021).~~specific model.
- ~~Posterior probability distribution: Bayes' theorem combines these two elements as:~~

~~$$\sigma(\boldsymbol{m}) = cL(\boldsymbol{m})\rho(\boldsymbol{m}) \tag{2}$$~~

- ~~where $\sigma(\boldsymbol{m})$ denote the posterior probability density,~~ c is a normalization constant~~, $L(\boldsymbol{m})$ is the likelihood,~~
- ~~$\rho(m) \sim N$~~And $p(\boldsymbol{d}_{obs}) = \int p(\boldsymbol{d}_{obs}|\boldsymbol{m})p(\boldsymbol{m})\, d\boldsymbol{m}$

2.1.1 Prior Distribution Specification

The prior distribution encodes constraints on the model parameters based on independent information such as well logs, regional geology, and rock physics relationships. We employ a Gaussian prior distribution $\rho(\boldsymbol{m}) \sim \mathcal{N}(\boldsymbol{m}_{prior}, \boldsymbol{C}_{\boldsymbol{m}})$ ~~is the prior probability density~~ (Bosch, 2004~~),~~; Buland and Omre, 2003):

$$\rho(\boldsymbol{m}) = c_1 \exp\left[-\frac{1}{2}(\boldsymbol{m} - \boldsymbol{m}_{prior})^T \boldsymbol{C}_m^{-1}(\boldsymbol{m} - \boldsymbol{m}_{prior})\right], \tag{3}$$

where ~~N~~$\mathcal{N}$ denotes a Gaussian probability distribution, with mean $\boldsymbol{m}_{prior}$ ~~and~~ $\in \mathbb{R}^{N_m}$, typically derived from low-frequency interpolation of well data or seismic velocities, covariance matrix given by ~~C_m~~$\boldsymbol{C}_m \in \mathbb{R}^{N_m \times N_m}$, which imposes spatial correlation and smoothness constraints, effectively regularizing the ill-posed inverse problem and c_1 an appropriate normalization constant.

~~In this paper, we will focus on~~**2.1.2 Likelihood Function and Data Errors**

Assuming Gaussian-distributed errors in Eq. (1) with zero mean and covariance matrix $\boldsymbol{C}_d$, the likelihood function takes the form:

$$L(\boldsymbol{m}) = \exp\left[-\frac{1}{2}\left(\boldsymbol{d}_{obs} - g(\boldsymbol{m})\right)^T \boldsymbol{C}_d^{-1}\left(\boldsymbol{d}_{obs} - g(\boldsymbol{m})\right)\right], \quad (4)$$

The likelihood function assumes Gaussian-distributed errors, consistent with standard formulations in Bayesian seismic inversion (Tarantola, 2005; Izzatullah et al., 2021). $g: \boldsymbol{m} \rightarrow \boldsymbol{d}_{obs}$, being the function solving the seismic forward problem and the data covariance matrix $\boldsymbol{C}_d$ accounts for measurement noise and uncertainties in the forward model. While often assumed diagonal (uncorrelated errors), it can incorporate more complex error structures when available.

2.1.3 Posterior Distribution and MCMC Sampling

Combining Eqs. (3) and (4), the posterior ~~probability~~ distribution~~, which~~ can be expressed ~~mathematically as~~through the negative log-posterior (misfit function) $S(\mathbf{m})$:

$$S(\boldsymbol{m}) = -\log p(\boldsymbol{m}|\boldsymbol{d}_{obs}) = \frac{1}{2}\left(\boldsymbol{d}_{obs} - g(\boldsymbol{m})\right)^T \boldsymbol{C}_d^{-1}\left(\boldsymbol{d}_{obs} - g(\boldsymbol{m})\right) + \frac{1}{2}\left(\boldsymbol{m} - \boldsymbol{m}_{prior}\right)^T \boldsymbol{C}_m^{-1}\left(\boldsymbol{m} - \boldsymbol{m}_{prior}\right), \quad (5)$$

Thus,

$$p(\boldsymbol{m}|\boldsymbol{d}_{obs}) \propto \exp[-S(\boldsymbol{m})], \quad (6)$$

In practice, the posterior distribution defined above is analytically intractable. Markov Chain Monte Carlo (MCMC) methods are therefore employed to generate samples $\{\boldsymbol{m}^{(i)}\}_{i=1}^{N}$ from Eq. (6), enabling comprehensive uncertainty quantification through estimation of posterior means, variances, and marginal distributions (Robert and Casella, 2004; Neal, 2012; Gebraad et al., 2020).

2.2 Model Parameterization and Forward Modelling

2.2.1 Logarithmic Parameterization

To ensure positivity of physical parameters ($V_P > 0, V_S > 0, \rho > 0$) and improve numerical stability, we adopt a logarithmic parameterization. For a 1D model with N layers, the parameter vector is (Buland and Omre, 2003):

$$\boldsymbol{m} = \left[\ln(V_{P,1}), \ldots, \ln(V_{P,N}), \ln(V_{S,1}), \ldots, \ln(V_{S,N}), \ln(\rho_1), \ldots, \ln(\rho_N)\right]^T, \quad (7)$$

This transformation also aligns naturally with the logarithmic form of the Aki-Richards reflectivity approximation used in our forward model.

2.2.1 AVO Forward Modeling Formulation

The forward operator $g(\boldsymbol{m})$ implements the Aki-Richards linearized approximation for PP reflectivity. The AVO method was created in the early 1980s to analyze the amplitudes of seismic CMP gathers as a function of angle to find hydrocarbons. The Aki-Richards equation (Aki and Richards, 2002) is the foundation of AVO analysis. The original form of the equation can be rewritten for a weak-contrast interface to give (Buland and Omre, 2003; Niu et al., 2020):

$$R_{pp}(\theta) = c_1(\theta)\frac{\Delta V_p}{\overline{V_p}} + c_2(\theta)\frac{\Delta V_s}{\overline{V_s}} + c_3(\theta)\frac{\Delta \rho}{\overline{\rho}}, \tag{8}$$

where

$$c_1(\theta) = \frac{1}{2}(1 + tan^2\theta), \tag{9}$$

$$c_2(\theta) = -4\frac{\overline{V_s}}{\overline{V_p}}sin^2\theta, \tag{10}$$

$$c_3(\theta) = \frac{1}{2}\left(1 - 4\frac{\overline{V_s}}{\overline{V_p}}sin^2\theta\right), \tag{11}$$

In equations (8 - 11), the incident angle θ is the angle at which a wave reflected off a surface. V_p, V_s and ρ represent the velocities of P-waves, S-waves, and the density of a material, respectively. ΔV_p, ΔV_s and $\Delta\rho$ are the changes in V_p, V_s and ρ across a reflective interface. $\overline{V_p}$, $\overline{V_s}$ and $\overline{\rho}$ are the average values of V_p, V_s and ρ, respectively.

To obtain the seismic trace for a certain theta angle we can use the approximation for small reflectivity (Russell et al., 2006),

$$T(\theta) = \frac{1}{2}c_1 W(\theta) D L_{V_p} + \frac{1}{2}c_2 W(\theta) D L_{V_s} + \frac{1}{2}c_3 W(\theta) D L_{\rho}, \tag{12}$$

where $L_{V_p} = ln(V_p)$, $L_{V_s} = ln(V_s)$, $L_{\rho} = ln(\rho)$, W is the wavelet matrix and D is the derivative matrix. Equation (12) can be implemented in matrix form as

$$\begin{bmatrix} T(\theta_1) \\ T(\theta_2) \\ \vdots \\ T(\theta_n) \end{bmatrix} = \frac{1}{2}\begin{bmatrix} c_1W(\theta_1)D & c_2W(\theta_1)D & c_3W(\theta_1)D \\ c_1W(\theta_2)D & c_2W(\theta_2)D & c_3W(\theta_2)D \\ \vdots & \vdots & \vdots \\ c_1W(\theta_n)D & c_2W(\theta_n)D & c_3W(\theta_n)D \end{bmatrix}\begin{bmatrix} L_{V_p} \\ L_{V_s} \\ L_{\rho} \end{bmatrix} \tag{13}$$

~~$\sigma(\boldsymbol{m}) = ce^{-s}$,~~

In discrete matrix form, for multiple angles $\{\theta_1, \theta_2, \dots, \theta_n\}$, the complete forward modeling operation can be expressed compactly as:

$$\boldsymbol{d}_{syn} = g(\boldsymbol{m}) = \boldsymbol{G}\boldsymbol{m}, \tag{14}$$

where:

~~$$S(\boldsymbol{m}) = \frac{1}{2}\left(\boldsymbol{d}_{obs} - g(\boldsymbol{m})\right)^T \boldsymbol{C}_d^{-1}\left(\boldsymbol{d}_{obs} - g(\boldsymbol{m})\right) + \frac{1}{2}\left(\boldsymbol{m} - \boldsymbol{m}_{prior}\right)^T \boldsymbol{C}_m^{-1}\left(\boldsymbol{m} - \boldsymbol{m}_{prior}\right), \quad (4)$$~~

~~$g: \boldsymbol{m} \rightarrow \boldsymbol{d}_{obs}$,~~ being $\boldsymbol{d}_{syn} = \begin{bmatrix} T(\theta_1) \\ T(\theta_2) \\ \vdots \\ T(\theta_n) \end{bmatrix}$; $\quad \boldsymbol{G} = \frac{1}{2}\begin{bmatrix} c_1 W(\theta_1) D & c_2 W(\theta_1) D & c_3 W(\theta_1) D \\ c_1 W(\theta_2) D & c_2 W(\theta_2) D & c_3 W(\theta_2) D \\ \vdots & \vdots & \vdots \\ c_1 W(\theta_n) D & c_2 W(\theta_n) D & c_3 W(\theta_n) D \end{bmatrix}$ (15)

and $\boldsymbol{m}$ was defined in equation (7). For how to construct W and D, known as the ~~function solving~~ wavelet matrix and the derivative matrix respectively, the reader is advised to review Hampson et al., (2005).

2.3 Inference Workflow and Uncertainty Quantification

The complete Bayesian inversion workflow implemented in this study is visually summarized in Figure 1. The process integrates prior geological knowledge, seismic data, physical modelling, and statistical sampling to produce probabilistic estimates of subsurface elastic properties.

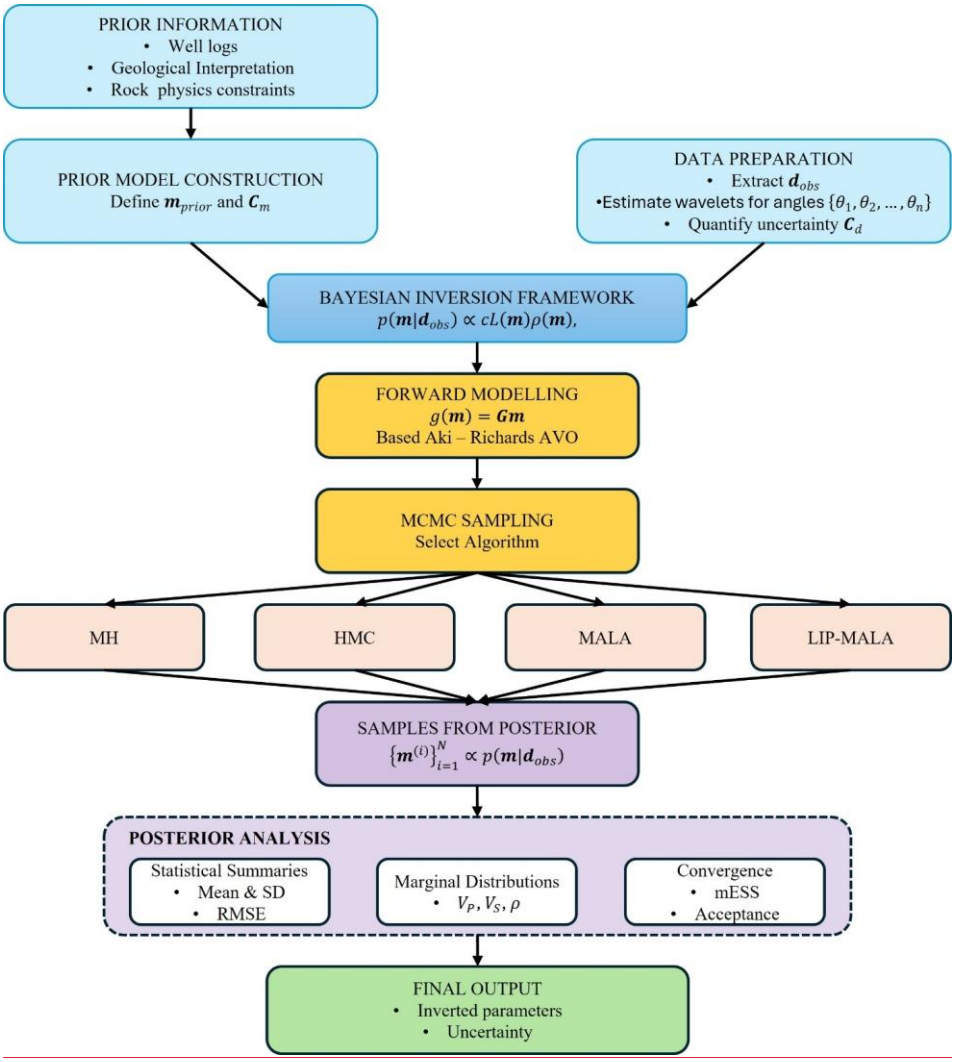

PRIOR INFORMATION
• Well logs
• Geological Interpretation
• Rock physics constraints
PRIOR MODEL CONSTRUCTION
Define $\boldsymbol{m}_{prior}$ and $\boldsymbol{C}_m$
DATA PREPARATION
• Extract $\boldsymbol{d}_{obs}$
•Estimate wavelets for angles $\{\theta_1, \theta_2, \ldots, \theta_n\}$
• Quantify uncertainty $\boldsymbol{C}_d$
BAYESIAN INVERSION FRAMEWORK
$p(\boldsymbol{m}|\boldsymbol{d}_{obs}) \propto cL(\boldsymbol{m})\rho(\boldsymbol{m})$,
FORWARD MODELLING
$g(\boldsymbol{m}) = \boldsymbol{G}\boldsymbol{m}$
Based Aki – Richards AVO
MCMC SAMPLING
Select Algorithm
MH
HMC
MALA
LIP-MALA
SAMPLES FROM POSTERIOR
$\{\boldsymbol{m}^{(i)}\}_{i=1}^{N} \propto p(\boldsymbol{m}|\boldsymbol{d}_{obs})$
POSTERIOR ANALYSIS
Statistical Summaries
• Mean & SD
• RMSE
Marginal Distributions
• V_P, V_S, ρ
Convergence
• mESS
• Acceptance
FINAL OUTPUT
• Inverted parameters
• Uncertainty

Figure 1: Bayesian seismic inversion workflow implemented in this study, integrating prior information, seismic data, forward ~~problem, $\boldsymbol{C_d}$ being the data~~ **modelling, MCMC sampling, and posterior analysis to produce probabilistic estimates of elastic parameters with quantified uncertainties.**

Bayesian inversion process implemented in this study follows a systematic workflow:

1. Prior Model Construction: Define $\boldsymbol{m}_{prior}$ and $\boldsymbol{C}_m$ using available well-log data, geological interpretation, and spatial correlation models.
2. Data Preparation and Uncertainty Quantification: Extract angle-stacked seismic traces $\boldsymbol{d}_{obs}$, estimate corresponding wavelets for multiple angles $\{\theta_1, \theta_2, \dots, \theta_n\}$, and characterize data uncertainties through the covariance matrix ~~that describes second-order statistics on the data uncertainties, and $\boldsymbol{C_m}$ an appropriate covariance matrix describing variability and correlation between parameters of $\boldsymbol{C}_d$~~.
3. Forward Operator Assembly: Construct the ~~medium and $\boldsymbol{m}_{prior}$ is a prior~~ matrix $\boldsymbol{G}$ according to Eqs. (8)-(15) to implement the forward modelling operator $g(\boldsymbol{m})$.
4. MCMC Sampling: Apply one of the four sampling algorithms (MH, HMC, MALA, or Lip-MALA) to generate a Markov chain of model~~.~~ samples from the posterior distribution $p(\boldsymbol{m}|\boldsymbol{d}_{obs})$.
5. Posterior Analysis: Compute statistical summaries (posterior means, standard deviations, correlation coefficients, RMSE) and marginal probability distributions from the sampled models to assess the inverted elastic parameters and their associated uncertainties.

This comprehensive framework provides a rigorous approach to seismic inversion where solutions are characterized not as single models but as probability distributions that fully quantify uncertainty, directly addressing the ill-posed nature of the inverse problem.

~~In practice, the posterior distribution defined above is analytically intractable. Therefore, it is approximated numerically using samples $\{\boldsymbol{m}_i\}_{i=1,\dots,N}$ generated by MCMC algorithms (Metropolis et al., 1953; Hastings, 1970; Estévez et al., 2012; Sanchez et al., 2016).~~

3 Theoretical Background for Metropolis-Hastings, Hamiltonian Monte Carlo and Langevin Diffusion

This section presents the theoretical foundations of the four MCMC algorithms compared in this study. A visual summary of their computational workflows and key characteristics is provided in Figure 2, followed by detailed mathematical descriptions and implementation considerations.

3.1 Algorithm Overview and Comparative Framework

Figure 2 provides a schematic comparative view of the four MCMC algorithms, highlighting their proposal mechanisms, gradient requirements, and tuning strategies. All methods share the common goal of generating samples from the posterior distribution $p(\boldsymbol{m}|\boldsymbol{d}_{obs})$ but employ different strategies for proposing and accepting new states.

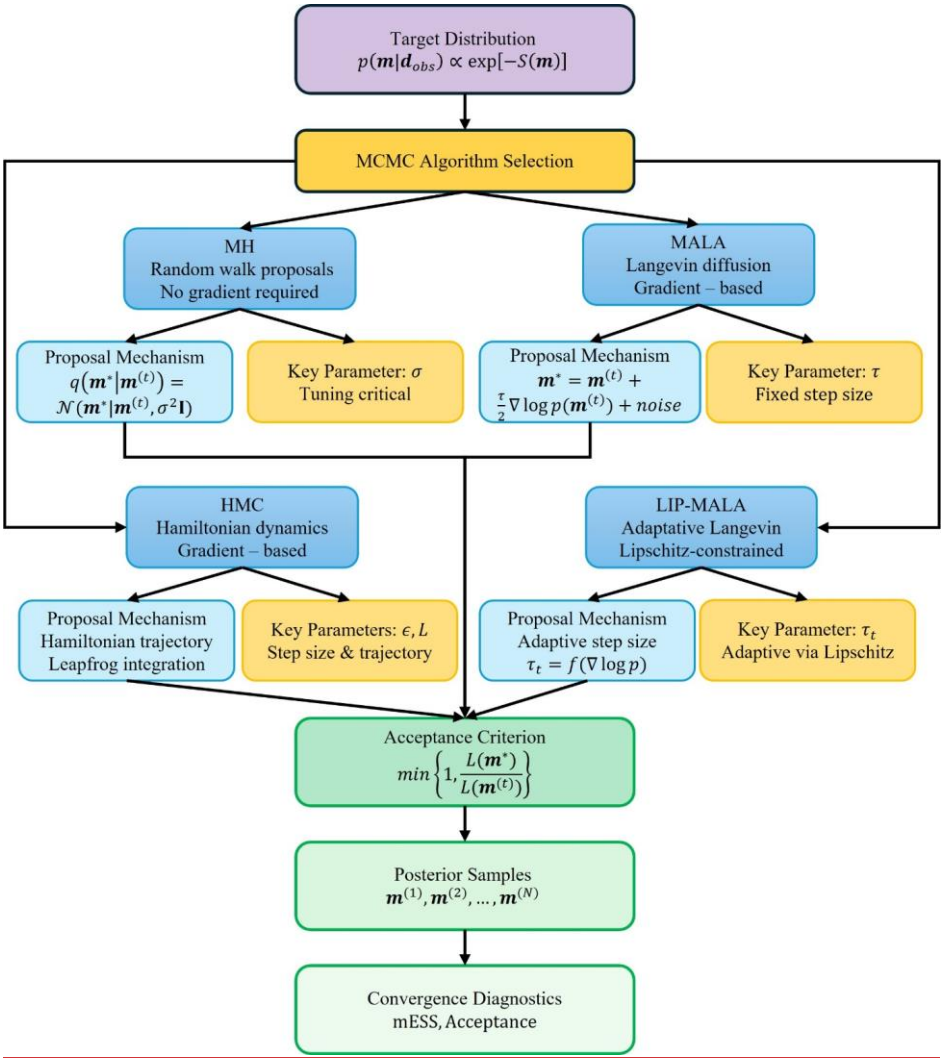


Figure 2: Comparative schematic of MCMC algorithms evaluated in this study. All methods generate samples from the posterior distribution but differ in proposal mechanisms, gradient requirements, and parameter tuning strategies.

3.2 Metropolis–Hastings ~~(MH)~~Algorithm

MH algorithm is one of the most widely used MCMC methods for sampling from complex probability distributions. Originally proposed by Metropolis et al. (1953) and generalized by Hastings (1970), MH defines a transition probability that ensures ergodicity, detailed balance, and reversibility of the chain (Chib and Greenberg, 1995).

~~**The algorithm generates a sequence of values $\boldsymbol{m}_t$ forming a Markov chain that approximates**~~**3.2.1 Mathematical Formulation**

Given the ~~target posterior density $\sigma(\boldsymbol{m})$. The~~ current state $\boldsymbol{m}^{(t)}$ at iteration t, MH proposes a new ~~value $\boldsymbol{m}_t$~~candidate $\boldsymbol{m}^*$ from a proposal distribution $q(\boldsymbol{m}^*|\boldsymbol{m}^{(t)})$. The candidate is ~~obtained according~~accepted with probability:

$$\alpha_{MH} = \min\left\{1, \frac{L(\boldsymbol{m}^*)}{L(\boldsymbol{m}^{(t)})} \cdot \frac{q(\boldsymbol{m}^{(t)}|\boldsymbol{m}^*)}{q(\boldsymbol{m}^*|\boldsymbol{m}^{(t)})}\right\}, \tag{16}$$

where $L(\cdot)$ represents the likelihood function (Eq. 4). For symmetric proposal distributions where $q(\boldsymbol{m}^*|\boldsymbol{m}^{(t)}) = q(\boldsymbol{m}^{(t)}|\boldsymbol{m}^*)$, this simplifies to the Metropolis ~~transition rule is~~ acceptance ratio.

3.2.2 Implementation Details

In our seismic inversion context, we employ a Gaussian random walk proposal:

$$q(\boldsymbol{m}^*|\boldsymbol{m}^{(t)}) = \mathcal{N}(\boldsymbol{m}^*|\boldsymbol{m}^{(t)}, \sigma^2\mathbf{I}), \tag{17}$$

Where σ is a tuning parameter controlling the proposal step size. The algorithm proceeds as follows:

Step 1. Initialization**:** ~~Choose a proposal function $\boldsymbol{q}(\boldsymbol{m}|\boldsymbol{m}_{t-1})$~~ Start with an initial ~~state $\boldsymbol{m}$ and the previous value $\boldsymbol{m}_{t-1}$.~~ model $\boldsymbol{m}^{(0)} = \boldsymbol{m}_{prior}$ and set $t = 0$.

Step 2. Propose a new state**:** ~~At iteration $\boldsymbol{t}$, a candidate values $\tilde{\boldsymbol{m}}$ is generated from the proposal probability density $\tilde{\boldsymbol{m}} \sim \boldsymbol{q}(\tilde{\boldsymbol{m}}|\boldsymbol{m}_{t-1})$, and generate~~ Generate $\boldsymbol{m}^* \sim \mathcal{N}(\boldsymbol{m}^{(t)}, \sigma^2\mathbf{I})$.

Step 3. Compute Acceptance: Calculate α_{MH} using Eq. (16).

~~Step 2.~~Step 4. Accept/Reject: Draw $u \sim$ ~~$\sim$~~ $Uniform(0,1)$~~, where $\boldsymbol{q}(\tilde{\boldsymbol{m}}|\boldsymbol{m}_{t-1}) = \boldsymbol{q}(\boldsymbol{m}_{t-1}|\tilde{\boldsymbol{m}})$, is a symmetric probability distribution.~~. If $u \leq \alpha_{MH}$, set $\boldsymbol{m}^{(t+1)} = \boldsymbol{m}^*$; otherwise $\boldsymbol{m}^{(t+1)} = \boldsymbol{m}^{(t)}$.

~~Step 3. Compute the acceptance probability:~~

~~$$p_{accept} = min\left(1, \frac{L(\tilde{\boldsymbol{m}})q(\tilde{\boldsymbol{m}}|\boldsymbol{m}_{t-1})}{L(\boldsymbol{m})q(\boldsymbol{m}_{t-1}|\tilde{\boldsymbol{m}})}\right)$$~~ ~~(5)~~

~~Step 4. If $u \leq p_{accept}$, accept $\tilde{\boldsymbol{m}}$ and set $\boldsymbol{m}_t = \tilde{\boldsymbol{m}}$. Otherwise reject $\tilde{\boldsymbol{m}}$ and set $\boldsymbol{m}_t = \boldsymbol{m}_{(t-1)}$.~~

~~Iteratively repeating the MH rule generates a chain that converges to a sample of the target probability density.~~

Step 5. Iterate: Increment t and repeat from step 2 until convergence.

3.2.3 Tuning Considerations and Limitations

The ~~efficiency~~performance of MH ~~strongly~~critically depends on the choice of ~~proposal distribution (e.g., random walk, independent, or symmetric MH). Calibration typically involves adjusting parameters such as step size~~ σ. Optimal scaling theory suggests aiming for acceptance rates of approximately 23.4% for random walk proposals in high dimensions (Roberts et al., 1997). In our implementation, $\boldsymbol{\sigma}$ was tuned during a 1000-iteration burn-in period to achieve this target.

While MH is conceptually simple and requires no gradient computations, its random walk behavior leads to ~~maximize acceptance rates and~~slow exploration of high-dimensional parameter spaces. The autocorrelation between samples remains high, requiring longer chains to achieve effective sample ~~size (Robert, 2016).~~sizes comparable to gradient-based methods.

3.~~2~~3 Hamiltonian Monte Carlo (HMC)

HMC is a sampling algorithm that was first introduced in molecular dynamics (Duane et al., 1987) and later adapted to Bayesian inference problems (Neal, 2012~~).~~, Fichtner and Zunino, 2019). It is particularly effective for high-dimensional posteriors when gradient information of the $\sigma(\boldsymbol{m})$ is available and it is more efficient than standard MH. ~~However, the cost of generating independent samples with HMC grows as $O(n^{5/4})$ (Neal, 2012) while with MH grows as $O(n^2)$ (Creutz, 1988) where n is the dimension of the model~~ It introduces an auxiliary momentum variable to simulate Hamiltonian dynamics in an augmented parameter space~~. Thus, HMC requires more computation~~.

3.3.1 Hamiltonian Dynamics Formulation

HMC is an MCMC algorithm that uses classical Hamiltonian mechanics (Landau and Lifshitz, 1976) to sample from an arbitrary n-dimensional probability density function (PDF~~) $\boldsymbol{p}(\boldsymbol{m}) = \sigma(\boldsymbol{m})$ (This notation is adopted in order to maintain consistency with the convention used in the field of Physics). HMC regards the current state $\boldsymbol{m}$ of the Markov chain as the location of a physical particle in n-dimensional space M. The particle moves under the influence of a potential energy, U, which is defined as the negative logarithm of the PDF (Gebraad et al., 2020):~~). HMC introduces momentum variables $\boldsymbol{p} \in \mathbb{R}^{N_m}$ and samples from the joint distribution:

$$p(\boldsymbol{m}, \boldsymbol{p}) = p(\boldsymbol{m})p(\boldsymbol{p}) \propto \exp(-U(\boldsymbol{m}) - K(\boldsymbol{p})) \tag{18}$$

where $U(\boldsymbol{m}) = -log(\boldsymbol{p}(\boldsymbol{m}))$ ~~(6)~~

~~If~~ $(p(\boldsymbol{m}))$, which is defined as the ~~probability density function $\boldsymbol{p}(\boldsymbol{m})$~~negative logarithm of the ~~subsurface model parameters is Gaussian, then the potential energy $U(\boldsymbol{m})$ of the system is equal to the least squares misfit $S(\boldsymbol{m})$, up to an additive constant. To make the system physically complete, we need to add momentum variables $\boldsymbol{p}$ (which is the same~~ PDF~~)~~ (Gebraad et al.,

2020), and mass matrices for each dimension of the model parameter space. The momentum variables represent the velocity of the Markov chain as it moves through the parameter space, and the mass matrix $\boldsymbol{M}$ of dimension $n \times n$ represents the resistance to change. The $K(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p}$ is the kinetic energy of the system is defined by the momenta and thewith mass matrix as $\boldsymbol{M}$.

$$K(\boldsymbol{p}) = \frac{1}{2}\boldsymbol{p}^T\boldsymbol{M}^{-1}\boldsymbol{p} \quad (7)$$

The HMC algorithm uses a random momentum $\boldsymbol{p}$, drawn from a multivariate Gaussian distribution with covariance matrix $\boldsymbol{M}$. The potential energy of the system depends on the location, and the kinetic energy depends on the momentum. The total energy of the system, also known as the Hamiltonian, is the sum of potential and kinetic energies,

$H(\boldsymbol{m},\boldsymbol{p}) =$ The Hamiltonian equations of motion are:

$$U(\boldsymbol{m}) + K(\boldsymbol{p}) \quad (8)$$

Hamilton's equations

$$\frac{d\boldsymbol{m}}{d\tau} = \frac{\partial H}{\partial \boldsymbol{p}}, \quad and = \boldsymbol{M}^{-1}\boldsymbol{p}, \quad \frac{d\boldsymbol{p}}{d\tau} = -\frac{\partial H}{\partial \boldsymbol{m}} = -\nabla U(\boldsymbol{m})$$

$$= \frac{-\partial H}{\partial \boldsymbol{m}} \quad (9) \qquad (19)$$

where $H(\boldsymbol{m},\boldsymbol{p}) = U(\boldsymbol{m}) + K(\boldsymbol{p})$ Making an analogy with the physical problem, we want to find how the particle's position changes over time, as represented by the artificial time variable τ. Hamilton's equations tell us how the position and momentum of a particle change over time, but they can be complicated. We can simplify them by using the fact that the kinetic energy of a particle depends only on its momentum and its potential energy depends only on its position,

$$\frac{d\boldsymbol{m}}{d\tau} = \boldsymbol{M}^{-1}\boldsymbol{p}, \quad and \quad \frac{d\boldsymbol{p}}{d\tau} = \frac{-\partial U}{\partial \boldsymbol{m}} \quad (10)$$

In HMC, the model parameters $\boldsymbol{m}$ and their moment $\boldsymbol{p}$ are represented as a state. It then evolves the state $(\boldsymbol{m},\boldsymbol{p})$ over time τ using Hamiltonian dynamics. This generates a distribution of the possible states of the system with new position $\widetilde{\boldsymbol{m}}$, momentum $\widetilde{\boldsymbol{p}}$, potential energy $\widetilde{U}$, and kinetic energy $\widetilde{K}$, which is a sample of the joint momentum and model space. Since we are only interested in the model parameters, we marginalize over the momenta to obtain a sample of the posterior distribution of the model parameters. This results in samples from the posterior distribution.

$\boldsymbol{p}(\boldsymbol{m}) = exp(-U(\boldsymbol{m}))$ is the Hamiltonian and τ is fictitious time.

3.3.2 Leapfrog Integration and Algorithm

Since analytical solutions to Eq. (19) are unavailable for nonlinear problems, HMC employs the symplectic leapfrog integrator with step size ϵ and L steps. For our seismic inversion, the algorithm proceeds as (Gebraad et al., 2020):

Step 1. Start with an initial model $\boldsymbol{m}^{(0)} = \boldsymbol{m}_{prior}$, random momenta $\boldsymbol{p}^{(0)} \sim \mathcal{N}(0,\boldsymbol{M})$, and set $t = 0$.

(11)

~~If we could solve Hamilton's equations exactly, we could generate an infinite number of valid samples of the posterior probability distribution of the subsurface model parameters $\boldsymbol{p}(\boldsymbol{m})$. However, Hamilton's equations cannot be solved analytically for nonlinear forward models, so we must use numerical integration. Suitable integrators for numerical integration are symplectic, which means that they preserve time reversibility, phase space partitioning, and volume (Neal, 2012; Fichtner and Zunino, 2019). However, explicit time stepping schemes do not exactly preserve the Hamiltonian. In this work, we use the leapfrog method, as described in (Neal, 2012). Since the Hamiltonian is not preserved exactly, the leapfrog method introduces a small error into the samples of $\boldsymbol{p}(\boldsymbol{m})$. The Metropolis-Hastings correction step is a way to "fine-tune" the results of numerical integration to make sure that they are as accurate as possible.~~

~~To summarize, samples of the model parameters are generated by starting with a random model $\boldsymbol{m}$ and then following these steps (Gebraad et al., 2020):~~

~~Step 1. Generate random momenta $\boldsymbol{m}$ values from a Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{M}$.~~

Step 2. Evaluate the Hamiltonian ~~H of model $\boldsymbol{m}$, using its momenta $\boldsymbol{p}$~~$H(\boldsymbol{m}^{(t)}, \boldsymbol{p}^{(t)})$ starting with $t = 0$.

~~Step 3. Given the current values of the model parameters $\boldsymbol{m}$ and $\boldsymbol{p}$, and a time step τ, use a numerical integrator to calculate the updated values of $\boldsymbol{m}$ and $\boldsymbol{p}$, $\widetilde{\boldsymbol{m}}$ and $\widetilde{\boldsymbol{p}}$, after a time period of τ.~~

Step 3. Leapfrog integration (repeated L times): Calculate:

$$\boldsymbol{p}^* = \boldsymbol{p}^{(t)} - \frac{\epsilon}{2}\nabla U\left(\boldsymbol{m}^{(t)}\right) \tag{20}$$

$$\boldsymbol{m}^* = \boldsymbol{m}^{(t)} + \epsilon \boldsymbol{p}^* \tag{21}$$

$$\boldsymbol{p}^* = \boldsymbol{p}^* - \frac{\epsilon}{2}\nabla U(\boldsymbol{m}^*) \tag{22}$$

Step 4. Calculate the Hamiltonian ~~$\widetilde{H}$ of the model $\widetilde{\boldsymbol{m}}$ with momenta $\boldsymbol{p}$.~~$H(\boldsymbol{m}^*, \boldsymbol{p}^*)$.

~~Step 5. Permit the suggested change from $\boldsymbol{m}$ to $\widetilde{\boldsymbol{m}}$ to occur with probability.~~

~~$$p_{accept} = min\left(1, exp\left(H - \widetilde{H}\right)\right), \tag{12}$$~~

~~Step 6. If the new state is better than the current state, accept $\widetilde{\boldsymbol{m}}$ and change it to the current state $\boldsymbol{m}$. Otherwise, keep the current state. Then go back to step 1.~~

Step 5. Compute Acceptance: Calculate α_{HMC} using:

$$\alpha_{HMC} = min\left(1, exp\left(H(\boldsymbol{m}^{(t)}, \boldsymbol{p}^{(t)}) - H(\boldsymbol{m}^*, \boldsymbol{p}^*)\right)\right), \tag{23}$$

Step 6. Accept/Reject: Draw $u \sim Uniform(0,1)$. If $u \leq \alpha_{HMC}$, set $\boldsymbol{m}^{(t+1)} = \boldsymbol{m}^*$ and $H\left(\boldsymbol{m}^{(t+1)}, \boldsymbol{p}^{(t+1)}\right) = H(\boldsymbol{m}^*, \boldsymbol{p}^*)$; otherwise $\boldsymbol{m}^{(t+1)} = \boldsymbol{m}^{(t)}$ and $\boldsymbol{p}^{(t+1)} \sim \mathcal{N}(0, \boldsymbol{M})$.

Step 7. Iterate: Increment t and repeat from step 2 until convergence.

The acceptance rate of the leapfrog integration algorithm commonly used in Step 3 is largely influenced by how well it conserves energy in the trajectory. If the time steps are too large or the gradients of the fitting function are incorrectly calculated,

the algorithm will save less energy, and the acceptance rate will decrease. Simply put, the leapfrog integration algorithm works by bouncing model parameters back and forth across the simulated energy landscape. The acceptance rate determines how often the algorithm accepts a new proposed model parameter. If the time steps are too large or the gradients are calculated incorrectly, the algorithm cannot follow the energy landscape accurately and will likely reject the proposed model parameters. This leads to slower convergence and increases computational cost.

3.3 ~~The~~.3 Parameter Tuning for Seismic Inversion

HMC requires careful tuning of two parameters: the step size $\boldsymbol{\epsilon}$ and trajectory length $\boldsymbol{L}$. We employed a dual adaptation strategy:

1. Step size (ϵ): Tuned during burn-in to achieve acceptance rates optimal for HMC.
2. Trajectory length (L): Fixed at $L = 10$ based on preliminary experiments showing adequate exploration without excessive computational cost.

3.4 Langevin ~~dynamics~~Diffusion-Based Algorithms

3.4.1 Langevin Dynamics

Langevin dynamics are a mathematical model of Brownian motion, named after the French physicist Paul Langevin (Lemons and Gythiel, 1997) who developed them in 1908. Langevin dynamics is a simplification of Albert Einstein's approach to Brownian motion, which is based on Newton's second law of motion. The Langevin dynamics for target distribution $\sigma(\boldsymbol{m}_t)$, is a continuous-time stochastic process $(\boldsymbol{m}_t)_{t\geq 0}$ in $\mathbb{R}^{Nm}$ that evolves following the stochastic differential equation (SDE) (Roberts and Stramer, 2002; Nemeth et al., 2016; Izzatullah et al., 2021) and (Sanchez et al ., 2016; Infante et al., 2019~~),~~):

$$~~d\boldsymbol{m}_t = -~~ d\boldsymbol{m}(\tau) = -\boldsymbol{\Sigma}\nabla log[~~\sigma(\boldsymbol{m}_t)~~]~~dt~~\big[p\big(\boldsymbol{m}(\tau)\big)\big]d\tau + \sqrt{2}\boldsymbol{\Sigma}^{\frac{-1}{2}}~~d\boldsymbol{W}_t~~ d\boldsymbol{W}(\tau) \quad (24)$$

where ~~$(\boldsymbol{W}_t)_{t\geq 0}$~~ $\boldsymbol{W}(\tau)$ is a standard n-dimensional Brownian motion, $\boldsymbol{\Sigma}$ is a symmetric positive definite matrix (in this paper we use $\boldsymbol{\Sigma} = \boldsymbol{I}$), $\nabla log[\sigma(\boldsymbol{m}_t)]$ is the drift term of the Brownian particle $\boldsymbol{m}_t$ and $\sigma(\boldsymbol{m}_t)$ is a stationary posterior distribution.

3.~~3.1~~4.2 Metropolis-adjusted Langevin algorithm (MALA)

In the practice, a standard approach is to discretize the equation (13) using the Euler-Maruyama discretization (Stuart et al., 2004) and we obtained the Unadjusted Langevin algorithm (ULA) given by

$$~~\boldsymbol{m} = \widetilde{\boldsymbol{m}} + \tau_\epsilon\boldsymbol{\Sigma}\nabla log[\sigma(\boldsymbol{m}_\epsilon)] + \sqrt{2\tau_\epsilon}\boldsymbol{\Sigma}^{\frac{-1}{2}}\boldsymbol{\epsilon}_\epsilon~~ \boldsymbol{m}^* = \boldsymbol{m}^{(t)} + \tau\boldsymbol{\Sigma}\nabla log\big[p\big(\boldsymbol{m}^{(t)}\big)\big] + \sqrt{2\tau}\boldsymbol{\Sigma}^{\frac{-1}{2}}\boldsymbol{\epsilon}_t,\ \boldsymbol{\epsilon}_t \sim N(0, I_{n\times n}) \quad (~~14~~25)$$

where τ is the step-length for each iteration. ULA is simple in its implementation, yet it introduces a bias, then we need to introduce the acceptance-rejection step through the MH algorithm. By introducing MH algorithm into ULA, we will obtain

the Metropolis-Adjusted Langevin algorithm (MALA), (Izzatullah et al., 2020, Izzatullah et al., 2021). The procedure consists of constructing a Markov chain at each step t, given ~~$\widetilde{\boldsymbol{m}}, \boldsymbol{m}^{(t)}$~~, a new observation ~~$\boldsymbol{m}$~~ is generated from a proposal density ~~$q(\boldsymbol{m}).q(\boldsymbol{m}^*)$~~. The candidate is then accepted with probability ~~p_{accept}~~α_{MALA} given by,

~~$$p_{accept} = min\left(1, \frac{L(\widetilde{\boldsymbol{m}})q(\widetilde{\boldsymbol{m}}|\boldsymbol{m}_{t-1})}{L(\boldsymbol{m})q(\boldsymbol{m}_{t-1}|\widetilde{\boldsymbol{m}})}\right)$$~~ ~~(15)~~

$$\alpha_{MALA} = \min\left\{1, \frac{L(\boldsymbol{m}^*)}{L(\boldsymbol{m}^{(t)})} \cdot \frac{q\left(\boldsymbol{m}^{(t)}|\boldsymbol{m}^*\right)}{q\left(\boldsymbol{m}^*|\boldsymbol{m}^{(t)}\right)}\right\}, \quad (26)$$

In summary, MALA algorithm is obtained as follows:

Step 1. Choose an initial solution ~~$\widetilde{\boldsymbol{m}} = \boldsymbol{m}_{prior}$~~ $\boldsymbol{m}^{(0)} = \boldsymbol{m}_{prior}$ and the discretization step-length τ.

Step 2. Draw $\boldsymbol{\epsilon}_t \sim N(0, I_{n\times n})$ and simulate a new sample from the Langevin ~~diffussion:~~diffusion (Eq. (25)).

~~$$\boldsymbol{m} = \widetilde{\boldsymbol{m}} + \tau_t \boldsymbol{\Sigma} \nabla log[\sigma(\boldsymbol{m}_t)] + \sqrt{2\tau_t}\boldsymbol{\Sigma}^{\frac{-1}{2}}\boldsymbol{\epsilon}_t,$$~~ ~~(16)~~

Step 3. Compute the accept-reject probability α_{MALA} using Eq. (26).

~~$$p_{accept} = min\left(1, \frac{L(\widetilde{\boldsymbol{m}})q(\widetilde{\boldsymbol{m}}|\boldsymbol{m}_{t-1})}{L(\boldsymbol{m})q(\boldsymbol{m}_{t-1}|\widetilde{\boldsymbol{m}})}\right)$$~~ ~~(17)~~

Step 4. Draw u from a uniform distribution $u \sim Uniform(0,1)$, if ~~p_{accept}~~$\alpha_{MALA} < u$ ~~then accept $\boldsymbol{m}_t = \widetilde{\boldsymbol{m}}$, else $\boldsymbol{m}_t = \boldsymbol{m}_{t-1}$~~set $\boldsymbol{m}^{(t+1)} = \boldsymbol{m}^*$; otherwise $\boldsymbol{m}^{(t+1)}$=$\boldsymbol{m}^{(t)}$.

Step 5. Then, repeat this process until the convergence.

~~The main advantage of the MALA algorithm is that high-dimensional density samples are obtained using the gradient of the logarithm of the posterior distribution. The MALA algorithm is a MCMC method that uses simulations from the discretization by the Euler-Maruyama algorithm of an SDE whose target density has a stationary distribution. The algorithm is inspired by stochastic models of molecular dynamics and is a multivariate extension of a Metropolis random walk, including partial derivatives to improve the mixing rate. It is general purpose, has good theoretical properties, in particular, it can scale better to high-dimensional problems than standard MCMC algorithms, geometric convergence is well established, has an acceptance rate between 40-80%. One drawback is that it requires calculating a gradient at each iteration and successively evaluating the objective function.~~

3.4.3~~.2~~ MALA with locally Lipschitz adaptive step size (Lip-MALA)

In the MALA algorithm, it is required to calibrate the step-size τ, because τ must decrease with dimension, n. then τ can be turned such that the MCMC achieve better mixing performance. An extension of ULA and similar in spirit with Stochastic Gradient Langevin Dynamics algorithm proposed by Welling and Teh, (2011) by suppressing the MH acceptance steps. See in (Izzatullah et al., 2021) proposed ULA variant, Lip-MALA, in which the step-length τ is adapted according to the Lipschitz condition,

$$\tau_t = \frac{1}{2}\frac{\|\boldsymbol{m}_{t+1} - \boldsymbol{m}_t\|_2}{\|\nabla log[\sigma(\boldsymbol{m}_{t+1})] - \nabla log[\sigma(\boldsymbol{m}_t)]\|_2}\frac{||\boldsymbol{m}^{(t)} - \boldsymbol{m}^{(t-1)}||_2}{||\nabla log[p(\boldsymbol{m}^{(t)})] - \nabla log[p(\boldsymbol{m}^{(t-1)})]||_2} \quad (1827)$$

The general steps for Lip-MALA MCMC with locally Lipschitz adaptive step size are:

Step 1. Choose an initial solution $\widetilde{\boldsymbol{m}} = \boldsymbol{m}_{prior}\boldsymbol{m}^{(0)} = \boldsymbol{m}_{prior}$, the discretization step-length τ, $\beta_0 = +\infty$ and $L_c = N_m^{-1/3}$.

Step 2. Draw $\boldsymbol{\epsilon}_t \sim N(0, I_{n\times n})$ and simulate a new sample from the Langevin diffussion:diffusion (Eq. (25)).

$$\boldsymbol{m} = \widetilde{\boldsymbol{m}} - \tau_t\boldsymbol{\Sigma}\nabla log[\sigma(\widetilde{\boldsymbol{m}})] + \sqrt{2\tau_t}\boldsymbol{\Sigma}^{\frac{-1}{2}}\boldsymbol{\epsilon}_t, \quad (19)$$

Step 3. Compute the accept-reject probability α_{MALA} using Eq. (26).

$$p_{accept} = min\left(1, \frac{L(\widetilde{\boldsymbol{m}})q(\widetilde{\boldsymbol{m}}|\boldsymbol{m}_{t+1})}{L(\boldsymbol{m})q(\boldsymbol{m}_{t+1}|\widetilde{\boldsymbol{m}})}\right) \quad (20)$$

Step 4. Draw u from a uniform distribution $u \sim Uniform(0,1)$, if $p_{accept} < u$ then accept $\boldsymbol{m}_t = \widetilde{\boldsymbol{m}}$,$\alpha_{MALA} < u$ set $\boldsymbol{m}^{(t+1)} = \boldsymbol{m}^*$, then update,

$$\tau_t = min\left\{\sqrt{1+\beta_{t-1}}\tau_{t-1}, L_c\frac{\|\boldsymbol{m}_{t+1} - \boldsymbol{m}_t\|}{\|\nabla log[\sigma(\boldsymbol{m}_{t+1})] - \nabla log[\sigma(\boldsymbol{m}_t)]\|}\right\} \quad (21)\ \left\{\frac{||\boldsymbol{m}^{(t)} - \boldsymbol{m}^{(t-1)}||_2}{||\nabla log[p(\boldsymbol{m}^{(t)})] - \nabla log[p(\boldsymbol{m}^{(t-1)})]||_2}\right\} \quad (28)$$

$$\beta_t = \frac{\tau_t}{\tau_{t-1}} \quad (2229)$$

Step 4.Step 5. Else reject $\boldsymbol{m}_t = \boldsymbol{m}_{t-1}$.$\boldsymbol{m}^{(t+1)} = \boldsymbol{m}^{(t)}$. Then, repeat this process until the convergence.

The resulting Lip-MALA algorithm follows the same steps as MALA but dynamically adjusts ϵ, improving stability and sampling efficiency in high dimensions.

3.5 Algorithm Comparison and Tuning Considerations

Table 1 summarizes the mainkey characteristics, advantages, and limitations of the methods used in this study, in order to providefour algorithms, supported by theoretical and empirical studies from the reader with a clear and concise reference that facilitates the comparative understanding of their scope and restrictionsliterature.

Table 1: Advantages and limitationsComparative analysis of the sampling methodsMCMC algorithms for seismic inversion

MethodAlgorithm	Advantages Key Mechanism	Limitations Gradient Required	Optimal Acceptance Rate	Tuning Parameters	Computational Cost per Iteration	Best Suited For

Celdas insertadas
Celdas insertadas
Celdas insertadas
Celdas insertadas

MH	~~MH is simple to implement, flexible (can target arbitrary distributions), and it has a better acceptance and rejection rate in high-dimensional spaces than other competing algorithms.~~ Random walk proposal	No	~23.4% (Roberts et al., 1997)	Step size σ	Low (O(N))	~~Convergence may be slow in high~~Low-dimensional ~~or multimodal~~ problems, ~~and poor tuning of the proposal step size can lead to highly correlated samples.~~rapid screening
HMC	~~HMC reduces random walk behavior, explores the posterior efficiently, and performs well in high dimensions.~~Hamiltonian dynamics	Yes	~65-80% (Beskos et al., 2013)	~~It requires computation of gradients and careful tuning of ϵ (step size) and L (trajectory length). Poor tuning reduces acceptance rates and increases~~	High (O(L·N) + gradient)	High-dimensional smooth posteriors

Celdas insertadas

Celdas insertadas

					computational cost.Step size ε, steps L		
MALA		MALA exploits gradient information, improves mixing, and scales better to high-dimensional problems.Langevin diffusion	Yes	~57.4% (Roberts & Rosenthal, 1998)	Step size τ	Requires gradient evaluations at each iteration, which can be computationally costly.Moderate (O(N) + gradient)	Moderately high dimensions with gradients
Lip MALA	Lip-MALA improves sampling efficiency by adapting the step size locally using a Lipschitz condition, which enhances stability and convergence while reducing the need for	Adaptive Langevin	Yes	~50-70% (Izzatullah et al., 2021)	Include the extra computational cost of estimating the Lipschitz constant, reliance on accurate gradient calculations, and potential difficulties when dealing with	Moderate-High (O(N) + gradient + adaptation)	High-dimensional, heterogeneous posteriors

Celdas insertadas

Celdas eliminadas

Celdas insertadas

Celdas insertadas

	~~manual tuning. This makes it more effective in high-dimensional and complex posterior spaces compared to standard MALA~~**Lip-MALA**				~~highly multi-modal posteriors.~~Initial τ, Lipschitz constant		

~~4 Forward modelling: The AVO method~~

~~The AVO method was created in the early 1980s to analyze the amplitudes of seismic CMP gathers as a function of angle to find hydrocarbons. The Aki-Richards equation (Aki and Richards, 2002) is the foundation of AVO analysis. The original form of the equation can be rewritten for a weak-contrast interface to give (Buland and Omre, 2003; Niu et al., 2020):~~

$$R_{pp}(\theta) = c_1(\theta)\frac{\Delta V_p}{\overline{V_p}} + c_2(\theta)\frac{\Delta V_s}{\overline{V_s}} + c_3(\theta)\frac{\Delta \rho}{\overline{\rho}}, \quad (23)$$

~~where~~

$$c_1(\theta) = \frac{1}{2}(1 + tan^2\theta), \quad (24)$$

$$c_2(\theta) = -4\frac{\overline{V_s}}{\overline{V_p}}sin^2\theta, \quad (25)$$

$$c_3(\theta) = \frac{1}{2}\left(1 - 4\frac{\overline{V_s}}{\overline{V_p}}sin^2\theta\right), \quad (26)$$

~~In equations (23 – 26), the incident angle θ is the angle at which a wave reflected off a surface. V_p, V_s and ρ represent the velocities of P-waves, S-waves, and the density of a material, respectively. ΔV_p, ΔV_s and $\Delta\rho$ are the changes in V_p, V_s and ρ across a reflective interface. $\overline{V_p}$, $\overline{V_s}$ and $\overline{\rho}$ are the average values of V_p, V_s and ρ,respectively.~~

~~To obtain the seismic trace for a certain theta angle we can use the approximation for small reflectivity (Russell et al., 2006),~~

$$T(\theta) = \frac{1}{2}c_1 W(\theta) D L_{V_p} + \frac{1}{2}c_2 W(\theta) D L_{V_s} + \frac{1}{2}c_3 W(\theta) D L_{\rho}, \quad (27)$$

where $L_{V_p} = ln(V_p)$, $L_{V_s} = ln(V_s)$, $L_{\rho} = ln(\rho)$, W is the wavelet matrix and D is the derivative matrix. Equation 27 can be implemented in matrix form as

$$\begin{bmatrix} T(\theta_1) \\ T(\theta_2) \\ \vdots \\ T(\theta_n) \end{bmatrix} = \frac{1}{2} \begin{bmatrix} c_1 W(\theta_1) D & c_2 W(\theta_1) D & c_3 W(\theta_1) D \\ c_1 W(\theta_2) D & c_2 W(\theta_2) D & c_3 W(\theta_2) D \\ \vdots & \vdots & \vdots \\ c_1 W(\theta_n) D & c_2 W(\theta_n) D & c_3 W(\theta_n) D \end{bmatrix} \begin{bmatrix} L_{V_p} \\ L_{V_s} \\ L_{\rho} \end{bmatrix} \quad (28)$$

A practical approach to solve equation 28 is to initialize the solution to, $[L_{V_p} \quad L_{V_s} \quad L_{\rho}]^T = [L_{V_{p0}} \quad L_{V_{s0}} \quad L_{\rho_0}]^T$ where $L_{V_{p0}}, L_{V_{s0}}$ and L_{ρ_0} is the prior model for P-wave and S-wave velocities and bulk density respectively, and then to iterate towards a solution using in our case MH, HMC, MALA and Lip-MALA.

53.5.1 Practical Implementation Notes

For our seismic inversion experiments, all algorithms were implemented with the following considerations:

1. Gradient Computation: The gradient $\nabla \log p(\mathrm{m}) = -\nabla S(\mathrm{m})$ was computed analytically, ensuring computational efficiency.
2. Step Size Tuning: Each algorithm's step size parameter was tuned during a preliminary burn-in phase to achieve target acceptance rates.
3. Convergence Diagnostics: We employed diagnostics using Effective Sample Size (ESS) and multivariate ESS (Vats et al., 2019)
4. Computational Resources: All experiments were conducted on a workstation with Intel Core i9-10900K, 64GB RAM, without GPU acceleration. Execution times were measured as wall-clock time for fair comparison.

The following sections present application of these algorithms to synthetic and real seismic data, evaluating their performance in terms of accuracy, uncertainty quantification, and computational efficiency.

4 Results

54.1 Synthetic test

We test our algorithms using noise-free synthetic seismic traces that were obtained from real data of V_p, V_s and ρ for which synthetic seismic traces were generated from the equation 28 for the angles $\theta_1 = 9^o, \theta_2 = 18.5^o$ and $\theta_3 = 27.5^o$and these synthetic seismic traces will be our observed data. We ran the sampling algorithms described in section 3, producing a large chain of realizations, starting from a prior model configuration corresponding to a low frequency model of V_p, V_s and ρ.

Figure ~~1~~3 shows the objective function variation curves for the different sampling algorithms. Each iteration involves randomly perturbing the velocities and density of a subset of layers and recalculation of seismic traces. The vertical axis represents the objective function calculated using Equation ~~4~~5. The horizontal axis shows the number of steps in the Markov chain, each associated with an accepted or rejected perturbation of the velocity and density configuration. The first stage of the chain, associated with the initial configuration and large residues, is called the burn-in or warm-up stage. After subtracting the residuals, the model realizations of velocities and densities satisfactorily explain the seismic data within the data errors. This is called the sampling phase. Realizations produced during the sampling phase are treated as samples from the probability density.

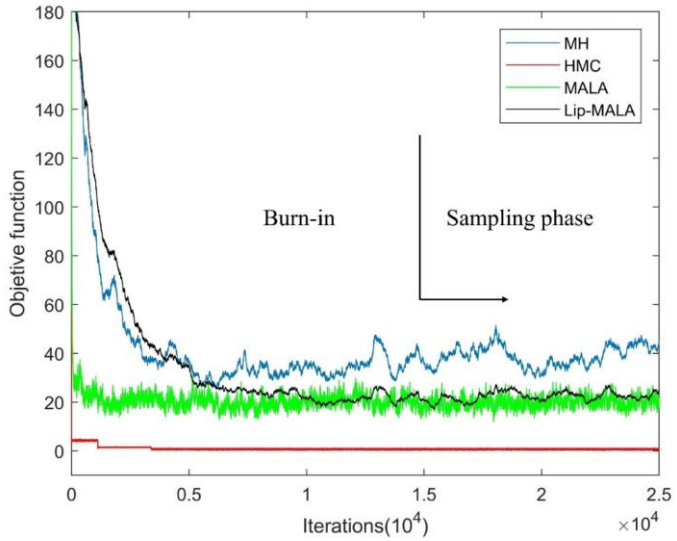


Figure ~~1~~3: Progress with iterations in the MH (blue line), HMC (red line), MALA (green line) and Lip-MALA (black line) sampling algorithms for synthetic test.

The model settings were modified during the sampling phase, but remain within the probability function, as shown in Figure ~~2~~4. Figure ~~3~~5 shows all realizations taken (gray area) in the chain sampling phase for the different algorithms tested in this work, all adjusting the observed seismic data and within the uncertainties of the data. These realizations indicate the features and variability of the velocities and density. Table 2 shows the statistical parameters of mean and standard deviation (SD) which we will compare then with the data obtained from the inference in the different algorithms used.

Table 2: Mean and Standard deviation of elastic parameters used in the synthetic test

Parameter	Mean	Sd

$\boldsymbol{V_p}$ (m/s)	3068.21	278.29
$\boldsymbol{V_s}$ (m/s)	1553.81	240.60
$\boldsymbol{\rho}$ (Kg/m^3)	2263.57	54.68

.

490

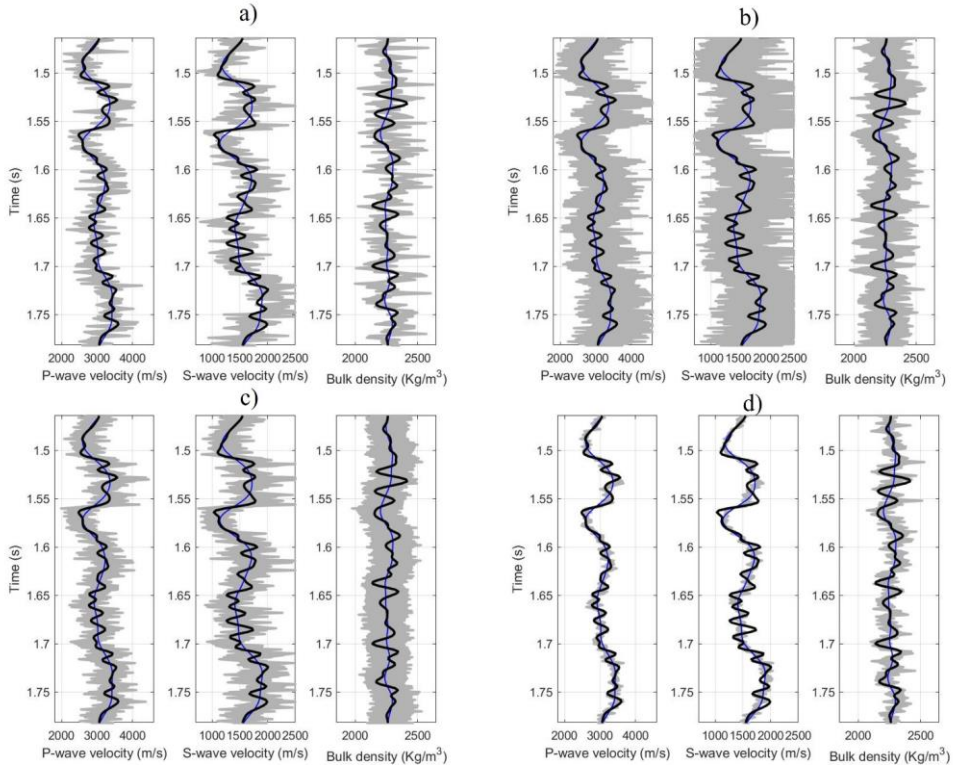


Figure 24: True data (black line), prior model (blue line), and accepted realizations of the model (gray) for the synthetic test where a) MH, b) HMC, c) MALA, and d) Lip-MALA.

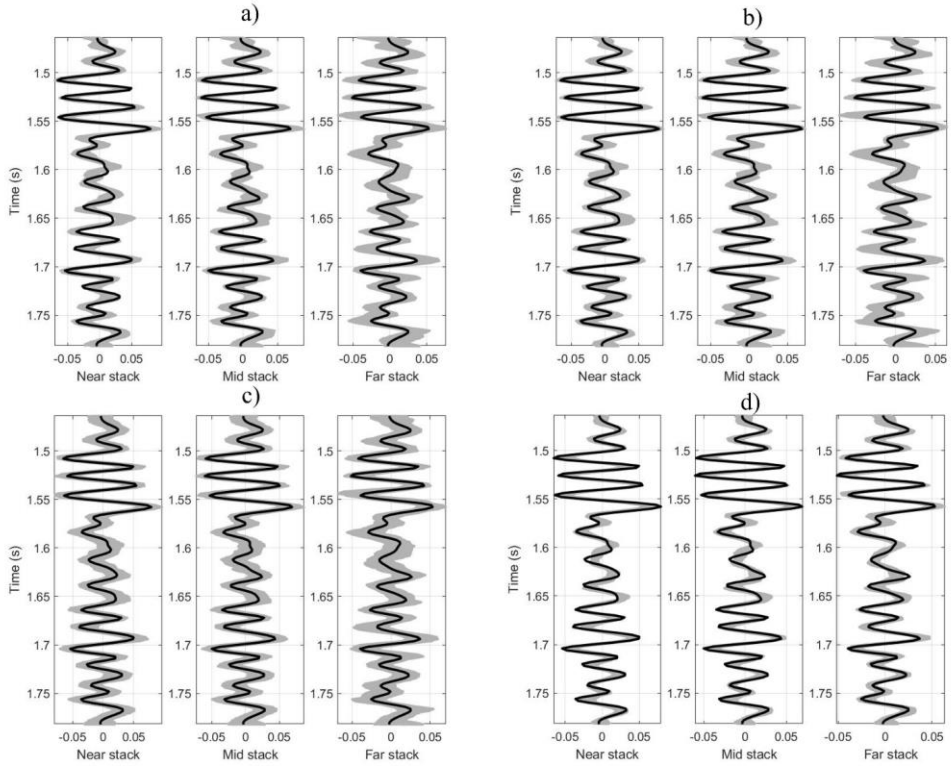


Figure ~~3~~5: Observed seismic data (black line) and seismic traces obtained from the accepted model realizations (gray) for the synthetic test where a) MH, b) HMC, c) MALA, and d) Lip-MALA.

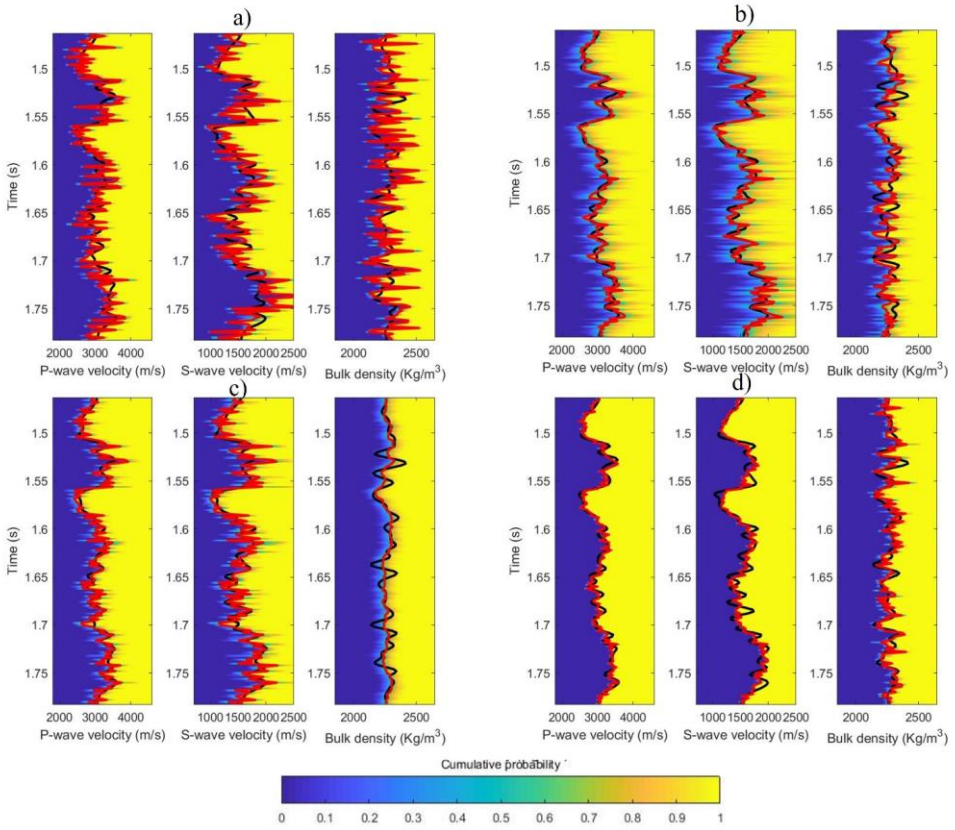


Figure 46: Marginal cumulative probability distributions (color map), true data (black line) and seismic inversion model result (red line) for the synthetic test where a) MH, b) HMC, c) MALA, and d) Lip-MALA.

Our chain sampling phase yielded 10,000 realizations. From these realizations, we calculated the expected values and marginal probabilities of P-wave and S-wave velocities and density as a function of two-way reflection time. These calculations were based on averaging the model performances over the sampling phase. Figure 46 presents the marginal cumulative probability distributions for P and S wave velocities and density, as estimated by the inversion, along with the actual P and S wave velocities and density of the synthetic test. The figure demonstrates the successful prediction of the actual values for all tested

algorithms, accurately identifying the main stratification characterized by high and low velocities and the corresponding high and low density.

Table 3 summarizes the performance of the different algorithms tested in predicting P-wave and S-wave velocities and density. The mean, Standard Deviation (SD), correlation, and Root Mean Squared Error (RMSE) are presented for each parameter. The mean and standard deviation values indicate that the predicted values are closely aligned with the true values. Regarding correlation, MH exhibits the lowest correlation for velocity prediction, while HMC achieves the highest. For density prediction, MH and HMC show correlations below 0.29, while MALA and Lip-MALA achieve correlations above 0.60.

In terms of RMSE, MH demonstrates the highest error for velocity prediction, while HMC achieves the lowest. For density prediction, MH and HMC exhibit errors above 75.75, while MALA and Lip-MALA maintain errors below 51.04.

Table 3: Statistical parameters for the results obtained for algorithms tested for the synthetic test.

Parameter	Mean	Sd	Corr	RMSE
MH				
$\boldsymbol{V_p}$ (m/s)	3058.90	394.79	0.64	302.96
$\boldsymbol{V_s}$ (m/s)	1577.99	359.38	0.64	277.26
$\boldsymbol{\rho}$ (Kg/m^3)	2278.00	112.32	0.29	110.39
HMC				
$\boldsymbol{V_p}$ (m/s)	3075.62	312.57	0.90	135.23
$\boldsymbol{V_s}$ (m/s)	1566.98	271.01	0.90	118.15
$\boldsymbol{\rho}$ (Kg/m^3)	2256.42	61.61	0.16	75.75
MALA				
$\boldsymbol{V_p}$ (m/s)	3051.93	326.28	0.85	174.52
$\boldsymbol{V_s}$ (m/s)	1544.13	277.40	0.80	165.30
$\boldsymbol{\rho}$ (Kg/m^3)	2262.27	29.39	0.68	40.64
Lip-MALA				
$\boldsymbol{V_p}$ (m/s)	3062.09	265.42	0.91	112.94
$\boldsymbol{V_s}$ (m/s)	1552.04	217.04	0.89	110.60
$\boldsymbol{\rho}$ (Kg/m^3)	2266.78	59.00	0.60	51.04

Table 4 presents various performance parameters, including acceptance rate and total execution time. Lip-MALA exhibits the highest acceptance rate, while HMC exhibits the lowest. Conversely, MH boasts the lowest total execution time, while HMC demonstrates the highest.

Table 4: Other parameters for synthetic test.

Method	Acceptance rate (%)	Total execution time (s)
MH	36.50	24.12
HMC	17.53	9356.99
MALA	25.28	694.54
Lip-MALA	38.49	3337.02

Finally, the convergence of the samples was analyzed a posteriori of the unknown parameters (seismic data parameters) m obtained from the different algorithms used. The multivariate effective sample size (mESS) statistic was used. The mESS is a measure that determines the size of an independent and identically distributed sample with the same covariance structure as the sample obtained from an MCMC method for the multivariate case- If we want to know if the chain converges by we can calculate minimum effective sample size (minESS) so that if mESS>minESS we say that the chain converges, if the reader is recommended to review Vats et al. (2019) to delve deeper into the convergence test used in this work. Table 5 shows the summary of mESS and minESS obtained for each method.

Table 5: Convergence test for synthetic data.

Method	mESS	minESS
MH	8150.89	7458
HMC	8561.10	7458
MALA	7472.03	7458
Lip-MALA	8119.88	7458

~~5~~4.2 Application to real data

To demonstrate the effectiveness of the algorithms, we applied them to a real dataset of an oil field in eastern Venezuela. The site is located in a formation dominated by clastic rocks, a type of sedimentary rock characterized by alternating layers of sand and shale. The fluids in the pore spaces of these rocks are brine water and oil, without gas. As a preliminary step, we upscaled the P-wave and S-wave velocities obtained from well log data to the corresponding seismic scale using a bandpass filter. This process ensures that the velocity data is consistent with the frequency range of seismic waves. Table 6 presents the descriptive

statistics, including mean and standard deviation (SD), for the real data. These values will serve as a baseline for comparison with the results obtained from the inference procedures employed by the various algorithms under consideration.

Table 6: Mean and Standard deviation of elastic parameters used for real data.

Parameter	Mean	Sd
$\boldsymbol{V_p}$ (m/s)	2642.92	249.40
$\boldsymbol{V_s}$ (m/s)	1289.86	205.84
$\boldsymbol{\rho}$ (Kg/m^3)	2180.06	111.89

The seismic traces were obtained from partial stacks for the angles θ_1= 19°, θ_2= 24°, and θ_3= 29°. Utilizing V_p, V_s and ρ logs in seismic scale and wavelets were extracted from the partial stacked seismic data using the frequency content of the data, the synthetic trace was generated using equation ~~28.~~(13). The synthetic trace obtained was correlated with observed traces for seismic well tie (see figure ~~5~~7) obtaining a correlation value of 0.55. The sampling algorithms described in section 3 were implemented, generating a large chain of realizations starting from a prior model configuration corresponding to a low-frequency model of V_p, V_s and ρ.

As depicted in figure ~~6~~8, the objective function variation curves for each sampling algorithm are presented. During each iteration, a subset of layers undergoes a random perturbation of their velocities and density, followed by a recalculation of the seismic trace. The objective function, calculated using equation ~~4~~5, is represented on the vertical axis, while the horizontal axis represents the number of steps in the Markov chain. Each step corresponds to an accepted or rejected perturbation of the velocities and density configuration.

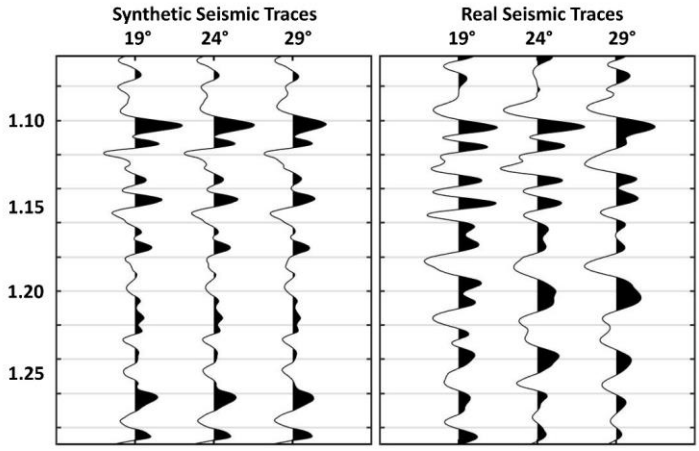
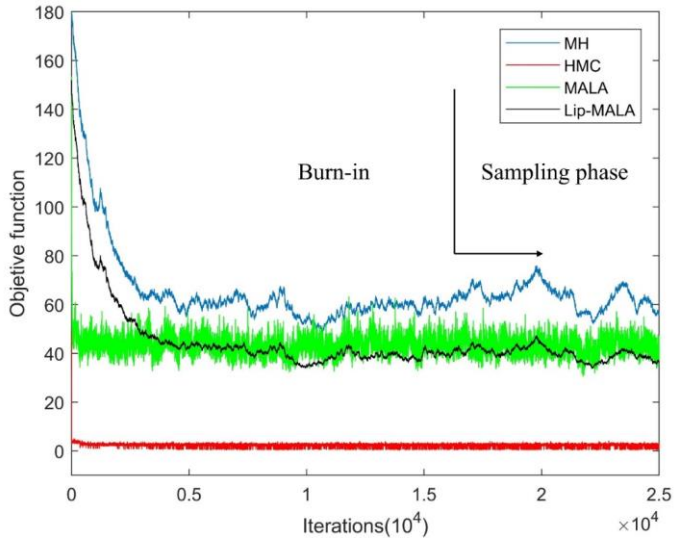


Figure ~~5~~7: Seismic well tie for real data used.



555 **Figure ~~6~~8: Progress with iterations in the MH (blue line), HMC (red line), MALA (green line) and Lip-MALA (black line) sampling algorithms for real data.**

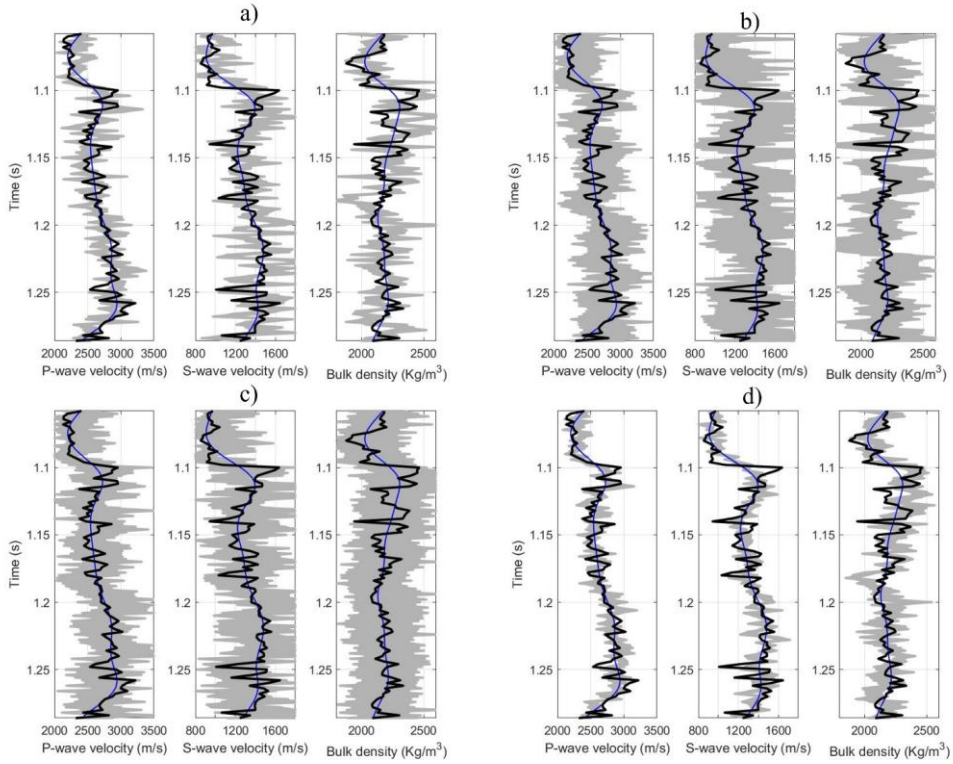


Figure ~~7:~~ 9: True data (black line), prior model (blue line), and accepted realizations of the model (gray) for real data where a) MH, b) HMC, c) MALA, and d) Lip-MALA.

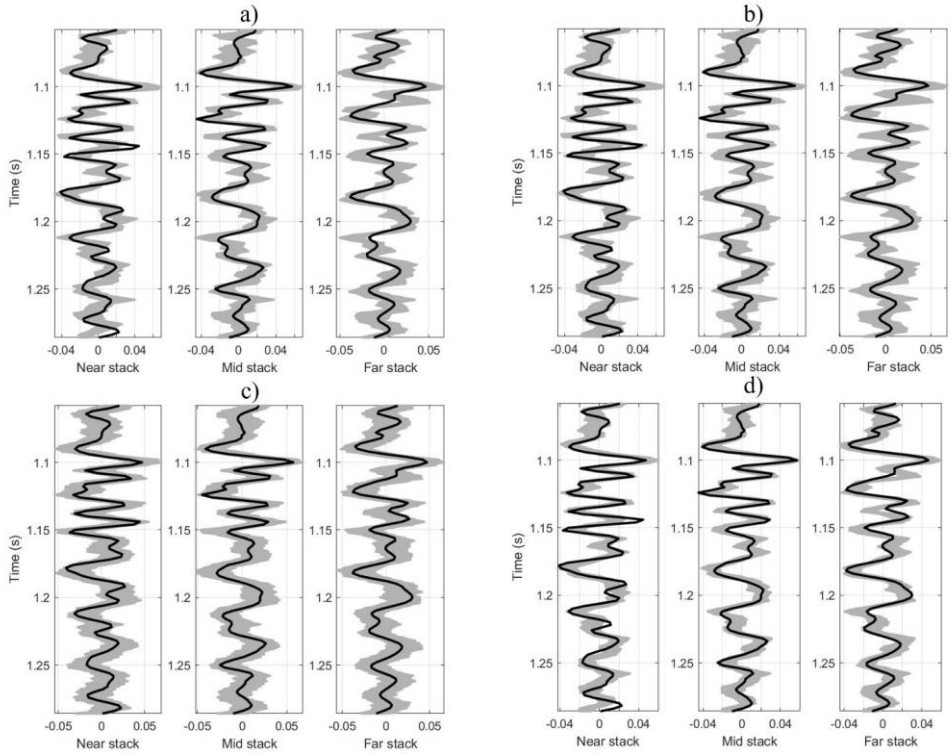


Figure ~~8:~~ 10: Observed seismic data (black line) and seismic traces obtained from the accepted model realizations (gray) for real data where a) MH, b) HMC, c) MALA, and d) Lip-MALA.

565

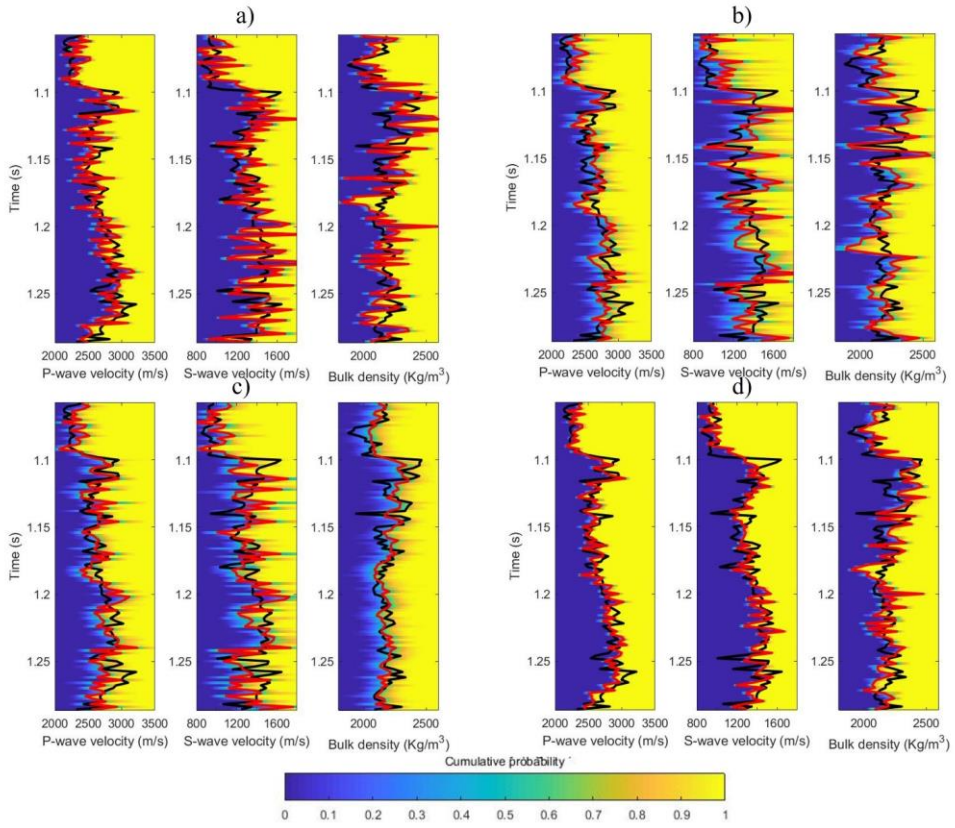


Figure ~~9:~~ 11: Marginal cumulative probability distributions (color map), true data (black line) and seismic inversion model result (red line) for real data where a) MH, b) HMC, c) MALA, and d) Lip-MALA.

The model settings were adjusted during the sampling phase, ensuring they remained within the probability function (Figure ~~7~~9). Figure ~~8~~10 illustrates all realizations sampled (gray area) in the chain sampling phase for the various algorithms tested in this study, all of which align with the observed seismic data and fall within the data's uncertainty bounds. These realizations highlight the characteristics and variability of the velocities and density.

Employing a chain sampling scheme, we generated 9,000 realizations from which we extracted the expected values and marginal probabilities of P-wave and S-wave velocities and density, all as functions of two-way reflection time. These calculations were derived by averaging the model performances across the sampling phase. Figure ~~9~~11 depicts the marginal cumulative

probability distributions for P and S wave velocities and density, as inferred from the inversion process, alongside the actual P and S wave velocities and density of the synthetic test.

Table 7 summarizes the performance of the tested algorithms in predicting P-wave and S-wave velocities and density. The mean, Standard Deviation (SD), correlation, and Root Mean Squared Error (RMSE) are presented for each parameter. The predicted values closely align with the true values as evidenced by the mean and standard deviation values. MH exhibits the lowest correlation for velocity prediction, while Lip-MALA achieves the highest. For density prediction, MH and HMC show correlations below 0.28, while MALA and Lip-MALA achieve correlations above 0.48. MH demonstrates the highest error for velocity prediction, while Lip-MALA achieves the lowest. For density prediction, MH and HMC exhibit errors above 151.41, while MALA and Lip-MALA maintain errors below 122.22.

Table 7: Statistical parameters for the results obtained for algorithms tested for real data.

Parameter	Mean	Sd	Corr	RMSE
MH				
$\boldsymbol{V_p}$ (m/s)	2634.66	255.65	0.64	215.01
$\boldsymbol{V_s}$ (m/s)	1327.43	241.58	0.51	224.74
$\boldsymbol{\rho}$ (Kg/m^3)	2197.22	170.73	0.35	168.44
HMC				
$\boldsymbol{V_p}$ (m/s)	2640.91	199.32	0.69	182.23
$\boldsymbol{V_s}$ (m/s)	1307.19	218.66	0.52	207.86
$\boldsymbol{\rho}$ (Kg/m^3)	2186.65	138.72	0.28	151.41
MALA				
$\boldsymbol{V_p}$ (m/s)	2634.50	217.55	0.65	196.22
$\boldsymbol{V_s}$ (m/s)	1283.90	202.36	0.55	193.84
$\boldsymbol{\rho}$ (Kg/m^3)	2177.61	72.40	0.65	84.42
Lip-MALA				
$\boldsymbol{V_p}$ (m/s)	2642.07	223.19	0.79	155.40
$\boldsymbol{V_s}$ (m/s)	1295.25	175.54	0.75	138.46
$\boldsymbol{\rho}$ (Kg/m^3)	2194.84	125.50	0.48	122.22

Table 8 presents various performance parameters, including acceptance rate and total execution time. Lip-MALA exhibits the highest acceptance rate, while HMC exhibits the lowest. Conversely, MH boasts the lowest total execution time, while HMC demonstrates the highest.

Table 8: Other parameters for real data.

Method	Acceptance rate (%)	Total execution time (s)
MH	32.66	15.48
HMC	3.94	3970.74
MALA	7.38	292.83
Lip-MALA	37.89	1215.87

And a final step, as in the synthetic data, was to test the convergence of the chains, this study employed a posteriori analysis to assess the convergence of samples obtained for the unknown seismic data parameters (denoted by m) using various algorithms. The multivariate effective sample size (mESS) statistic served as the convergence metric. The mESS quantifies the equivalent size of an independent and identically distributed (iid) sample possessing the same covariance structure as the sample generated by a Markov Chain Monte Carlo (MCMC) method in the multivariate case.

To formally determine chain convergence, a minimum effective sample size (minESS) threshold can be established. If the mESS value surpasses the minESS threshold, convergence is achieved. For a more in-depth exploration of the convergence test employed in this work, readers are referred to Vats et al. (2019). Table 9 summarizes the mESS and minESS values obtained for each method.

Table 9: Convergence test for real data.

Method	mESS	minESS
MH	7936.83	7555
HMC	10405.54	7555
MALA	10146.90	7555
Lip-MALA	7979.45	7555

~~5~~4.3 Two-dimensional test with real data

In line with the main objective of this study, which is to comparatively evaluate the performance of different MCMC algorithms applied to prestack seismic inversion in geological contexts of varying complexity, this section includes an additional test using real data in a two-dimensional setting. This extension allows us to validate the scalability and robustness of the methods beyond idealized one-dimensional cases, providing evidence of their practical applicability in environments with greater structural and lithological heterogeneity, typical of exploration in real reservoirs.

The study area corresponds to a sector of the Eastern Basin of Venezuela, characterized by clastic lithology with alternating sandstones and shales. The seismic data used was acquired using conventional reflection techniques and subsequently

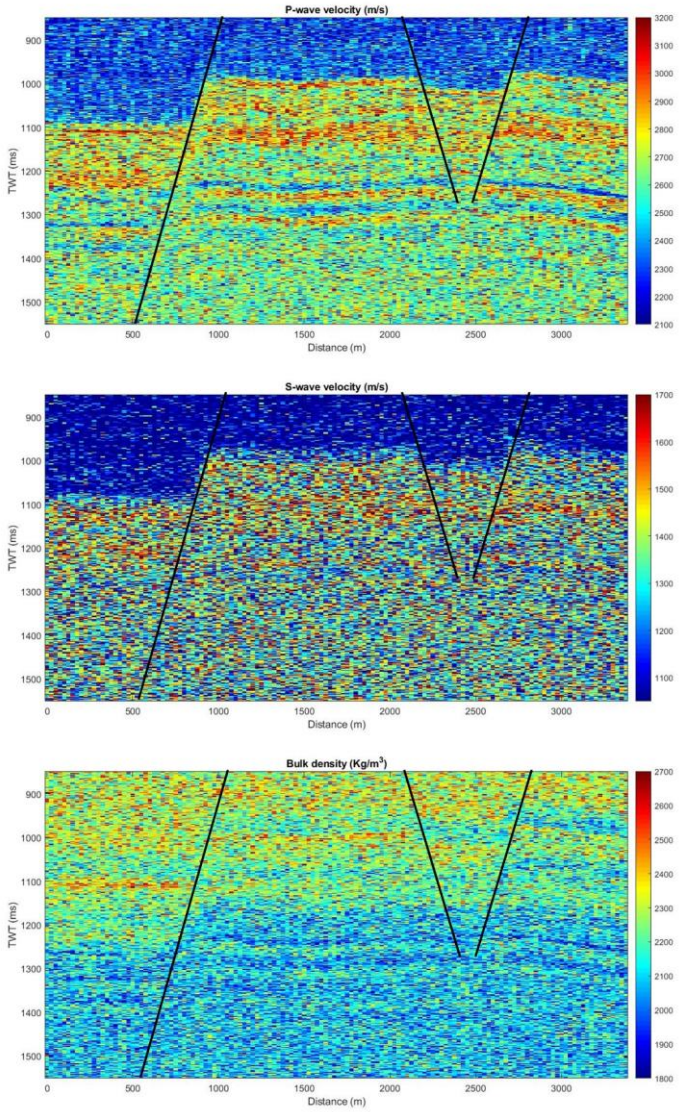
reprocessed to generate prestack gathers organized by incidence angle. Three partial stacks were selected, corresponding to central angles of approximately 19°, 24°, and 29°, and used in the inverse process.

The two-dimensional inversion grid was designed with a vertical resolution of 351 samples, corresponding to the two-way time axis from 850 to 1550 ms, and a horizontal extent of 136 cells from 0 to 340 m. Three fundamental elastic parameters were considered: V_p, V_s and ρ. The prior model was built by integrating the structural interpretation of horizons and faults obtained from seismic data, together with low-frequency interpolation of the elastic properties V_p, V_s and ρ. derived from the well logs available in the study area. This approach allowed the establishment of a consistent geological model that served as an initial reference for the Bayesian inversion process.

Table 10: Posterior standard deviations (2D real data).

Method	**Sd $V_p(m/s)$**	**Sd $V_s(m/s)$**	**Sd $\rho(Kg/m^3)$**
MH	253.31	234.06	162.53
HMC	170.86	145.72	95.51
MALA	243.09	231.35	109.96
Lip-MALA	195.97	159.14	124.74

The four MCMC sampling algorithms analyzed were applied in 1D: MH, HMC, MALA, and Lip-MALA. The resulting realizations were used to calculate the posterior marginal distributions for each parameter, as well as their summary statistics (mean and standard deviation).



~~Figure 10: Inverted models of V_p,~~ Figure 12: Inverted models of V_p, V_s and ρ obtained using the MH algorithm.

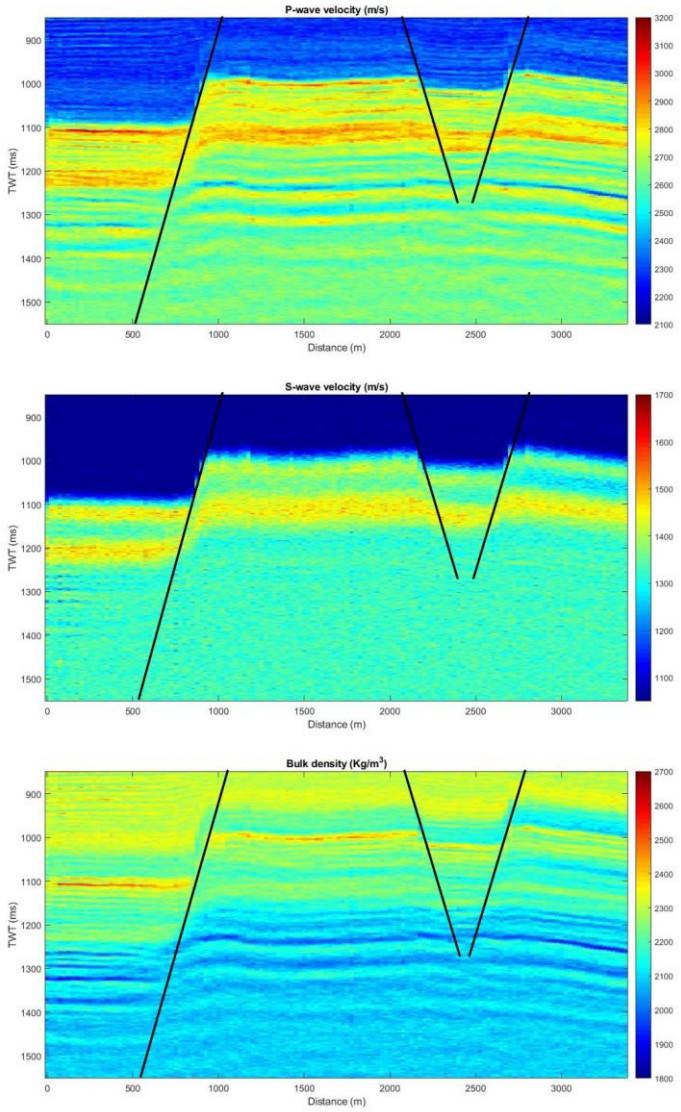
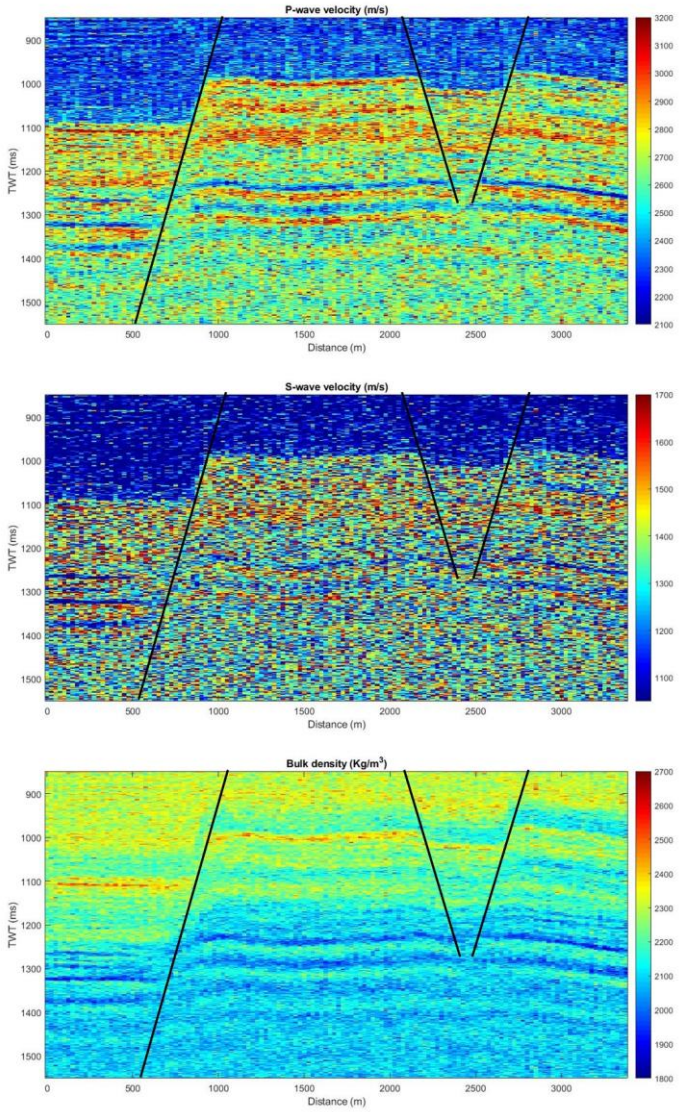


Figure ~~11~~13: Inverted models of V_p, V_s and ρ obtained using the HMC algorithm.



~~Figure 12: Inverted models of V_p,~~ Figure 14: Inverted models of V_p, V_s and ρ obtained using the MALA algorithm.

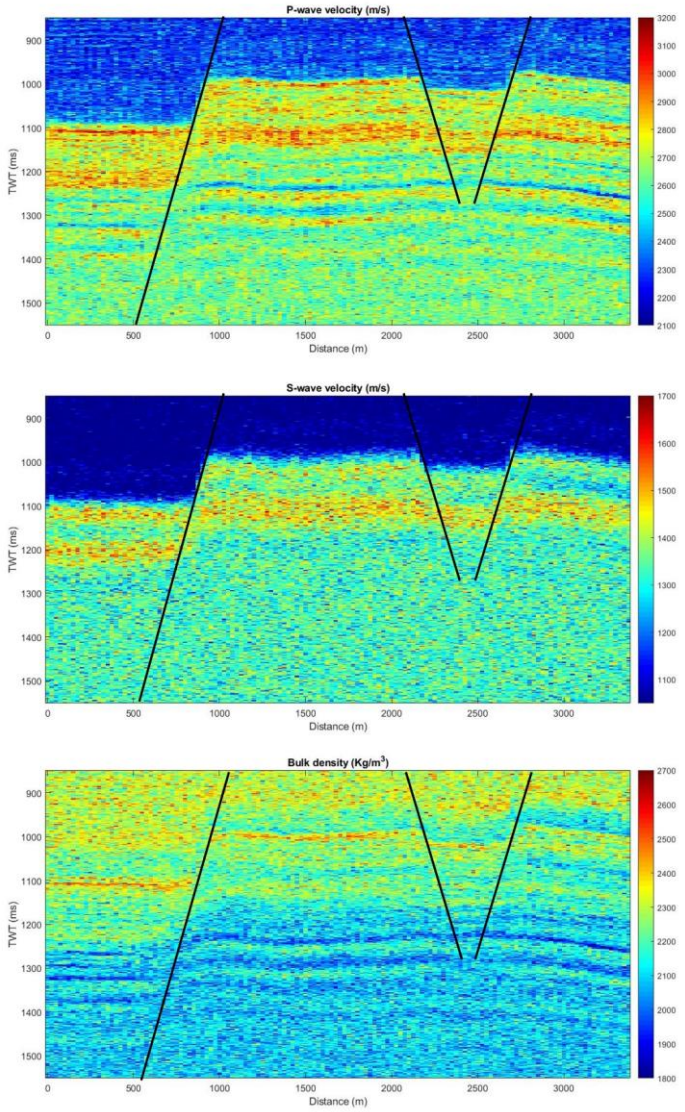


Figure ~~13~~15: Inverted models of V_p, V_s and ρ obtained using the Lip-MALA algorithm.

Quantitative analysis of the figures ~~10~~12 to ~~13~~15 reveals important differences between the algorithms. The HMC method presents the lowest standard deviations for all three parameters, indicating greater accuracy and lower uncertainty in the estimation, although at the cost of higher computational costs. MALA and Lip-MALA also offer robust results, with Lip-MALA demonstrating a better compromise between accuracy and computational efficiency, especially in density estimation. In contrast, the MH algorithm shows the greatest uncertainties, confirming its limitations in contexts of higher dimensionality and complexity.

From an applied perspective, these results are relevant for choosing the most appropriate method in exploratory contexts. In situations where high accuracy is required and sufficient computational resources are available, HMC may be the preferred option. On the other hand, in scenarios where computational time is a constraint, methods such as Lip-MALA offer an efficient and stable alternative. Density (ρ) remains the most challenging parameter to accurately recover, especially for MH and HMC, which is consistent with the results obtained in the one-dimensional case.

Finally, the incorporation of this two-dimensional test with real data demonstrates the practical applicability of the evaluated algorithms, not only under controlled conditions but also in situations representative of real-world geology. Furthermore, it reinforces the validity of the study's overall conclusions by confirming that gradient-based algorithms are more appropriate for complex inverse scenarios, and that algorithm selection must consider both the level of uncertainty and computational cost.

~~6~~5 Discussion

This study presents a comparative analysis of Markov Chain Monte Carlo (MCMC) methods for estimating elastic properties from seismic amplitudes. We demonstrate the application of these methods in a field case, employing the following assumptions: (1) a one-dimensional reservoir model represented by stacked seismic traces, (2) seismic data simulation using the small reflectivity approximation, and (3) the Aki-Richards equation for weak contrast to establish the relationship between seismic data and elastic parameters. Notably, the proposed general formulation transcends these assumptions, allowing for the integration of more sophisticated seismic simulation techniques and comprehensive petrophysical models within a similar framework. Why does Lip-MALA yield smaller errors in the theoretical (1D) test while HMC performs better in the two-dimensional real case? In our experiments, two factors explain this: (i) the evaluation metric and (ii) the dimensionality/structure of the posterior.

(i) Metric. In 1D (synthetic and real), we compared primarily RMSE along the trace; Lip-MALA, by enforcing a Lipschitz-based step control, stabilizes local moves and reduces pointwise error (RMSE) in V_p and V_s. In 2D, we mainly reported posterior SD per parameter; HMC, via long Hamiltonian trajectories, explores connected valleys of the posterior more efficiently and achieves lower global uncertainty (SD), even if RMSE 2D is not emphasized in the main text.

(ii) Dimensionality/heterogeneity. In real 2D, lateral couplings and heterogeneity induce a rough/multimodal posterior; HMC traverses basins more effectively than local Lip-MALA steps, reducing posterior variance.

The four methods studied demonstrate acceptable performance, but in-depth analysis reveals notable differences:

- Velocity estimation: In both the synthetic and real-world scenarios, methods that incorporate gradient calculations (HMC, MALA, and Lip-MALA) outperform MH in estimating velocities.
- Density estimation: Density estimation proves to be the most challenging parameter, with MH and HMC exhibiting unsatisfactory results. However, MALA and Lip-MALA showcase more promising performance.
- Execution time: A significant difference emerges in execution time between methods. MH and MALA exhibit shorter execution times compared to HMC and Lip-MALA, which are considerably more time-consuming.

A natural progression of this research would be to invert prestack seismic data to extract additional elastic parameters and reservoir properties, revealing a more comprehensive subsurface understanding. Similarly, incorporating well log conditioning into the model holds promise, as it could enhance vertical resolution near wells and guarantee that the model aligns with well data at drilling locations.

From an applied perspective, the results obtained in this study are relevant for decision-making in real-world exploration settings. For example, in frontier areas with poor well control, the MH algorithm could be used as a rapid evaluation tool due to its low computational cost, albeit with accuracy limitations. In contrast, methods such as HMC or Lip-MALA would be more suitable for mature fields where higher fidelity in estimating elastic properties is required, despite their greater computational demand. The choice of algorithm should be guided not only by statistical metrics but also by the specific requirements of the geophysical project, the geological setting, and the time and resource constraints available.

The results obtained in this work show consistency with previous research. For example, Gebraad et al. (2020) highlights the effectiveness of the HMC algorithm in full-waveform elastic inversion problems, particularly due to its ability to efficiently explore the posterior space. However, unlike their FWI-oriented approach, our AVO inversion results indicate that Lip-MALA achieves a better balance between accuracy and computational cost, particularly in density estimation, which is crucial in clastic media with gradual transitions. Similarly, Izzatullah et al. (2021) demonstrated that Langevin dynamics-based methods, such as MALA and its adaptive variants, are more efficient in high-dimensional spaces, which is reflected in our study by better definition of lithological boundaries in inverted images. These parallels confirm that Langevin-derived methods are viable and robust options in real-world seismic scenarios where efficiency and stability are practical priorities.

~~Regarding runtime, it is important to contextualize these values in terms of their operational applicability. The algorithms were run on a workstation equipped with an Intel Core i9 processor, 64 GB of RAM, and without the use of GPU acceleration.~~

Although the HMC algorithm presents significantly longer runtimes, these may be acceptable within a seismic interpretation workflow that includes validation and multidisciplinary analysis phases. In contexts where the inversion must be performed in near-real time, such as during well drilling (geosteering), methods such as MH or MALA may be more appropriate despite their lower resolution. Therefore, computation time should not be evaluated in isolation, but rather based on operational

priorities, available computational resources, and the criticality of the information to be estimated at each stage of the geophysical project.

~~7~~6 Conclusions

This study compares various pre-stack inversion methods under an MCMC framework for the estimation of elastic parameters. We invert pre-stacked seismic data to infer velocities ~~()~~(V_P and V_s) and density ~~()~~(ρ), which are linked to the seismic data via the Aki-Richards equation. All methods employed effectively handle the inherent uncertainties associated with seismic and elastic data.

The proposed algorithms allow estimating several important aspects of the posterior distribution, such as the means and standard deviations of the posterior parameters. We rigorously validated the algorithms by measuring the quality of the MCMC sample through correlations, plotting the objective function, seismic traces and estimating the RMSE.

The four methods evaluated in this study exhibit acceptable performance overall, but a closer examination reveals notable differences in their specific capabilities. Velocity estimation: In both the simulated and real-world scenarios, methods that leverage gradient calculations (HMC, MALA, and Lip-MALA) demonstrate superior performance in estimating velocities compared to MH. Density estimation: Density estimation poses the most significant challenge, with MH and HMC exhibiting unsatisfactory results. However, MALA and Lip-MALA demonstrate more promising performance in this area. Execution time: A clear distinction emerges in execution time between the methods. MH and MALA exhibit significantly shorter execution times compared to HMC and Lip-MALA, which are considerably more time-consuming.

Furthermore, the results of the two-dimensional test with real data showed that in situations where high accuracy is required and sufficient computational resources are available, HMC may be the preferred option. On the other hand, in scenarios where computational time is a constraint, methods such as Lip-MALA offer an efficient and stable alternative. This validation in a context closer to real geology strengthens the study's conclusions. The choice of algorithm must consider not only statistical metrics but also the geophysical context, resource availability, and project purpose. In summary we have:

- Balance runtime vs accuracy: HMC yields lower SD at higher cost; Lip-MALA provides strong local accuracy efficiently.
- Choose metrics deliberately: report both RMSE (fit) and SD (uncertainty) to avoid metric-induced contradictions.
- Use HMC for 2D (and higher) problems to reduce global posterior SD and traverse multi-basin landscapes.
- Use MALA/Lip-MALA to stabilize density estimation and reduce pointwise errors.
- Use Lip-MALA when local accuracy (lower RMSE) in VP/VS is prioritized on 1D settings.

Author contributions

RPR and SI designed the study, performed the research, analyzed data, and wrote the paper. GB and RM contributed to refining the ideas, proof the results, carrying out additional analyses, and finalizing this paper.

Use of IA tools

The authors declare that we use AI tools like ChatGPT and others to improve the writing, structure and make the article more readable for readers.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

The authors would like to thank Yachay Tech University for funding this study through the Research Project REG-INV-18-02946.

References

Aki, K., Richards, P., 2002. Quantitative Seismology. University Science Books.

~~Bishop, C.M., 2006. Pattern Recognition and Machine Learning (Information Science and Statistics). Springer-Verlag, Berlin, Heidelberg.~~

Beskos, A & Pillai, N & Roberts, G & Sanz-Serna, J. & Stuart, A. (2013). Optimal tuning of the hybrid Monte Carlo algorithm. Bernoulli. 19.

Bosch, M., 2004. The optimization approach to lithological tomography: Combining seismic data and petrophysics for porosity prediction. GEOPHYSICS, 69, 1272–1282. doi:10.1190/1.1801944.

Bosch, M., Cara, L., Rodrigues, J., Navarro, A., Díaz, M., 2007. A monte carlo approach to the joint estimation of reservoir and elastic parameters from seismic amplitudes. Geophysics 72. doi:10.1190/1.2783766.

Buland, A., Omre, H., 2003. Bayesian linearized avo inversion. GEOPHYSICS, 68, 185–198. doi:10.1190/1.1543206.

Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. The American Statistician, 49(4), 327–335. https://doi.org/10.2307/2684568

Clochard, V., Delépine, N., Labat, K., Ricarte, P., 2009. Post-stack versus pre-stack stratigraphic inversion for monitoring purposes: A case study for the saline aquifer of the Sleipner field. pp. 2417–2421. doi:10.1190/1.3255345.

~~Creutz, M., 1988. Global monte carlo algorithms for many-fermion systems. Phys. Rev. D 38, 1228–1238. doi:10.1103/PhysRevD.38.1228~~

de Lima, P.D.S., Corso, G., Ferreira, M.S., de Araújo, J.M., 2023. Acoustic full waveform inversion with hamiltonian monte carlo method. Physica A: Statistical Mechanics and its Applications 617, 128618. doi:10.1016/j.physa.2023.128618.

Duane, S., Kennedy, A.D., Pendleton, B.J., Roweth, D., 1987. Hybrid monte carlo. Physics Letters B 195, 216 – 222. doi:10.1016/0370-2693(87)91197-X.

~~Dubbeldam, D., Calero, S., Ellis, D., Snurr, R., 2016. Raspa: Molecular simulation software for adsorption and diffusion in flexible nanoporous materials. Molecular Simulation 42, 81–101. doi:10.1080/08927022.2015.1010082.~~

~~Estévez, G., Infante, S., Sáez, F., 2012. Estimación de modelos de equilibrio general en economías dinámicas por métodos de monte carlo y cadenas de markov. Revista de Matemática Teoría y Aplicaciones 19(1), 7–36. doi:10.15517/rmta.v19i1.2102.~~

~~Fichtner, A., Simuté, S., 2018. Hamiltonian monte carlo inversion of seismic sources in complex media. Journal of Geophysical Research: Solid Earth, 123, 2984–2999. doi:https://doi.org/10.1002/2017JB015249.~~

Fichtner, A., Zunino, A., 2019. Hamiltonian nullspace shuttles. Geophysical Research Letters 46, 644–651. doi:10.1029/2018GL080931.

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

Goodway, B., 2001. Avo and lame constants for rock parameterization and fluid detection. CSEG Recorder 26, 39–60.

Hampson, D. P., Russell, B. H., and Bankhead, B., 2005, Simultaneous inversion of prestack seismic data: SEG Technical Program Expanded Abstracts 2005, Society of Exploration Geophysicists, 1633-1637. doi: https://doi.org/10.1190/1.2148008

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.

Helland-Hansen, D., Magnus, I., Edvardsen, A., Hansen, E., 1997. Seismic inversion for reservoir characterization and well planning in the snorre field. The Leading Edge 16, 269–274. doi:10.1190/1.1437616.

Infante, S., ~~s~~Sanchez, l., Hernández, A., 2019. Stochastic models to estimate population dynamics. Statistics, Optimization & Information Computing 7, 311–328. doi:10.19139/soic.v7i2.538.

Izzatullah, M., van Leeuwen, T., & Peter, D. (2020). Accelerated Langevin Dynamics for Bayesian Seismic Inversion. Journal of Geophysical Research: Solid Earth, 125(3), e2019JB018428. https://doi.org/10.1029/2019JB018428

Izzatullah, M., van Leeuwen, T., Peter, D., 2021. Bayesian seismic inversion: a fast sampling Langevin dynamics Markov chain Monte Carlo method. Geophysical Journal International 227, 1523–1553. doi:10.1093/gji/ggab287.

Landa, E., Treitel, S., 2016. Seismic inversion: What it is, and what it is not. The Leading Edge 35, 277–279. doi:10.1190/tle35030277.1.

Landau, L., Lifshitz, E., 1976. Mechanics: Volume 1. Course of theoretical physics, Elsevier Science.

Lemons, D.S., Gythiel, A., 1997. Paul Langevin's 1908 paper "On the Theory of Brownian Motion" ["Sur la th´eorie du mouvement brownien," C. R. Acad. Sci. (Paris) 146, 530–533 (1908)]. American Journal of Physics 65, 1079–1081. doi:10.1119/1.18725.

Ma, X., 2002. Simultaneous inversion of prestack seismic data for rock properties using simulated annealing. GEOPHYSICS 67, 1877–1885. doi:10.1190/1.1527087.

MacKay, D. J. C. (2003). Information Theory, Inference, and Learning Algorithms. Cambridge University Press.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21, 1087–1092. doi:10.1063/1.1699114.

Neal, R. M. (1996). Bayesian Learning for Neural Networks. Lecture Notes in Statistics, 118. Springer. https://doi.org/10.1007/978-1-4612-0745-0

Neal, R., 2012. Mcmc using hamiltonian dynamics. Handbook of Markov Chain Monte Carlo doi:10.1201/b10905-6.

Nemeth, C., Sherlock, C., Fearnhead, P., 2016. Particle metropolis-adjusted langevin algorithms. Biometrika 103, 701–717.

Niu, L., Geng, J., Wu, X., Zhao, L., Zhang, H., 2020. Data-driven method for an improved linearised AVO inversion. Journal of Geophysics and Engineering 18, 1–22. doi:10.1093/jge/gxaa065.

Nemeth, C., & Fearnhead, P. (2021). Stochastic Gradient Markov Chain Monte Carlo. Journal of the American Statistical Association, 116(533), 433–450. https://doi.org/10.1080/01621459.2020.1847120

Robert, C.-P. ~~(2016). The Metropolis–Hastings algorithm. arXiv Preprint.~~and Casella, G. (2004) Monte Carlo Statistical Methods. Springer, Berlin. https://~~arxiv~~dx.doi.org/~~abs/1504.01896v3~~10.1007/978-1-4757-4145-2

Roberts, G. O., & Tweedie, R. L. (1996). Exponential Convergence of Langevin Distributions and Their Discrete Approximations. Bernoulli, 2(4), 341–363. https://doi.org/10.2307/3318418

Roberts, G.O., Gelman, A. and Gilks, W.R. (1997) Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms. Annals of Applied Probability, 7, 110-120. https://doi.org/10.1214/aoap/1034625254

Roberts, G., Stramer, O., 2002. Langevin diffusions and metropolis-hastings algorithms. Methodology And Computing In Applied Probability 4, 337–357. doi:10.1023/A:1023562417138.

Russell, B., Hampson, D., , Bankhead, B., 2006. An Inversion Primer. CSEG RECORDER 31, 96–103.

Stuart, A., Voss, J., Wiberg, P., 2004. Conditional path sampling of sdes and the langevin mcmc method. Commun. Math. Sci. 2. doi:10.4310/CMS.2004.v2.n4.a7.

Sánchez, L., Infante, S., Griffin, V., Rey, D., 2016. Spatio-temporal dynamic model and parallelized ensemble kalman filter for precipitation data. Brazilian Journal of Probability and Statistics 30. doi:10.1214/15-BJPS297.

~~Sarkka, Simo & Merkatas, Christos & Karvonen, Toni. (2021). Gaussian Approximations of SDES in Metropolis-Adjusted Langevin Algorithms. 1-6. 10.1109/MLSP52302.2021.9596301.~~

Tarantola, A., 2005. Inverse Problem Theory and Methods for Model Parameter Estimation. Society for Industrial and Applied Mathematics.

Vats, D., Flegal, J. M., & Jones, G. L. (2019). Multivariate output analysis for Markov chain Monte Carlo. Biometrika, 106(2), 321-337.

Welling, M., Teh, Y., 2011. Bayesian learning via stochastic gradient langevin dynamics, pp. 681–688.

Wu, H., Chen, Y., Li, S., Peng, Z., 2019. Acoustic impedance inversion using gaussian metropolis–hastings sampling with data driving. Energies 12. doi:10.3390/en12142744.