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 Pere-Roa (rperez@yachaytech.edu.ec)

Abstract. Seismic ~~data~~ inversion for estimating elastic properties is a ~~crucial~~ key technique for ~~characterizing~~ reservoir ~~properties post-~~ characterization after drilling. The choice of inversion method ~~significantly impacts~~ strongly influences the accuracy, efficiency, and reliability of results. Bayesian inference based on Markov ~~chain~~ Chain Monte Carlo (MCMC) algorithms ~~enable Bayesian inference,~~ provides a robust framework for incorporating ~~seismic~~ data uncertainty and ~~expert information via~~ prior ~~distribution. This~~ geological knowledge. In this study ~~compares~~, we compare the performance of four inversion methods—Metropolis-Hastings (MH), Hamiltonian Monte Carlo (HMC), ~~and two Lagrangian Diffusion variants [MALA (~~ the Metropolis-~~adjusted~~Adjusted Langevin ~~algorithm)~~ Algorithm (MALA), and its variant Lip-MALA ~~(Lipschitz-MALA)],~~ —in prestack seismic inversion~~,~~ using both synthetic models and real ~~-world~~ data from an eastern Venezuelan hydrocarbon reservoir. ~~All four methods show acceptable performance but differ in specific strengths and weaknesses. Gradient~~ Results indicate that gradient-based methods (HMC, MALA, ~~and~~ Lip-MALA) outperform MH in velocity estimation~~. Density estimation is~~, while density inversion remains more challenging~~;~~. MH and ~~HMC yield unsatisfactory results~~ MALA achieve shorter execution times, whereas ~~MALA~~ HMC and Lip-MALA ~~show promise. Execution time varies significantly: MH and MALA are substantially faster than HMC and Lip-MALA. Therefore, both accuracy and~~ improve accuracy at higher computational ~~efficiency should be considered when choosing a method. The study~~ cost. This analysis evaluates ~~the~~ mean values and standard ~~deviations of the subsequent parameters:~~ deviation (SD) estimates for P-wave ~~(V_p),~~ velocity, S-wave velocity ~~(V_s)~~, and density ~~(ρ). The~~, with quality ~~of the MCMC sample is checked using correlations~~ assessed through correlation metrics, objective function ~~plots~~ behavior, seismic ~~trace~~ traces, and Root Mean Square Error (RMSE~~) estimation. Acceptance rate and execution time assessments reveal HMC has the lowest acceptance rate, and MH the shortest execution time. In addition, a~~). A two-dimensional inversion ~~test~~ with real data ~~was included to evaluate the performance of the~~ further demonstrates algorithms ~~in a more~~ performance under complex geological ~~scenario, 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. Operational implications for method selection are also discussed depending on computational constraints and project objectives~~ 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 ~~crucial~~fundamental for effective reservoir management. This ~~pro-cess~~task requires ~~the integration of~~integrating two ~~distinct~~complementary sources of information~~ sets~~: general reservoir knowledge and reservoir-specific observations. General ~~reservoir~~ knowledge ~~encompasses insights gleaned~~arises from analogous reservoir studies~~, coupled with~~ and established principles in seismic and rock physics~~. In contrast, reservoir-~~, whereas specific observations include ~~direct measurements of the reservoir under study, including well data,~~ well logs, seismic surveys, and ~~historical~~ production ~~data~~history.

~~To optimally analyse this source of information on lithologically complex reservoirs, advanced statistical techniques such as Bayesian inversion methods are required. MCMC algorithms allow uncertainty to be quantified and address nonlinear problems (Bosch et al., 2007).~~

~~This work seeks to answer the following question: Which of the commonly used MCMC algorithms for Bayesian inversion offers more reliable and computationally efficient results in the context of prestack seismic inversion, particularly in lithologically complex reservoirs?~~

~~Consequently, the specific objective of this study is to evaluate the comparative performance of the MH, HMC, MALA and Lip-MALA algorithms in estimating elastic parameters from synthetic and real seismic data, interpreting their strengths and limitations in terms of uncertainty quantification, geological realism and computational feasibility.~~

Seismic data play ~~an important~~a central role in reservoir characterization ~~due to~~because of their extensive spatial coverage. Unlike well logs, which are ~~limited~~restricted to individual ~~well locations~~boreholes, seismic surveys provide a ~~comprehensive picture~~continuous description of the ~~entire~~ reservoir~~...~~. To leverage this information~~ for reservoir characterization, we require methods to transform~~, seismic amplitudes must be transformed into ~~rock properties relevant for reservoir description. Seismic inversion stands is a prominent technique for extracting those elastic and/or~~elastic and petrophysical properties ~~from~~relevant to reservoir description. This transformation is performed through seismic ~~data.~~

~~Seismic~~ inversion~~ is~~, a geophysical inverse problem~~. It~~ that aims to ~~indirectly extract information about the~~ infer subsurface ~~medium (elastic~~ properties (e.g., velocities, density, lithology~~, etc.)~~) from observed seismic data. ~~This requires a robust mathematical framework, typically represented by an equation or system of equations, that accurately describes the physical relationship between the medium (geological model) properties and the recorded seismic response.~~ The ~~process of mapping the parameters of a geological model to quantities in the data space is known as~~ forward modelling~~, generally of the type (Tarantola, 2005):~~

~~$d_{obs} = g(\mathbf{m}) + \xi$, (1)~~

~~where d_{obs} is the observed data, g is the function~~ process that relates ~~the~~model parameters ~~of the medium **m** with~~to recorded data can be expressed as (Tarantola, 2005):

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

where $\boldsymbol{d}_{obs}$ are the observed data, g is the forward operator linking the model parameters $\boldsymbol{m}$ to the data, and ~~ξ~~ξ represents ~~noise due to data~~ measurement and ~~/or modelling~~ modeling errors. ~~For the particular case of~~

In prestack amplitude versus offset (AVO) ~~prestack seismic~~ inversion (Helland-Hansen et al., 1997; Ma, 2002; Buland and Omre, 2003~~)~~), the problem is ~~an~~ ill-posed ~~problem~~ (Landa and Treitel, 2016~~), there is usually inconsistency, and the solution is~~): solutions are non-unique and extremely sensitive ~~and unstable~~ to measurement errors. The most important ~~physical~~ elastic parameters ~~for seismic inversion~~ are P-wave ~~(V_P) and~~ velocity (V_P), S-wave velocity ~~(V_S)~~ (V_S), and density ~~(ρ). These~~ (ρ), from which Lamé parameters can be ~~used to derive Lame parameters, which~~ derived. These are sensitive to rock fluid content and saturation ~~in rocks~~ (Clochard et al., 2009~~). Petrophysical~~) and can be further related to petrophysical parameters~~,~~ such as porosity, sand/shale ratio, and gas saturation~~, can then be estimated from the inverted Lame parameters~~ (Goodway, 2001). ~~Petrophysical parameters are very important in the interpretation of seismic data, which is a crucial process in oil exploration and production projects. By understanding the petrophysical properties of the earth's surface, geologists and engineers can better identify potential reserves of oil and gas~~ Accurately estimating these parameters is therefore crucial for identifying hydrocarbon accumulations.

Bayesian inference provides a robust framework for seismic inversion, as it explicitly accounts for uncertainty and allows the combination of prior geological knowledge with the likelihood of seismic data. Monte Carlo methods, particularly Markov Chain Monte Carlo (MCMC) algorithms, are widely used to sample from posterior distributions, address nonlinearity, and quantify uncertainty (Bosch et al., 2007).

~~The approach of this work is based on computer statistics, which allows us to include uncertainty in seismic data, prior knowledge of model parameters, and through the application of Monte Carlo methods, to generate samples that allow to estimate the posterior distribution (solutions of the inverse problem). Bayesian inference takes into account a likelihood of the seismic data, a prior distribution containing rock property information, combining these two sources of information for later applies Bayes' Theorem (or Bayes' Rule) to approximate the solution.~~

~~There is extensive literature related to MCMC that explores spaces in high dimensions. The MH algorithm was popularized by Metropolis et al., (1953) and Hastings, (1970), initially it was used to simulate the distribution of states of a system of idealized molecules. The MH is a method that facilitates the construction of a stationary Markov chain that converges to a posterior distribution. A more general algorithm is the HMC that was originally developed in the context of lattice quantum chromodynamics (e.g., Duane et al., 1987). Subsequently, the method was extended to Bayesian Neural Networks (Neal, 1996) and incorporated into widely recognized textbooks (MacKay, 2003; Bishop, 2006). The HMC is applied in many disciplines such as: Neural Networks and Machine Learning, (Bishop, 2006); in molecular simulations, (Dubbeldam et al., 2016). In inverse geophysics problems, Bosch et al., (2007) solved an inverse problem following the MCMC methodology, where they quantify the uncertainties of geophysical data and petrophysical properties, combining seismic information with powerful~~

computational methods, establishing a relationship between porosity and acoustic impedance in reservoir areas. In Wu et al., (2019) proposed a MCMC method to reduce the sampling range and improve the efficiency and resolution of impedance inversion, using a Gaussian MH algorithm with data handling for the sampling function [Gaussian MH sampling with data driving (GMHDD) approach]. In Gebraad et al., (2020) developed a Bayesian inversion methodology to treat the full elastic waveform, their proposal is based on HMC sampling of the posterior distribution, use adjoint techniques, and compute the mass matrix considering different sensitivities of seismic velocities and densities. In Izzatullah et al., (2021) studied the seismic inversion problem under a Bayesian approach, implement a MCMC algorithm inspired by Langevin dynamics, and propose a rule for determining the adaptive step size in MCMC algorithm that replaces the MH acceptance step. In Fichtner and Simuté, (2018) developed a model of probabilistic inversion that considers the heterogeneous 3D structure of the earth, the method is based on numerical simulations of wave fields in complex media and on HMC sampling. In de Lima et al., (2023) used a full waveform inversion (FWI) method, the proposed technique is of high resolution and is used in geophysics to evaluate the physical parameters and build subsurface models in a noisy scenario and with limited data, proposed a new way to adjust the mass matrix based on the seismic survey acquisition geometry, and demonstrate significant improvements of the ability of the HMC method in reconstructing reasonable seismic models with manageable computational costs. Both MALA and Lip-MALA algorithms belong to the family of Langevin Monte Carlo algorithms, which are derived from the Langevin dynamics. Lip-MALA algorithm implements a locally Lipschitz adaptive step-size (Izzatullah et al., 2021, Welling and Teh, 2011, Roberts and Tweedie 1996; Nemeth and Fearnhead 2020).

This article studiesSeveral MCMC algorithms have been proposed and applied in geophysics. The Metropolis–Hastings (MH) algorithm (Metropolis et al., 1953; Hastings, 1970) constructs a stationary Markov chain that converges to the target posterior distribution. A more general method is 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 (Gebraad et al., 2020; 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 Metropolis-adjusted Langevin algorithm (MALA) (Roberts and Tweedie, 1996; Welling and Teh, 2011) exploits gradient information to accelerate convergence. Lip-MALA, a more recent variant, incorporates a locally Lipschitz adaptive step size to enhance sampling efficiency and stability (Nemeth and Fearnhead, 2020; Izzatullah et al., 2021).

Building upon this background, the present study investigates the impact of the inversion method choice of theon prestack seismic inversion method on the results of the inversion. We. Specifically, we compare the performance of four algorithms: MH, HMC, MALA, and Lip-MALA for prestack seismic inversion. We validate the algorithms within estimating elastic parameters from synthetic and real seismic data and measure. The analysis evaluates their performance in terms of uncertainty

4

quantification, geological realism, and computational feasibility. Diagnostic methods are used to assess the quality of ~~the samples~~ generated ~~by the MCMC algorithms through diagnostic methods. This article is structured by first reviewing~~samples.

The rest of this paper is organized as follows. In Section 2 a review of the theory of seismic inversion~~, then we review the theory of the~~ and amplitude versus offset (AVO) analysis. Section 3 outlines the four MCMC algorithms considered. Section 4 ~~methods used and the AVO theory, in~~presents the results ~~part we show what was obtained~~ for synthetic ~~data~~ and real data ~~and finally we have the discussion and~~. Section 5 the results are analysed. Finally, Section 6 summarizes the main conclusions.

2 The ~~seismic inversion problem~~Seismic Inversion Problem

Seismic inversion ~~is a way to use~~uses seismic ~~waves~~traces to ~~understand the~~infer subsurface ~~$\mathbf{m} \in R^{N_m}$~~,properties model, denote usually by $\boldsymbol{m} \in \mathbb{R}$, such as ~~seismic~~ velocities and densities from observed seismic data $\boldsymbol{d}_{obs} \in \mathbb{R}^{N_D}$, where N_m and N_D are the dimensions of the model parameters and the observed seismic data ~~$d_{obs} \in R^{N_D}$, where N_m and N_D are the dimensions of the model parameters and the observed seismic data. This can be solved using~~, respectively. This problem can be formulated within a Bayesian framework ~~that treats the~~, where inversion ~~problem~~is treated as a statistical inference problem.

In Bayesian inference, we ~~start from the~~begin with a prior probability distribution of the subsurface parameters ~~of the subsurface models. This prior distribution represents our~~model, representing existing geological knowledge ~~of the ground~~ before ~~seeing~~considering the seismic observations data. ~~We then update this prior distribution with the~~By incorporating seismic data ~~using Bayes'~~through Bayes' theorem, this prior distribution is updated to obtain the posterior probability distribution ~~of the subsurface model parameters.~~

~~The posterior probability distribution encodes~~ , which reflects the degree of confidence in ~~subsurface~~the estimated model ~~parameter estimate. This~~parameters. The posterior distribution allows ~~us to quantify the~~ a rigorous quantification of uncertainty ~~of the underground parameters, considering the~~, as it integrates contributions from seismic data, ~~the~~ prior ~~data~~ information, and the forward ~~model~~modelling operator.

~~To fully characterize~~Fully characterizing the posterior ~~probability~~ distribution~~, we usually need to estimate~~ generally requires sampling the model ~~parameters. This can be~~space, which is computationally ~~expensive.~~
demanding. In this ~~section, we present a general framework for seismic~~work, we adopt the Bayesian inference~~, as established in previous studies~~ framework commonly used in geophysical inversion (Tarantola, 2005~~,~~; Bosch, 2004; Izzatullah et al., 2021 ~~and~~; de Lima et al., 2023~~). This framework~~), which provides a practical ~~basis~~foundation for implementing seismic Bayesian inversion in real-world geophysical applications.

2.1 Bayesian ~~inference framework~~Inference Framework for ~~seismic data~~Seismic Data

Bayesian ~~methods attempt~~inversion aims to ~~sample~~characterize the ~~full~~ posterior distribution ~~over the parameters and possibly latent variables which provides a way to assert uncertainty in the model.~~
~~Under the statistical approach~~ of ~~Bayesian inversion, the objective is to find the posterior distribution of the latent states (~~unknown parameters~~) $\mathbf{m}$~~ given ~~the~~ observed ~~data d_{obs}. To solve Bayesian~~ seismic ~~inversion, we need to know about the prior probability density $\rho(\mathbf{m})$ and the likelihood function $\mathrm{L}(\mathbf{m})$.~~ data. This requires specifying two key components:

- Prior probability density ~~tells us how confident we are~~$\rho(\boldsymbol{m})$: which encodes our confidence in ~~our knowledge of~~ the subsurface ~~model~~ parameters~~,~~ before ~~we look at the~~incorporating seismic data. ~~The likelihood function tells us how likely it is that a particular set of~~
- Likelihood function $L(\boldsymbol{m})$: which quantifies the probability of observing the data given a candidate subsurface model ~~parameters produce the seismic data that we observed.~~ (Izzatullah et al., 2021).

~~Bayes' theorem combines the prior probability density and the likelihood function to give us the posterior probability density $\sigma(\mathbf{m})$. The posterior probability density tells us how confident we are in our knowledge of the subsurface model parameters, after we have looked at the seismic data (Bosch, 2004).~~

~~$\sigma(\mathbf{m}) = \mathrm{c}\mathrm{L}(\mathbf{m})\rho(\mathbf{m})$ (2)~~

- Posterior probability distribution: Bayes' theorem combines these two elements as:

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

where ~~$\sigma(\mathbf{m})$ is~~$\sigma(\boldsymbol{m})$ denote the posterior probability density, c is a normalization constant, ~~$\mathrm{L}(\mathbf{m})$~~ $L(\boldsymbol{m})$ is the likelihood ~~and $\rho(\mathbf{m}) \sim$~~.

$\rho(m) \sim N$~~$(\mathbf{m}_{prior} | C_m)$~~$(\boldsymbol{m}_{prior}, \boldsymbol{C}_m)$ is the prior probability density~~.~~ (Bosch, 2004), where N denotes a Gaussian probability distribution, with mean $\boldsymbol{m}_{prior}$ and covariance matrix given by C_m

In this paper, we will focus on posterior probability distribution, which can be expressed mathematically as

~~$\sigma(\mathbf{m}) = \mathrm{c}\exp(-\mathrm{S})$, (3)~~

~~with the half-sum of squares S being:~~

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

~~with $\mathrm{g}: \mathbf{m} \rightarrow \mathrm{d}_{obs}$~~

$$\sigma(\boldsymbol{m}) = ce^{-S}, \tag{3}$$

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 the function solving the seismic forward problem, ~~C_d~~ $\boldsymbol{C}_d$ being the data covariance matrix that describes second-order statistics on the data uncertainties, ~~C_m~~ and $\boldsymbol{C}_m$ an appropriate covariance matrix describing variability and correlation between parameters of the medium and ~~$\mathbf{m}_{prior}$~~ $\boldsymbol{m}_{prior}$ is a prior model.

~~Usually~~In practice, the ~~equation given in (3)~~posterior distribution defined above is analytically intractable~~, but~~. Therefore, it ~~can be~~is approximated numerically ~~by~~ using ~~the simulated~~ samples ~~$\mathbf{m} \sim \sigma(\mathbf{m})$, using~~ $\{\boldsymbol{m}_i\}_{i=1,..,N}$ generated by MCMC ~~computational~~ 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

3.1 Metropolis~~-~~–Hastings (MH)

MH algorithm is ~~a MCMC method~~one of the most widely used ~~to generate samples~~ MCMC methods for sampling from ~~Markov chains. MH algorithm, introduced~~complex probability distributions. Originally proposed by Metropolis et al. (1953) and ~~later~~ generalized by Hastings (1970), MH defines a transition probability that ensures ~~the Markov chain is ergodic, satisfies~~ergodicity, detailed balance, and ~~exhibits~~ reversibility~~,~~ of the chain (Chib and Greenberg, 1995).~~ This~~

The algorithm generates a sequence of values ~~m~~$\boldsymbol{m}_t$ forming a Markov chain~~, which can be used to approximate a~~ that approximates the target posterior density ~~$\sigma(\mathbf{m})$.~~

~~In MH, a candidate configuration is produced from a source sampling distribution, which is not the target distribution. The source sampling distribution can be anything, but it is desirable for the efficiency of the algorithm that it is somehow close to the target distribution, which is to be sampled. The algorithm is based on comparing the candidate configuration and the current configuration, to decide whether the candidate is accepted as the next step of the chain or if it is rejected, repeating the current configuration as the new link.~~

~~In order to establish this comparison, it is necessary to calculate the multivariate density for both configurations, or the ratio between them. Consider a prior model, $\rho(\mathbf{m})$, and a likelihood function, $L(\mathbf{m})$, where $\mathbf{m}$ is a point in the sample space. Starting from any configuration in the parameter space, and with the chain in the configuration corresponding to the step, $\mathbf{m}^{(t-1)}$, the MH defines a chain that converges to the target density $\sigma(\mathbf{m})$.~~

~~A variety of proposal functions can be used and the speed of convergence and the quality of the estimators obtained substantially depend on the quality of this proposal (Symmetric MH, Independent MH, Random walk MH). The acceptance rate and the effective sample size are used to calibrate the MH algorithm. This means that a range of values of the parameters involved in the proposal must be compared and the value that maximizes the objective function must be selected, (Robert 2016).~~

~~The configuration for the new step $\mathbf{m}^{(t)}$~~ $\sigma(\boldsymbol{m})$. The new value $\boldsymbol{m}_t$ is obtained according to the Metropolis transition rule is as follows:

~~1.~~Step 1. Initialization**:** Choose a proposal function ~~$q\left(\mathbf{m}/\mathbf{m}^{(t-1)}\right)$,~~ $\boldsymbol{q}(\boldsymbol{m}|\boldsymbol{m}_{t-1})$ an initial state ~~m.~~$\boldsymbol{m}$ and the previous value $\boldsymbol{m}_{t-1}$.

~~2.~~Step 2. Propose a new state**:** At iteration ~~t generate~~$\boldsymbol{t}$, a candidate values ~~$\widetilde{\mathbf{m}}$~~$\widetilde{\boldsymbol{m}}$ is generated from the proposal probability density ~~$\widetilde{\mathbf{m}} \sim q\left(\widetilde{\mathbf{m}}/\mathbf{m}^{(t-1)}\right)$,~~ $\widetilde{\boldsymbol{m}} \sim \boldsymbol{q}(\widetilde{\boldsymbol{m}}|\boldsymbol{m}_{t-1})$. and generate ~~$u \sim U$~~ $\sim Uniform(0,1)$, where ~~$q\left(\mathbf{m}/\mathbf{m}^{(t-1)}\right) = q\left(\mathbf{m}^{(t-1)}/\mathbf{m}\right)$, is~~ $\boldsymbol{q}(\widetilde{\boldsymbol{m}}|\boldsymbol{m}_{t-1}) = \boldsymbol{q}(\boldsymbol{m}_{t-1}|\widetilde{\boldsymbol{m}})$, is a symmetric probability distribution.

~~3.~~Step 3. **C**ompute the acceptance probability~~,~~:

~~$$p_{accept} = \min\left(1, \frac{L(\widetilde{\mathbf{m}})}{L(\mathbf{m})}\right) \quad (5)$$~~

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

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

Iteratively repeating the MH rule generates a chain that converges to a sample of the target probability density. ~~The MH algorithm has some advantages and disadvantages allows sampling from arbitrary objective distributions, it is not necessary to determine the marginals, it is simple to implement, and it has a better acceptance and rejection rate in high-dimensional spaces than other competing algorithms. In addition, it can have a poor convergence rate when samples are correlated, it has problem when the target distribution is multimodal, and it is sensitive to the step size between draws, choosing too large or small a step can affect the convergence of the parameters.~~

The efficiency of MH strongly depends on the choice of proposal distribution (e.g., random walk, independent, or symmetric MH). Calibration typically involves adjusting parameters such as step size to maximize acceptance rates and effective sample size (Robert, 2016).

3.2 Hamiltonian Monte Carlo (HMC)

HMC is a sampling algorithm that was ~~originally developed for~~first introduced in molecular dynamics (Duane et al., 1987~~). It is now commonly used for sampling~~) and later adapted to Bayesian inference problems ~~where the gradients~~(Neal, 2012). It is particularly effective for high-dimensional posteriors when gradient information of the ~~posterior probability distribution $\sigma(\mathbf{m})$ with respect to the model parameters $\mathbf{m}$ are easy to compute. HMC~~$\sigma(\boldsymbol{m})$ is available and it is more efficient than standard ~~Metropolis-Hastings for high-dimensional problems.~~MH. However, the cost of generating independent samples with HMC grows ~~faster than the cost of generating samples with Metropolis-Hastings. Specifically, the cost of generating independent samples with HMC grows~~ as ~~$O(n^{5/4})$~~$O(n^{5/4})$ (Neal, 2012~~),~~) while with MH grows as $O(n^2)$ (Creutz, 1988) where n is the dimension of the model parameter space. ~~The cost of generating independent samples with MH grows as $\mathcal{O}(n^2)$ (Creutz, 1988).~~Thus, HMC requires more computation.

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) ~~$\mathrm{p}(\mathbf{m}) = \sigma(\mathbf{m})$~~ $\boldsymbol{p}(\boldsymbol{m}) = \sigma(\boldsymbol{m})$ (This notation is adopted in order to maintain consistency with the convention used in the field of Physics). HMC regards the current state ~~$\mathbf{m}$~~ $\boldsymbol{m}$ of the Markov chain as the location of a physical particle in n-dimensional space ~~$\mathbb{M}$~~ 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):

~~$\mathrm{U}(\mathbf{m}) = -\ln \mathrm{p}(\mathbf{m}) \quad (6)$~~

$$U(\boldsymbol{m}) = -log\left(\boldsymbol{p}(\boldsymbol{m})\right) \quad (6)$$

If the probability density function ~~p~~ $\boldsymbol{p}(\boldsymbol{m})$ of the subsurface model parameters is Gaussian, then the potential energy ~~U~~ $U(\boldsymbol{m})$ of the system is equal to the least squares misfit ~~$\mathrm{S}(\mathbf{m})$~~ $S(\boldsymbol{m})$, up to an additive constant. To make the system physically complete, we need to add momentum variables ~~p~~ $\boldsymbol{p}$ (which is the same PDF) 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 ~~M~~ $\boldsymbol{M}$ of dimension $n \times n$ represents the resistance to change. The kinetic energy of the system is defined by the momenta and the mass matrix as

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

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

The HMC algorithm uses a random momentum ~~p~~ $\boldsymbol{p}$, drawn from a multivariate Gaussian distribution with covariance matrix ~~M~~ $\boldsymbol{M}$. 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,

~~$\mathrm{H}(\mathbf{m}, \mathrm{p}) = \mathrm{U}(\mathbf{m}) + \mathrm{K}(\mathrm{p}) \quad (8)$~~

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

Hamilton's equations

~~$\frac{d\mathbf{m}}{d\tau} = \frac{\partial \mathrm{H}}{\partial \mathrm{p}}, \quad \frac{d\mathrm{p}}{d\tau} = -\frac{\partial \mathrm{H}}{\partial \mathbf{m}} \quad (9)$~~

$$\frac{d\boldsymbol{m}}{d\tau} = \frac{\partial H}{\partial \boldsymbol{p}}, \quad and \quad \frac{d\boldsymbol{p}}{d\tau} = \frac{-\partial H}{\partial \boldsymbol{m}} \quad (9)$$

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~~ 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,

~~$$\frac{d\mathbf{m}}{d\tau} = M^{-1}p, \quad \frac{dp}{d\tau} = -\frac{\partial U}{\partial \mathbf{m}} \tag{10}$$~~

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

In HMC, the model parameters ~~m~~ $\boldsymbol{m}$ and their moment ~~p~~ $\boldsymbol{p}$ are represented as a state. It then evolves the state (~~m, p~~ $\boldsymbol{m}, \boldsymbol{p}$) over time ~~τ~~ τ using Hamiltonian dynamics. This generates a distribution of the possible states of the system with new position ~~m̃~~ $\tilde{\boldsymbol{m}}$, momentum ~~p̃~~ $\tilde{\boldsymbol{p}}$, potential energy ~~Ũ~~ $\tilde{U}$, and kinetic energy ~~K̃~~ $\tilde{K}$, 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.

~~$$p(\mathbf{m}) = \exp(-U(\mathbf{m})) \tag{11}$$~~

$$\boldsymbol{p}(\boldsymbol{m}) = exp(-U(\boldsymbol{m})) \tag{11}$$

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 ~~p(m).~~ $\boldsymbol{p}(\boldsymbol{m})$. 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 ~~p(m).~~ $\boldsymbol{p}(\boldsymbol{m})$. 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 **m** and then following these steps (Gebraad et al., 2020):

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

~~2.~~ Step 2. Evaluate the Hamiltonian H of model $\boldsymbol{m}$, using its momenta $\boldsymbol{p}$.

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

~~4.~~ Step 4. Calculate the Hamiltonian ~~H̃~~ $\tilde{H}$ of the model ~~m̃~~ $\tilde{\boldsymbol{m}}$ with momenta ~~p̃~~.

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

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

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

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

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 results in lower acceptance~~This leads to slower convergence and increases computational cost.

3.3 The 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 ~~$\sigma(\mathbf{m_t})$~~$\sigma(\boldsymbol{m}_t)$, is a continuous-time stochastic process ~~$(\mathbf{m_t})_{t\geq 0}$~~$(\boldsymbol{m}_t)_{t\geq 0}$ in ~~$\mathbb{R}^{N_m}$~~$\mathbb{R}^{N_m}$ that evolves following the stochastic differential equation (SDE) (Roberts and Stramer, 2002; Nemeth et al., 2016; Izzatullah et al., 2021) and (Infante et al., 2019),

~~$$d\mathbf{m_t} = \Sigma\nabla \log \sigma(\mathbf{m_t})\, dt + \sqrt{2}\,\Sigma^{-\frac{1}{2}} dW_t \qquad (13)$$~~

$$d\boldsymbol{m}_t = -\boldsymbol{\Sigma}\nabla log\,[\sigma(\boldsymbol{m}_t)]dt + \sqrt{2}\boldsymbol{\Sigma}^{\frac{-1}{2}} d\boldsymbol{W}_t \qquad (13)$$

where ~~$(W_t)_{t\geq 0}$~~$(\boldsymbol{W}_t)_{t\geq 0}$ is a standard n-dimensional Brownian motion, ~~Σ~~$\boldsymbol{\Sigma}$ is a symmetric positive definite matrix, ~~$\nabla \log \sigma(\mathbf{m_t})$~~$\nabla log\,[\sigma(\boldsymbol{m}_t)]$ is the drift term of the Brownian particle ~~$\mathbf{m_t}$~~$\boldsymbol{m}_t$ and σ~~(.)~~$(\boldsymbol{m}_t)$ is a stationary posterior distribution.

3.3.1 Metropolis-adjusted Langevin algorithm (MALA)

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

~~$$\mathbf{m} = \tilde{\mathbf{m}} + \tau_t \Sigma\nabla \log \sigma(\mathbf{m_t}) + \sqrt{2\tau_t}\,\Sigma^{-\frac{1}{2}}\epsilon_t, \quad \epsilon_t \sim N(0, I_{n\times n}) \qquad (14)$$~~

$$\boldsymbol{m} = \tilde{\boldsymbol{m}} + \tau_t \boldsymbol{\Sigma}\nabla log\,[\sigma(\boldsymbol{m}_t)] + \sqrt{2\tau_t}\Sigma^{\frac{-1}{2}}\boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(0, I_{n\times n}) \qquad (14)$$

where ~~τ~~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 ~~$\tilde{\mathbf{m}}$~~$\tilde{\boldsymbol{m}}$, a new observation $\boldsymbol{m}$ is generated from a proposal density q($\boldsymbol{m}$). The candidate is then accepted with probability ~~p_{accept}~~ p_{accept} given by,

~~$$p_{accept} = \min\left(1, \frac{L(\tilde{\mathbf{m}})q(\tilde{\mathbf{m}}, \mathbf{m_{t-1}})}{L(\mathbf{m})q(\mathbf{m_{t-1}}, \tilde{\mathbf{m}})}\right) \qquad (15)$$~~

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

In summary, MALA algorithm is obtained as follows:

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

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

~~$$\mathbf{m} = \widetilde{\mathbf{m}} + \tau_t \Sigma \nabla \log \sigma(\mathbf{m}_t) + \sqrt{2\tau_t}\, \Sigma^{-\frac{1}{2}} \epsilon_t, \quad (16)$$~~

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

~~3.~~Step 3. Compute the accept-reject probability

~~$$\mathrm{p_{accept}} = \min\left(1, \frac{L(\widetilde{\mathbf{m}})}{L(\mathbf{m})}\right) \quad (17)$$~~

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

~~4.~~Step 4. Draw u from a uniform distribution ~~$u \sim U$~~ $\sim Uniform(0,1)$, if ~~$\mathrm{p_{accept}}$~~$p_{accept} < u$ then accept ~~$\mathbf{m_t} = \widetilde{\mathbf{m}}$~~$\boldsymbol{m}_t = \widetilde{\boldsymbol{m}}$, else ~~$\mathbf{m_t} = \mathbf{m_{t-1}}$~~$\boldsymbol{m}_t = \boldsymbol{m}_{t-1}$

~~5.~~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.

3.3.2 MALA with locally Lipschitz adaptive step size (Lip-MALA)

In the MALA algorithm, it is required to calibrate the step-size ~~τ~~τ, because ~~τ~~τ must decrease with dimension, n. then ~~τ~~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. ~~In~~See in (Izzatullah et al., 2021) ~~propose~~proposed ULA ~~with~~variant, Lip-MALA, in which the step-length ~~τ based on~~τ is adapted according to the Lipschitz condition,

~~τ_t~~ τ_t

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

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

~~1.~~Step 1. Choose an initial solution ~~$\mathbf{m}_{prior}$~~ $\widetilde{\boldsymbol{m}} = \boldsymbol{m}_{prior}$, the discretization step-length ~~τ, β_0~~ τ, $\beta_0 = +\infty$ and $L_c =$ ~~$N_{\mathbf{m}}^{-1/3}$~~ $N_m^{-1/3}$.

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

~~$\mathbf{m} = \widetilde{\mathbf{m}}$~~ $\boldsymbol{m}$

$$= \widetilde{\boldsymbol{m}} - \tau_t \Sigma\nabla \log \sigma(\widetilde{\mathbf{m}})\ \boldsymbol{\Sigma}\nabla log[\sigma(\widetilde{\boldsymbol{m}})]$$

$$+ \sqrt{2\tau_t}\,\Sigma^{-\frac{1}{2}}\epsilon_t\,\boldsymbol{\Sigma}^{\frac{-1}{2}}\boldsymbol{\epsilon}_t, \qquad ~~(19)~~(19)$$

~~3.~~Step 3. Compute the accept-reject probability

~~$p_{accept} = \min\left(1, \frac{L(\widetilde{\mathbf{m}})}{L(\mathbf{m})}\right)$ (20)~~

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

Draw ~~$u \sim U$~~ $uU(0,1)$, if ~~$p_{accept} <$~~ $p_{accept} < u$ then accept ~~$\mathbf{m}_t = \widetilde{\mathbf{m}}$~~ $\boldsymbol{m}_t = \widetilde{\boldsymbol{m}}$, then update,

~~$\tau_t = \min\{\sqrt{1+\beta_{t-1}}\tau_{t-1}, L_c \frac{\|\mathbf{m}_t - \mathbf{m}_{t-1}\|}{|\Sigma\nabla \log\sigma(\mathbf{m}_t) - \Sigma\nabla\log\sigma(\mathbf{m}_{t-1})|}\}$~~

$$\tau_t = min\{\sqrt{1+\beta_{t-1}}\tau_{t-1}, L_c \frac{\|\boldsymbol{m}_{t+1} - \boldsymbol{m}_t\|}{|\boldsymbol{\Sigma}\nabla log[\sigma(\boldsymbol{m}_{t+1})] - \boldsymbol{\Sigma}\nabla log[\sigma(\boldsymbol{m}_t)]|}\} \qquad (21)$$

$$\beta_t = \frac{\tau_t}{\tau_{t-1}} \qquad (22)$$

~~4.~~Step 4. Else reject ~~m_t~~ $\boldsymbol{m}_t =$ ~~m_{t-1}~~ $\boldsymbol{m}_{t-1}$. Then, repeat this process until the convergence.

~~4~~The resulting Lip-MALA algorithm follows the same steps as MALA but dynamically adjusts ϵ, improving stability and sampling efficiency in high dimensions.

Table 1 summarizes the main advantages and limitations of the methods used in this study, in order to provide the reader with a clear and concise reference that facilitates the comparative understanding of their scope and restrictions.

Table 1: Advantages and limitations of the sampling methods

Method	**Advantages**	**Limitations**
MH	MH is simple to implement, flexible (can target arbitrary distributions), and it has a better acceptance and rejection rate in high-	Convergence may be slow in high-dimensional or multimodal problems, and poor

	dimensional spaces than other competing algorithms.	tuning of the proposal step size can lead to highly correlated samples.
HMC	HMC reduces random walk behavior, explores the posterior efficiently, and performs well in high dimensions.	It requires computation of gradients and careful tuning of ϵ (step size) and L (trajectory length). Poor tuning reduces acceptance rates and increases computational cost.
MALA	MALA exploits gradient information, improves mixing, and scales better to high-dimensional problems.	Requires gradient evaluations at each iteration, which can be computationally costly.
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 manual tuning. This makes it more effective in high-dimensional and complex posterior spaces compared to standard MALA	Include the extra computational cost of estimating the Lipschitz constant, reliance on accurate gradient calculations, and potential difficulties when dealing with highly multimodal posteriors.

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):

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

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

where

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

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

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

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

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

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

In equations (23 - 26), the incident angle θ θ is the angle at which a wave hitsreflected off a surface. Vp, Vs V_p, V_s and ρ ρ represent the velocities of P-waves, S-waves, and the density of a material, respectively. ΔVp, ΔVs $\Delta V_p, \Delta V_s$ and Δρ $\Delta\rho$ are the changes in V_p, V_s and ρ across a reflective interface. $\overline{V_p}, \overline{V_s}$ $\overline{V_p}, \overline{V_s}$ and $\overline{\rho}$ $\overline{\rho}$ 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),

$$T(\theta) = \frac{1}{2}c_1W(\theta)DL_{V_p} + \frac{1}{2}c_2W(\theta)DL_{V_s} + \frac{1}{2}c_3W(\theta)DL_{\rho}, \quad (27)$$

where $L_{V_p} = \ln(V_p)$, $L_{V_s} = \ln(V_s)$, $L_\rho = \ln(\rho)$,

$$T(\theta) = \frac{1}{2}c_1W(\theta)DL_{V_p} + \frac{1}{2}c_2W(\theta)DL_{V_s} + \frac{1}{2}c_3W(\theta)DL_{\rho}, \quad (27)$$

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

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

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 $L_{\rho 0}$ $[L_{V_{p0}} \quad L_{V_{s0}} \quad L_{\rho_0}]^T$ 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.

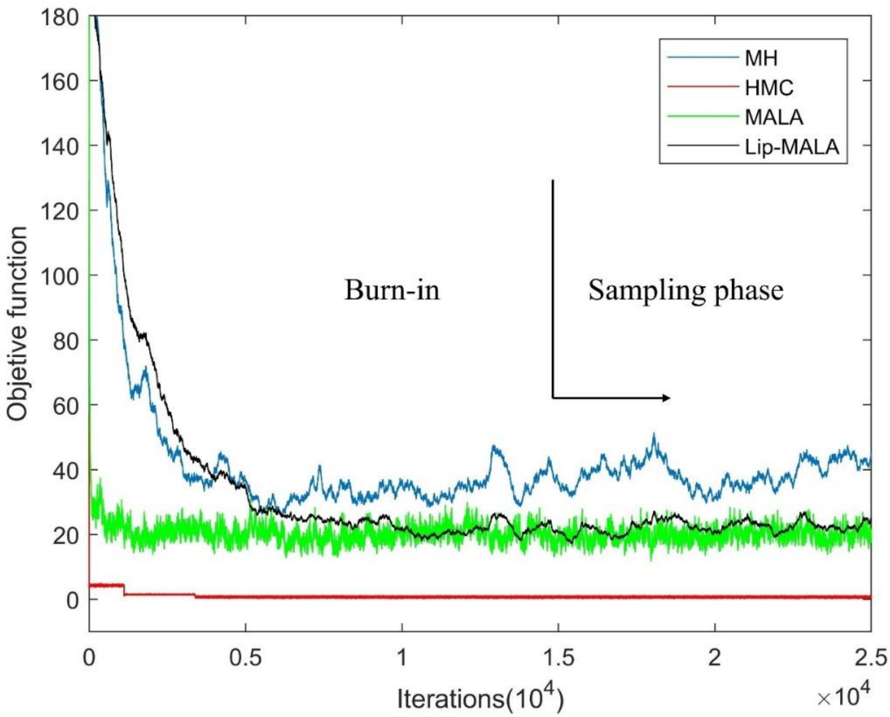
5 Results

5.1 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 θ1 θ_1 = 9°, θ2 $9^o, \theta_2$ = 18.5° 18.5^o and θ3 θ_3 = 27.5° 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 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. 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.



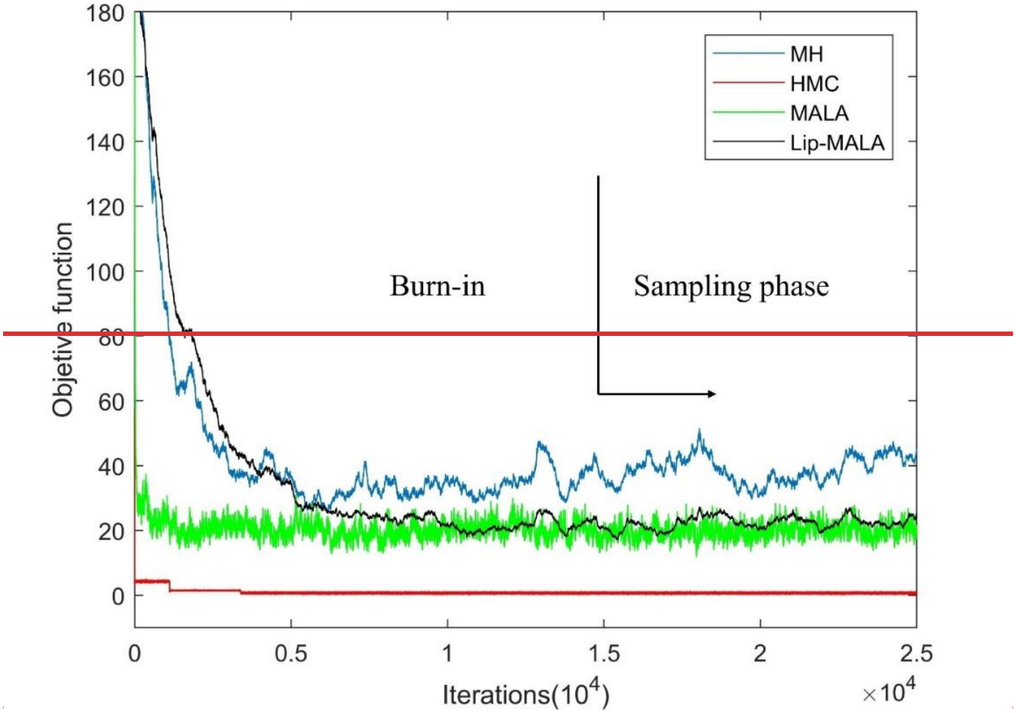


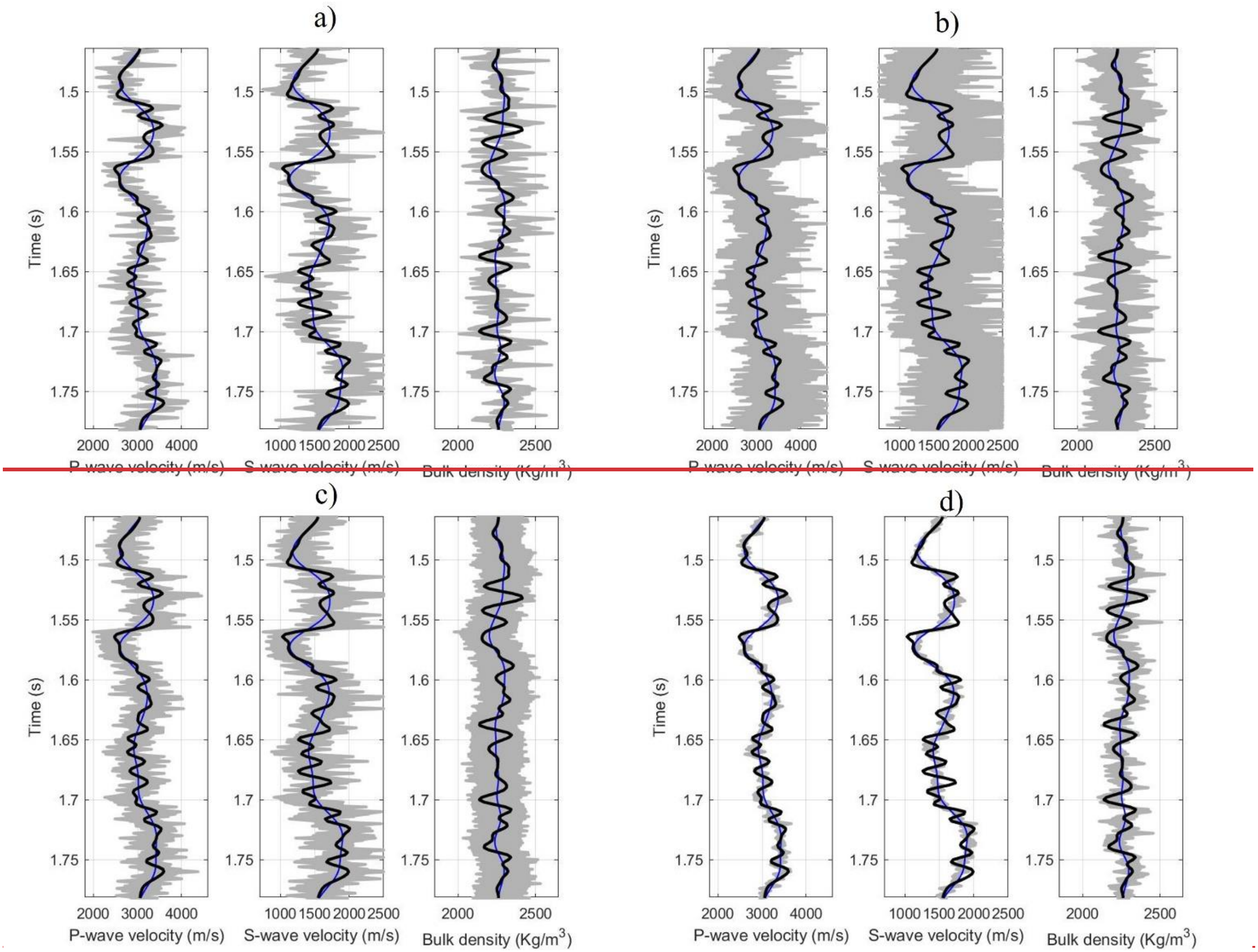
Figure 1: Progress with iterations in the MH (blue line), HMC (red line), MALA (green line) and Lip-MALA (black line) sampling algorithms for synthetic test.

The model settings were modified during the sampling phase, but remain within the probability function, as shown in Figure 2. Figure 3 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 12 shows the statistical parameters of mean and standard deviation (SdSD) which we will compare then with the data obtained from the inference in the different algorithms used.

Table 12: 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

.


a)
b)
c)
d)
Time (s)
1.5
1.55
1.6
1.65
1.7
1.75
2000 3000 4000
P-wave velocity (m/s)
1000 1500 2000 2500
S-wave velocity (m/s)
2000 2500
Bulk density (Kg/m^3)

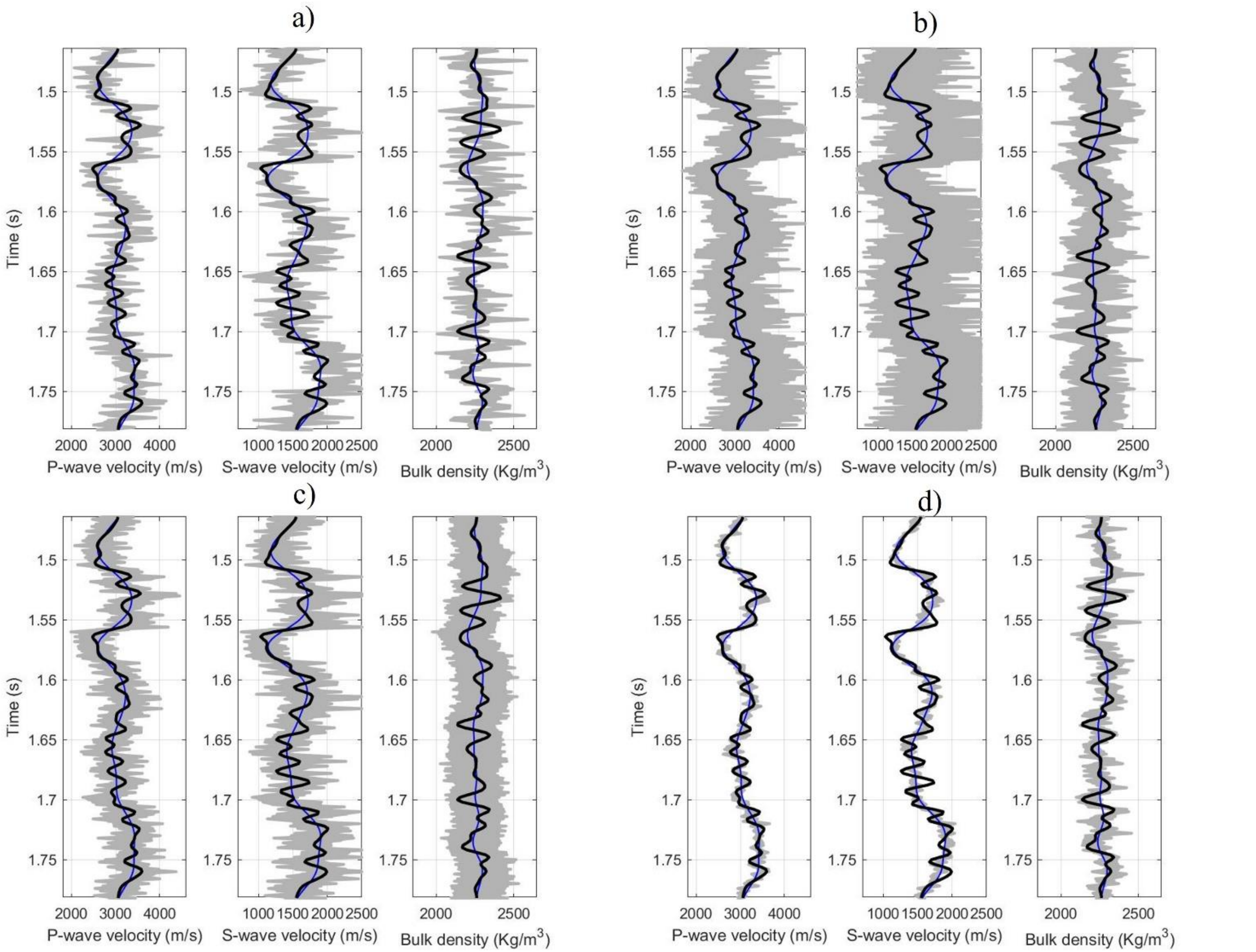
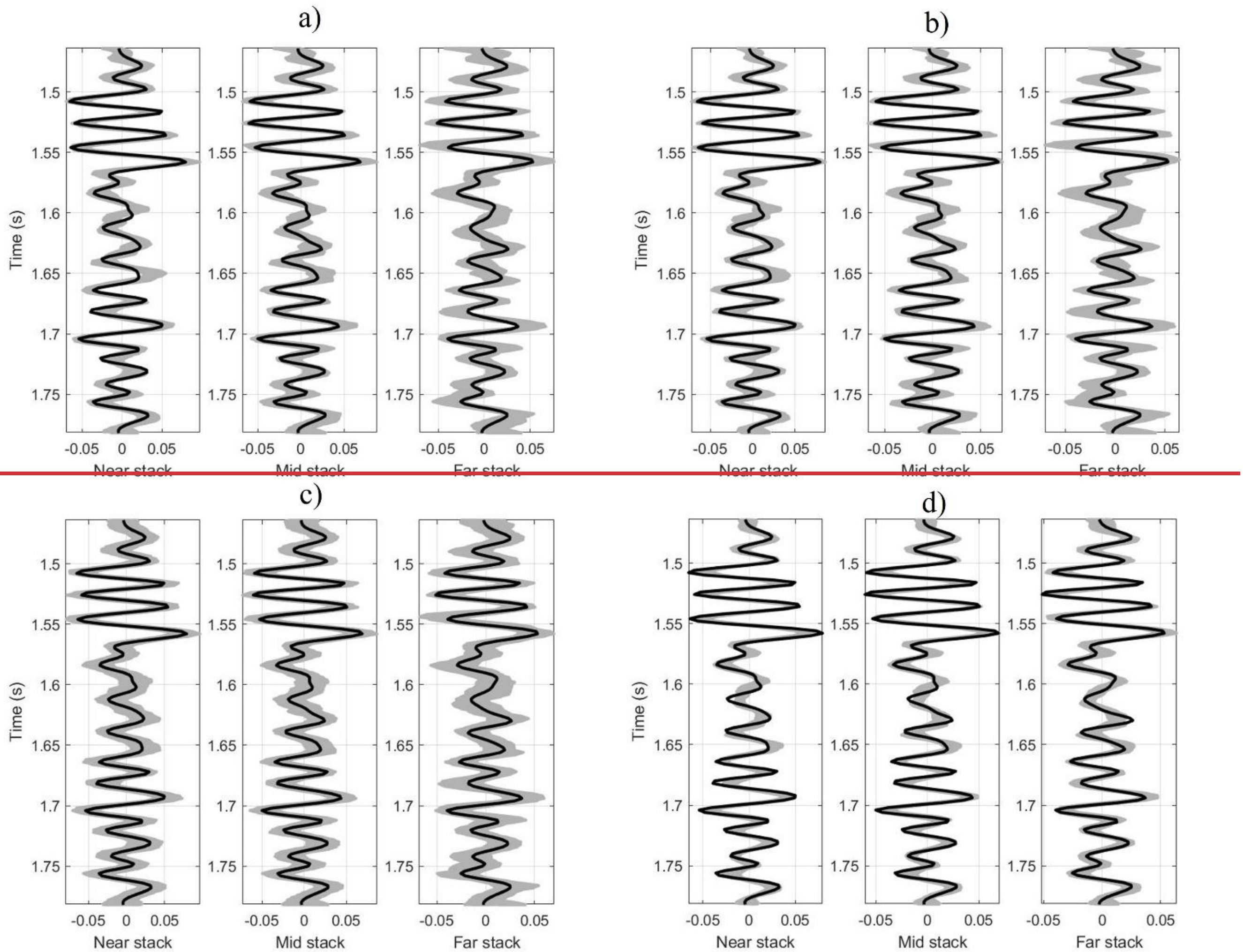


Figure 2: True data (black line), prior model (blue line), and accepted realizations of the model (gray) for the synthetic test where a) MH, b) HMC, c) MALA, and d) Lip-MALA.


a)
b)
c)
d)
Time (s)
1.5
1.55
1.6
1.65
1.7
1.75
-0.05
0
0.05
Near stack
Mid stack
Far stack

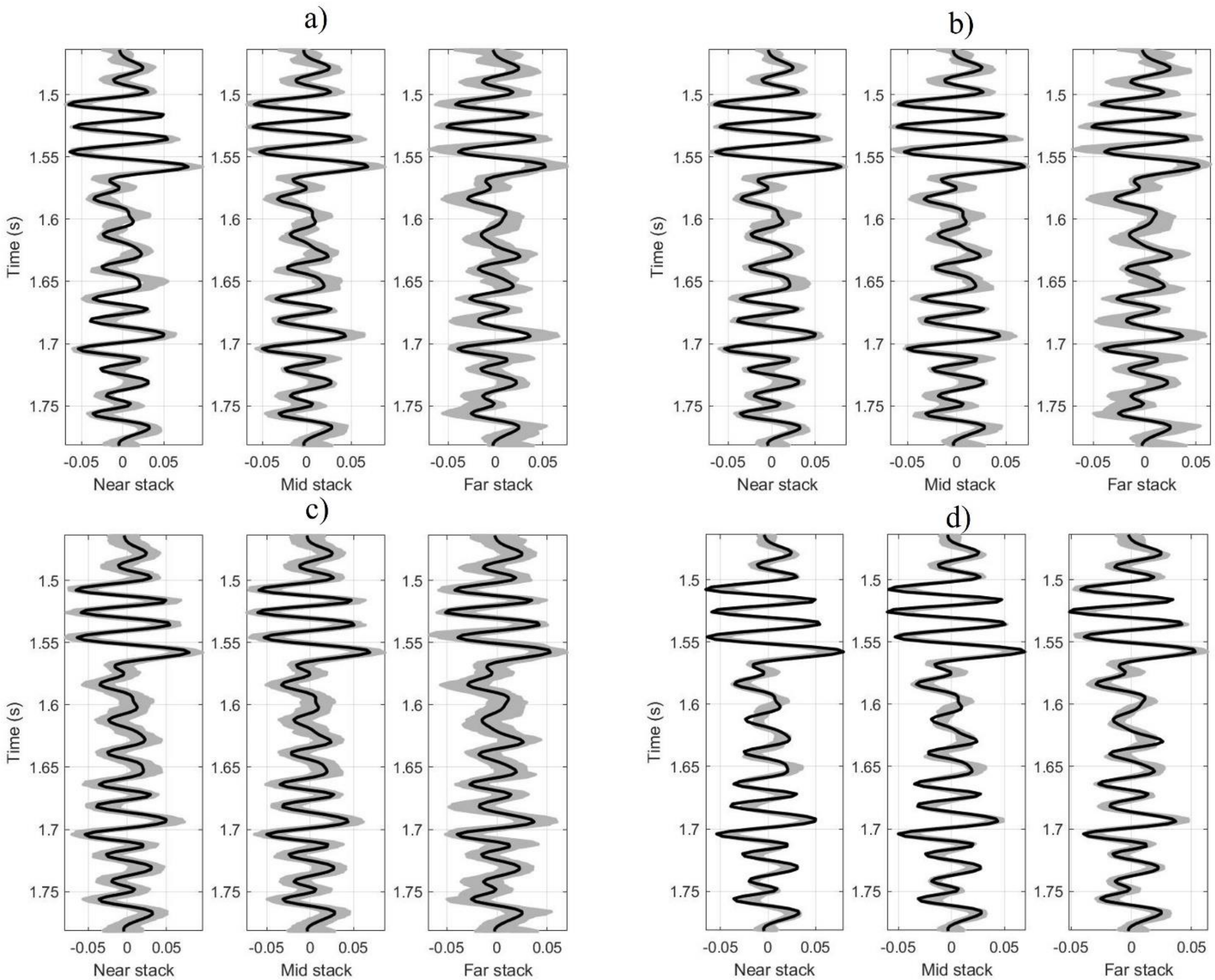
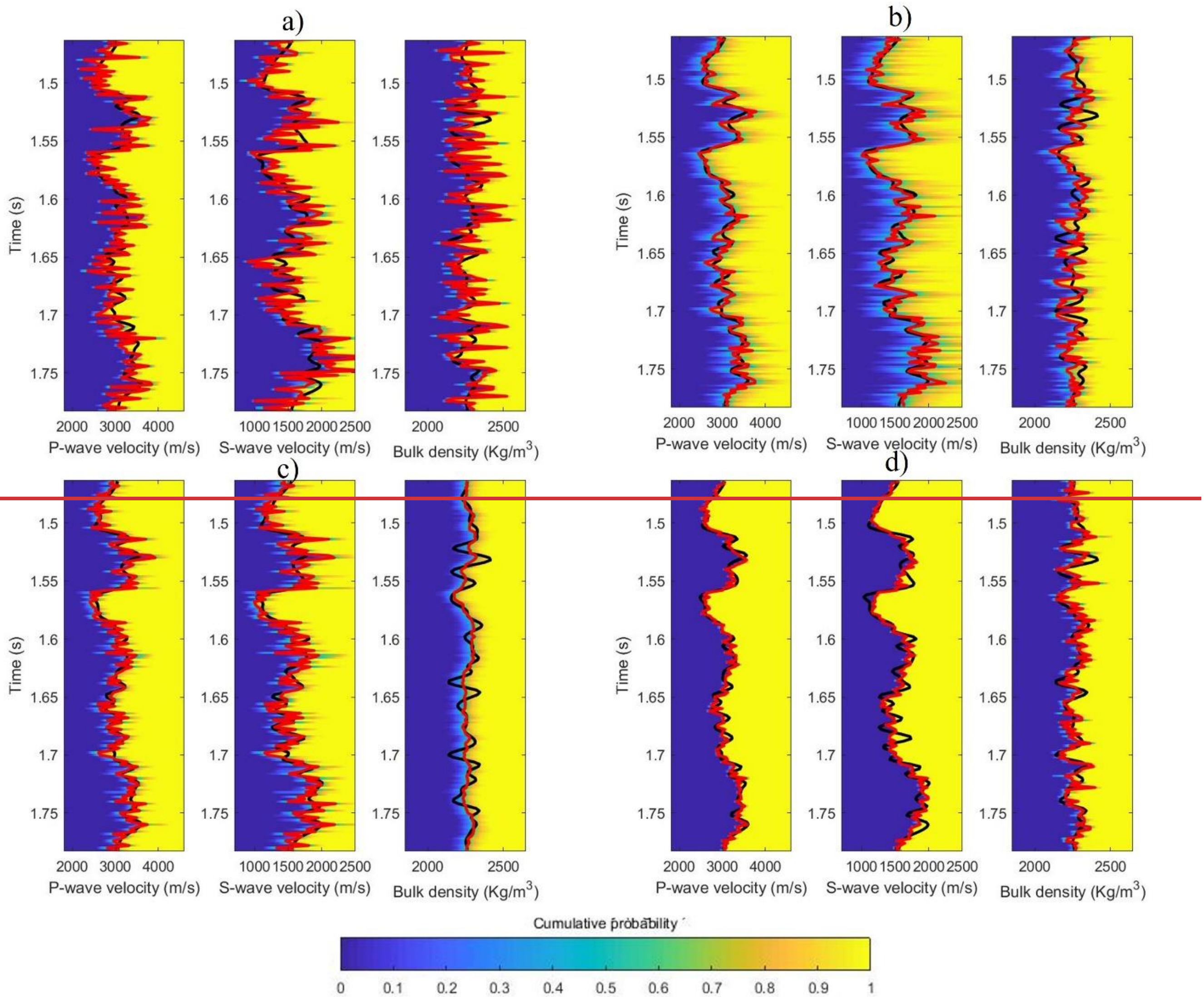


Figure 3: Observed seismic data (black line) and seismic traces obtained from the accepted model realizations (gray) for the synthetic test where a) MH, b) HMC, c) MALA, and d) Lip-MALA.


a)
b)
c)
d)
Time (s)
1.5
1.55
1.6
1.65
1.7
1.75
2000 3000 4000
P-wave velocity (m/s)
1000 1500 2000 2500
S-wave velocity (m/s)
2000 2500
Bulk density (Kg/m^3)
Cumulative probability
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

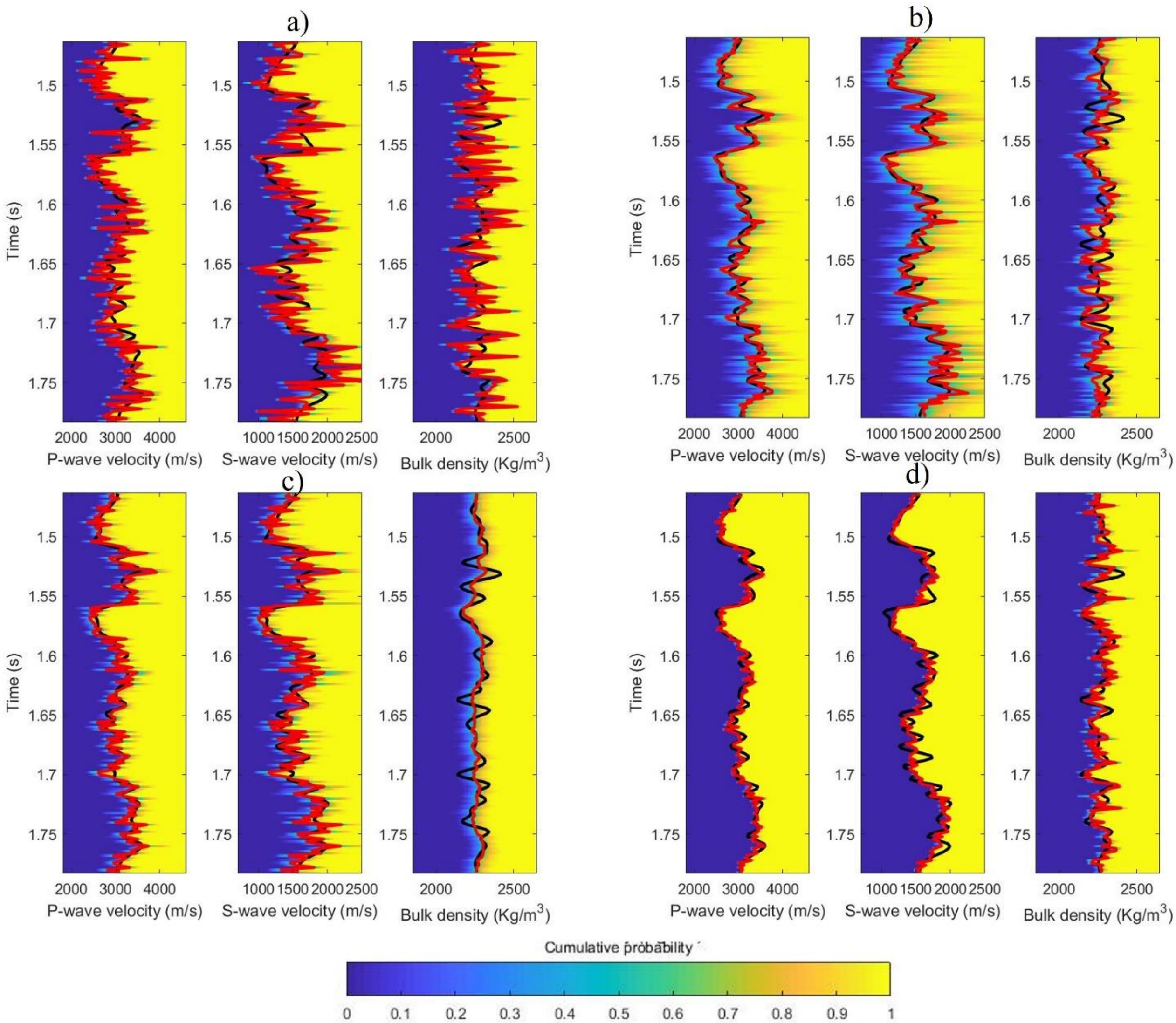


Figure 4: Marginal cumulative probability distributions (color map), true data (black line) and seismic inversion model result (red line) for the synthetic test where a) MH, b) HMC, c) MALA, and d) Lip-MALA.

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 4 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 ~~2~~3 summarizes the performance of the different algorithms tested in predicting P-wave and S-wave velocities and density. The mean, Standard Deviation (~~Sd~~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 ~~2~~3: Statistical parameters for the results obtained for algorithms tested for the synthetic test.

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

Table 34 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 34: 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 45 shows the summary of mESS and minESS obtained for each method.

Table 45: 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

5.2 Application to real data

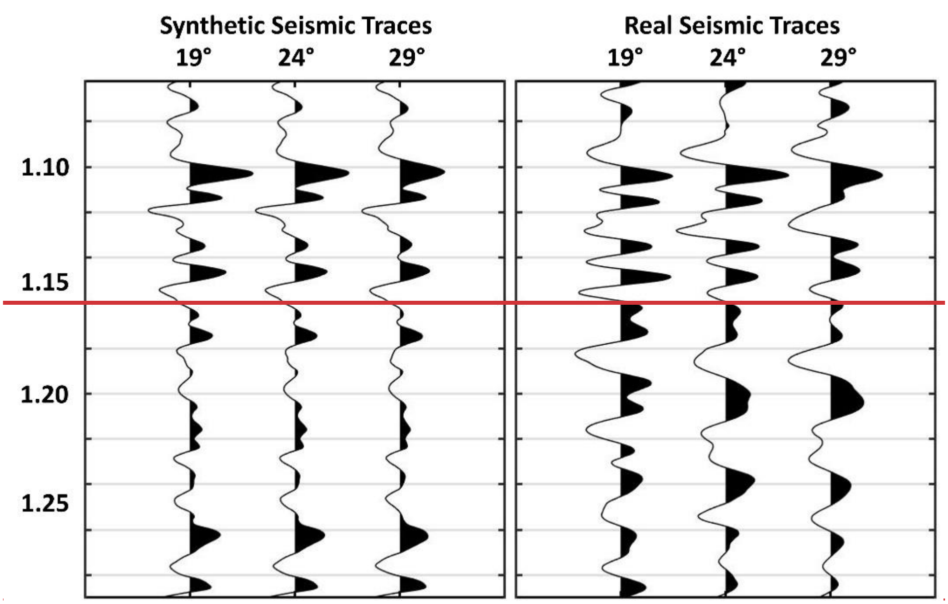
To demonstrate the effectiveness of the algorithms, we applied them to a real dataset of an oil field in eastern Venezuela. The site is located in a formation dominated by clastic rocks, a type of sedimentary rock characterized by alternating layers of sand and shale. The fluids in the pore spaces of 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 56 presents the descriptive statistics, including mean and standard deviation (~~Sd~~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 ~~5~~6: Mean and Standard deviation of elastic parameters used for real data.

Parameter	Mean	Sd
V_p (m/s)	2642.92	249.40
V_s (m/s)	1289.86	205.84
ρ (Kg/m^3)	2180.06	111.89

The seismic traces were obtained from partial stacks for the angles ~~θ_1 =~~ 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. The synthetic trace obtained was correlated with observed traces for seismic well tie (see figure 5) 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, 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, 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.



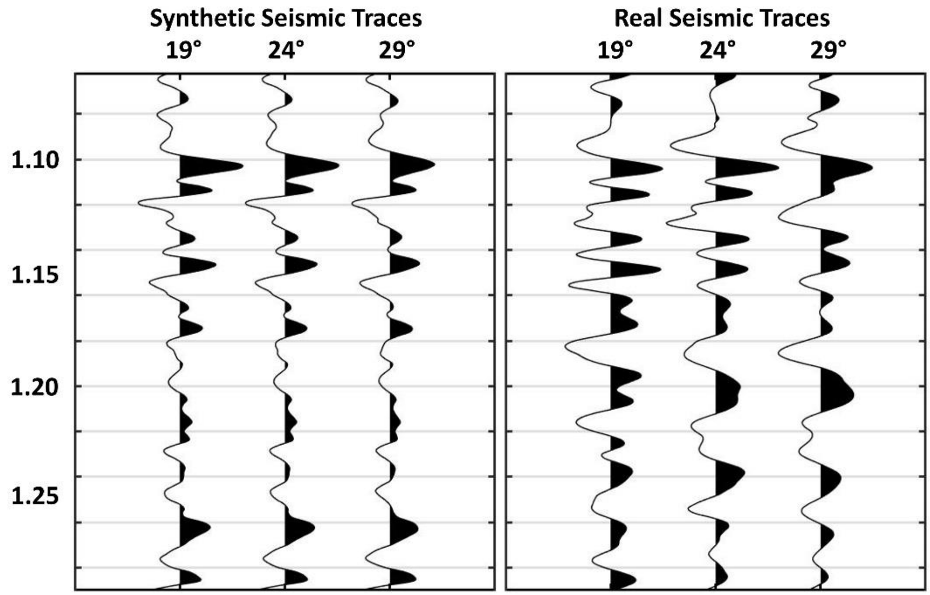
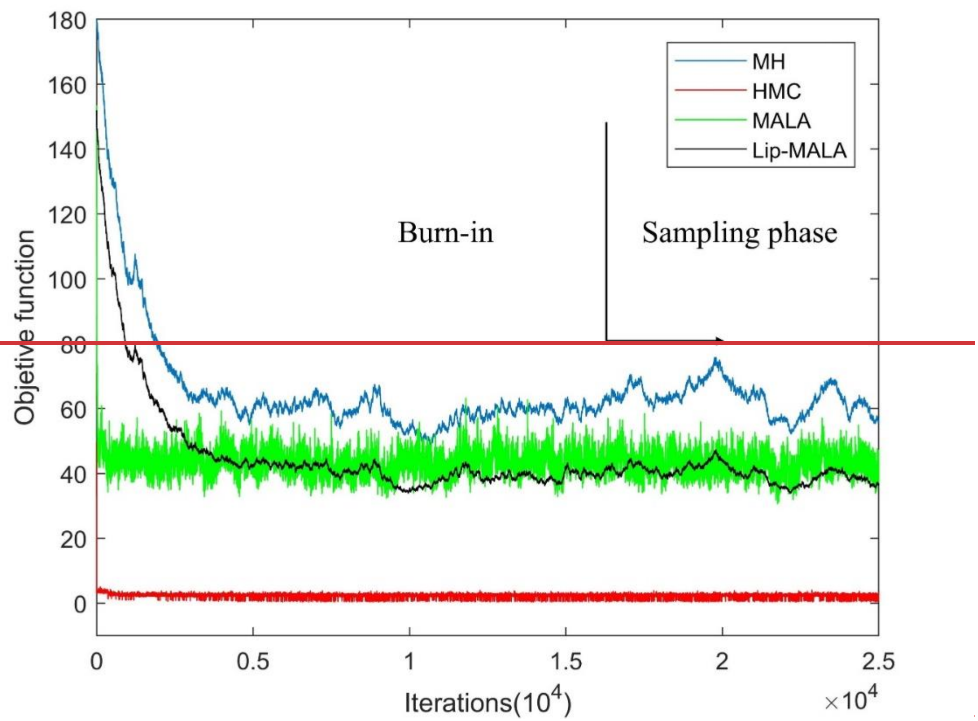


Figure 5: Seismic well tie for real data used.



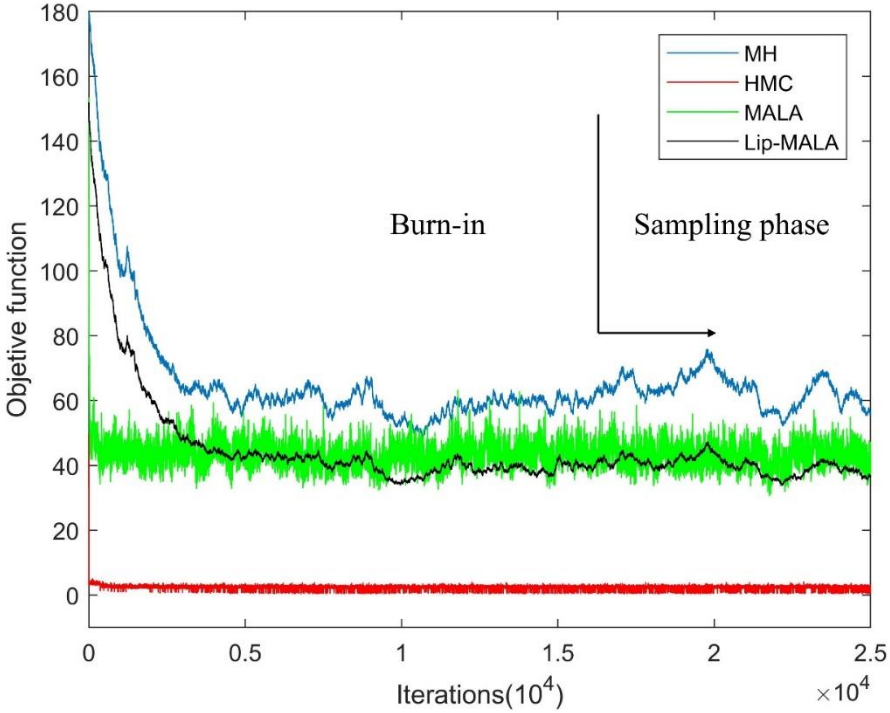
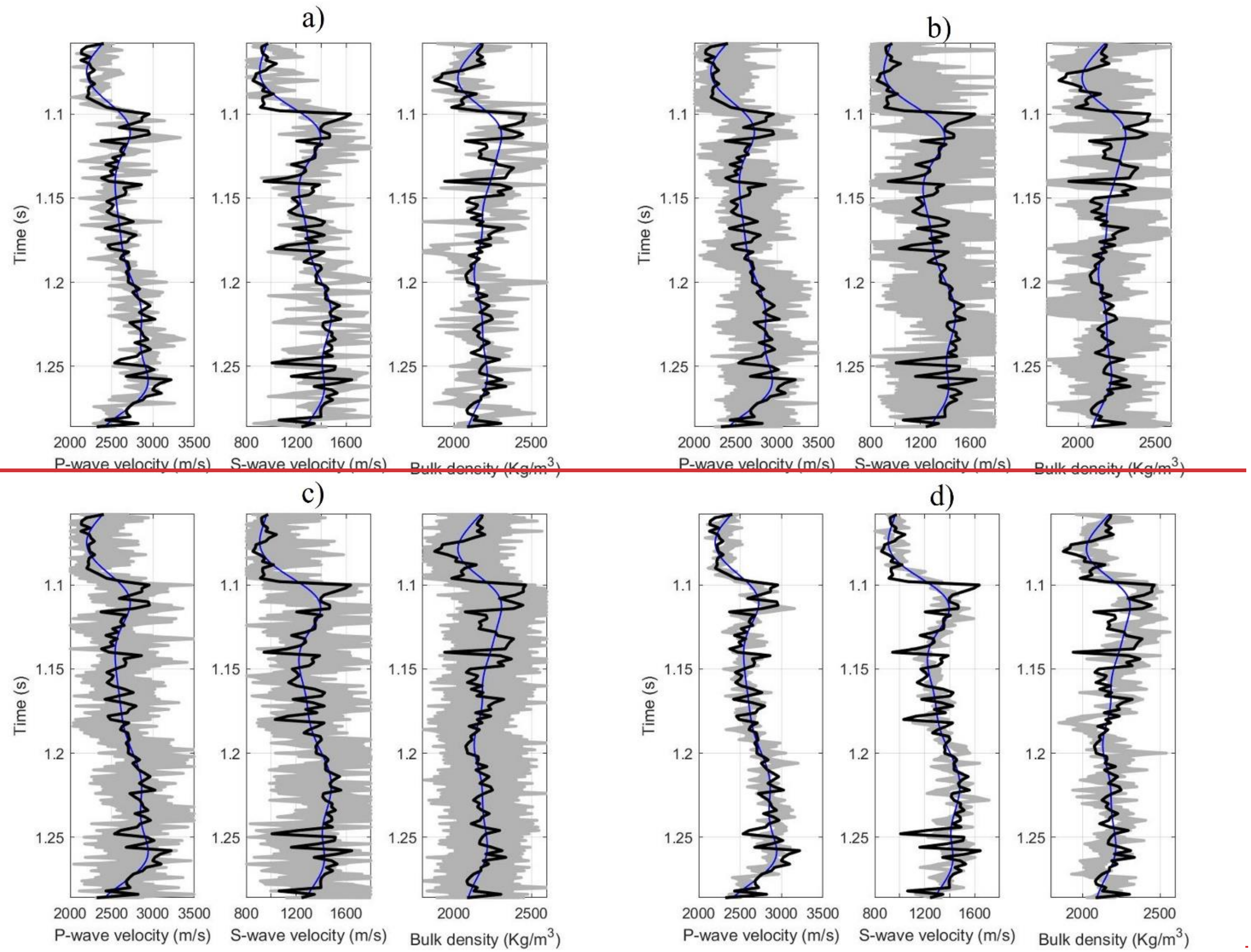


Figure 6: Progress with iterations in the MH (blue line), HMC (red line), MALA (green line) and Lip-MALA (black line) sampling algorithms for real data.


a)
b)
c)
d)
Time (s)
1.1
1.15
1.2
1.25
2000 2500 3000 3500
P-wave velocity (m/s)
800 1200 1600
S-wave velocity (m/s)
2000 2500
Bulk density (Kg/m^3)

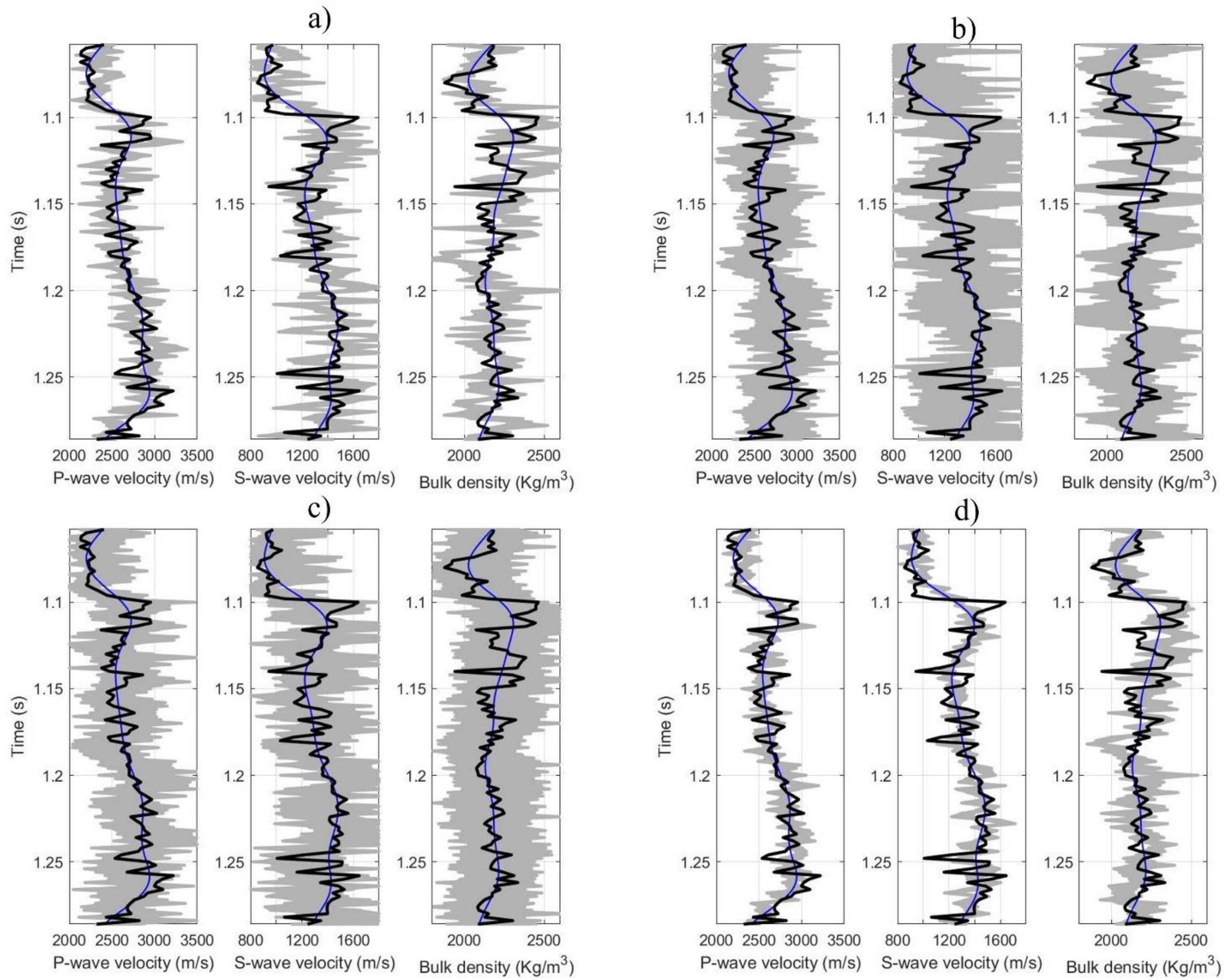
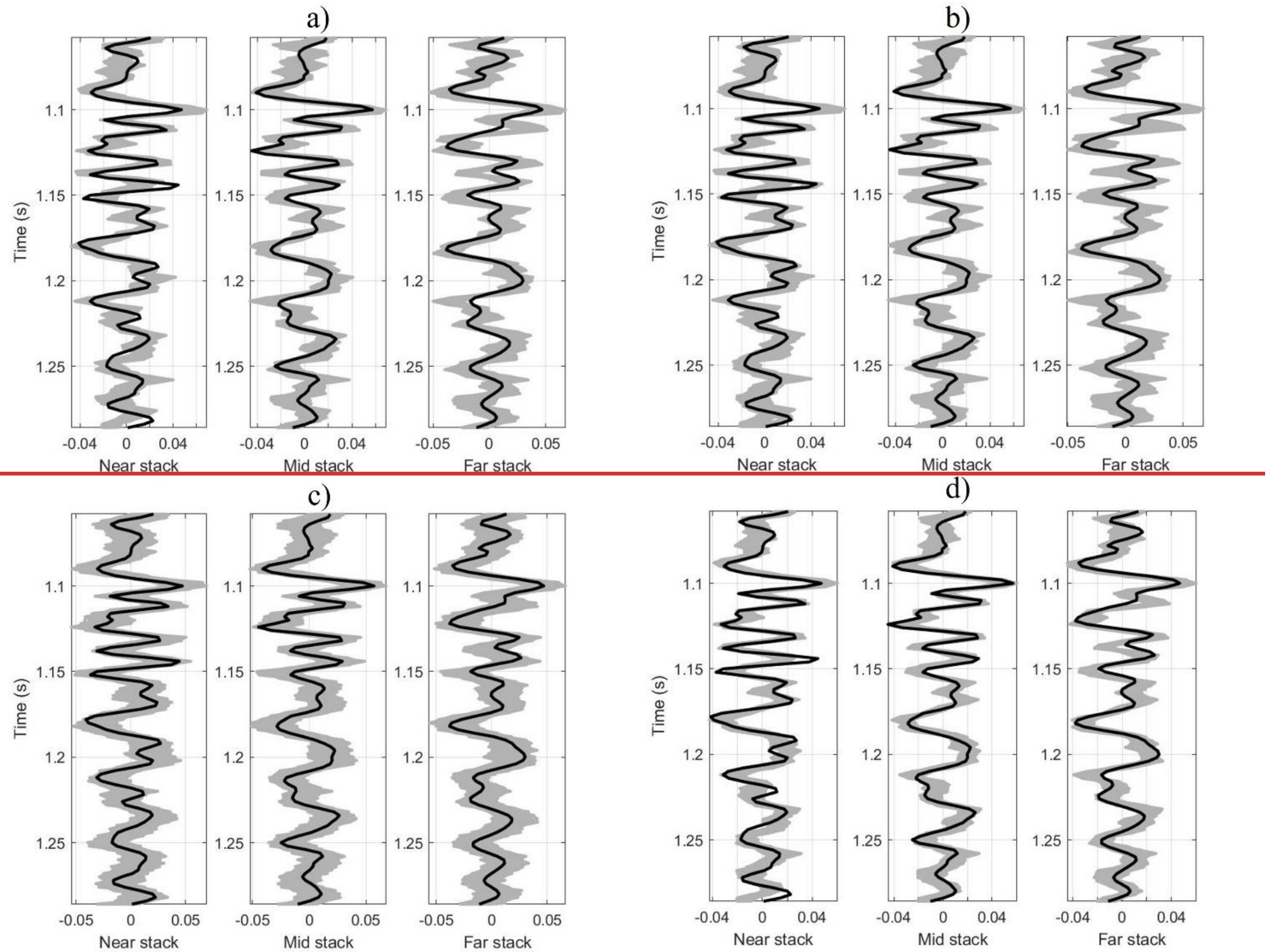


Figure 7: True data (black line), prior model (~~red~~ blue line), and accepted realizations of the model (gray) for real data where a) MH, b) HMC, c) MALA, and d) Lip-MALA.


a)
b)
c)
d)
Time (s)
1.1
1.15
1.2
1.25
-0.04
0
0.04
-0.05
0.05
Near stack
Mid stack
Far stack

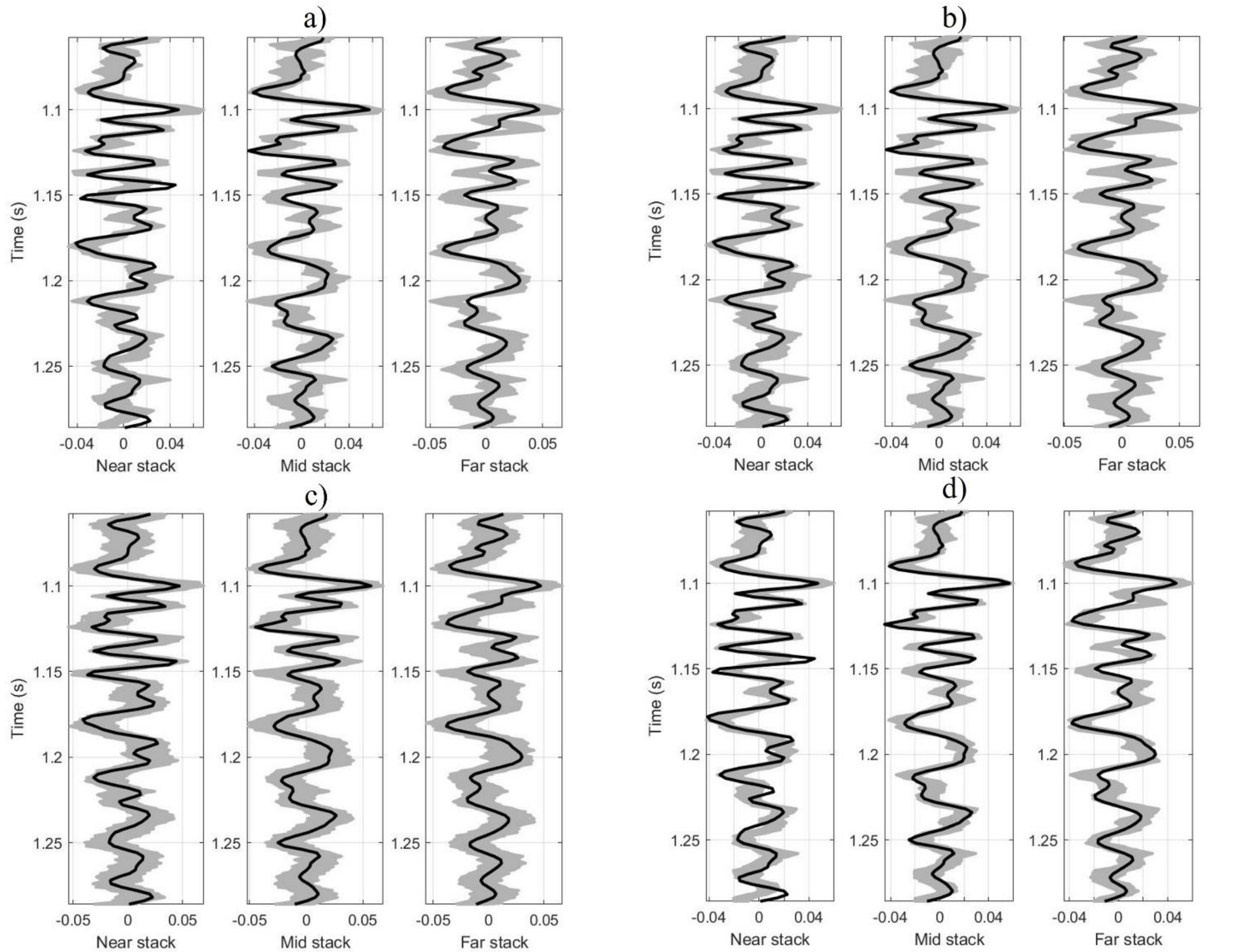
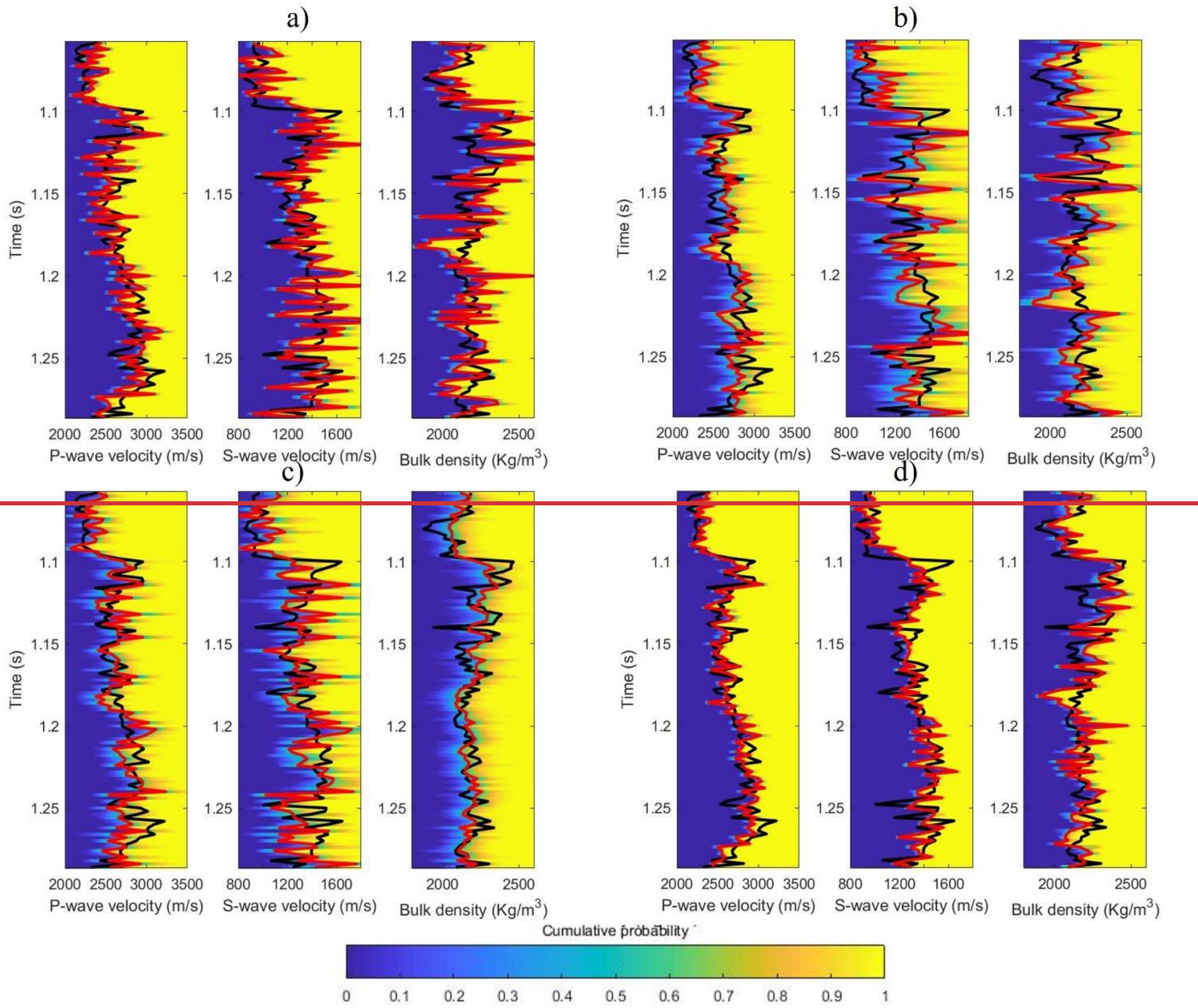


Figure 8: Observed seismic data (black line) and seismic traces obtained from the accepted model realizations (gray) for real data where a) MH, b) HMC, c) MALA, and d) Lip-MALA.


a)
b)
c)
d)
Time (s)
1.1
1.15
1.2
1.25
2000 2500 3000 3500
P-wave velocity (m/s)
800 1200 1600
S-wave velocity (m/s)
2000 2500
Bulk density (Kg/m^3)
Cumulative probability
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

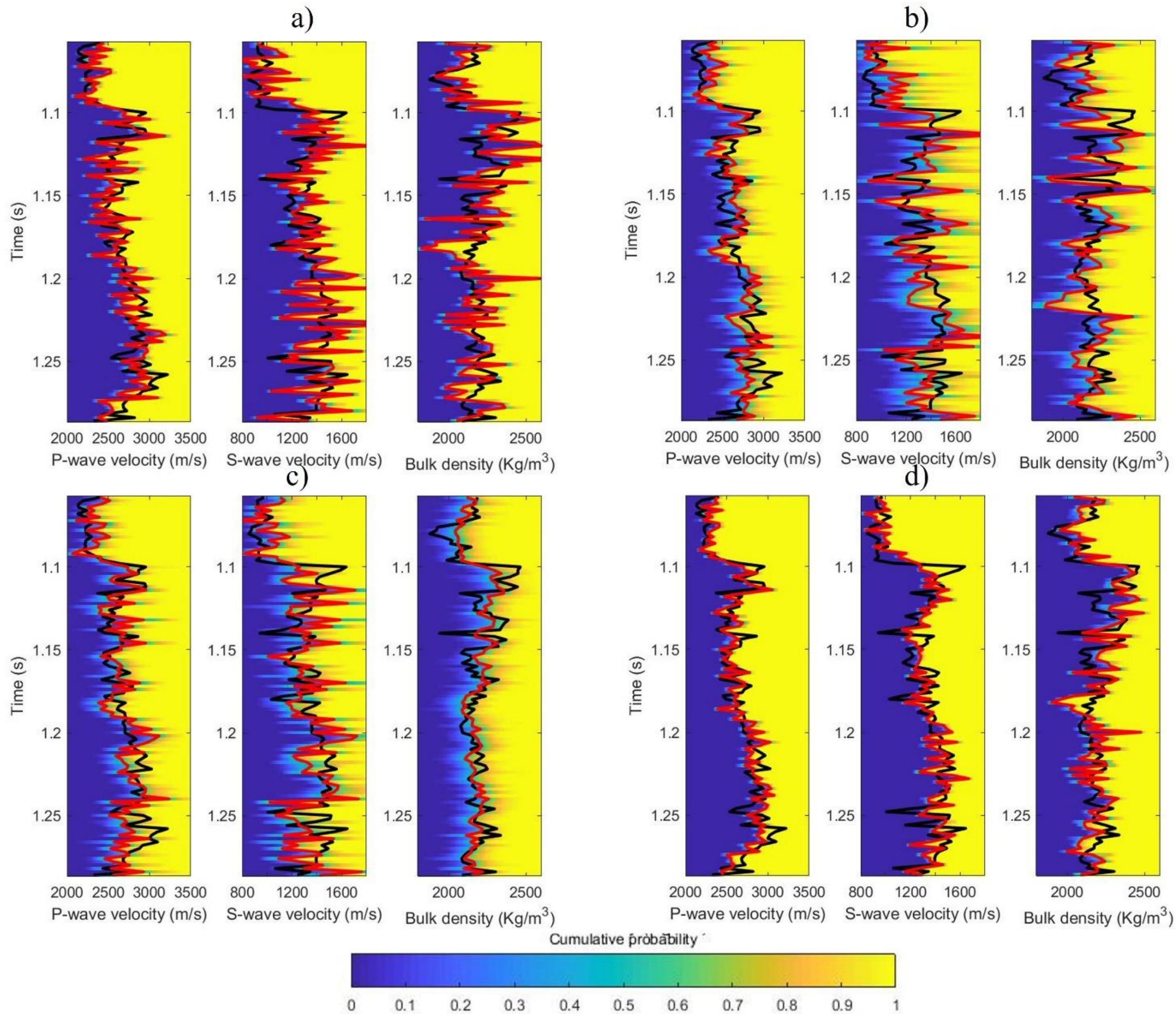


Figure 9: Marginal cumulative probability distributions (color map), true data (black line) and seismic inversion model result (red line) for real data where a) MH, b) HMC, c) MALA, and d) Lip-MALA.

The model settings were adjusted during the sampling phase, ensuring they remained within the probability function (Figure 7). Figure 8 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

34

were derived by averaging the model performances across the sampling phase. Figure 9 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 ~~6~~7 summarizes the performance of the tested algorithms in predicting P-wave and S-wave velocities and density. The mean, Standard Deviation (~~Sd~~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 ~~6~~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 ~~7~~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 ~~7~~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 ~~8~~9 summarizes the mESS and minESS values obtained for each method.

Table ~~8~~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

5.3 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.

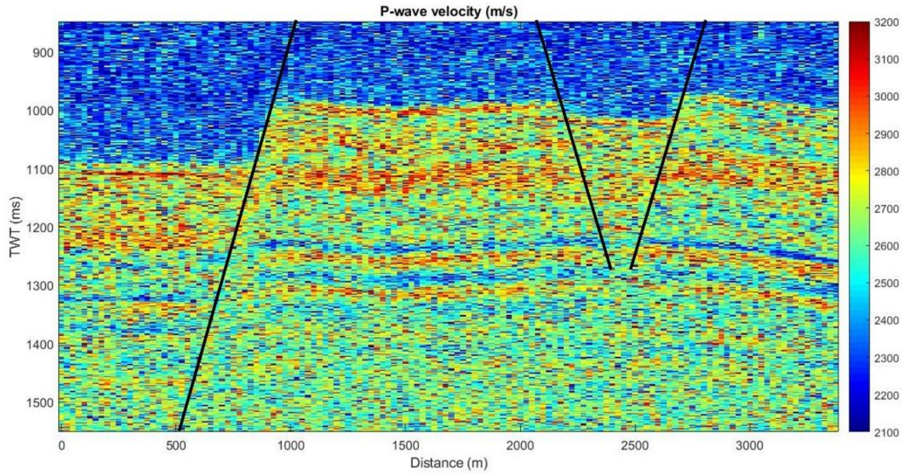
The study area corresponds to a sector of the Eastern Basin of Venezuela, characterized by clastic lithology with alternating sandstones and shales. The seismic data used was acquired using conventional reflection techniques and subsequently reprocessed to generate prestack gathers organized by incidence angle. Three partial stacks were selected, corresponding to central angles of approximately 19°, 24°, and 29°, and used in the inverse process.

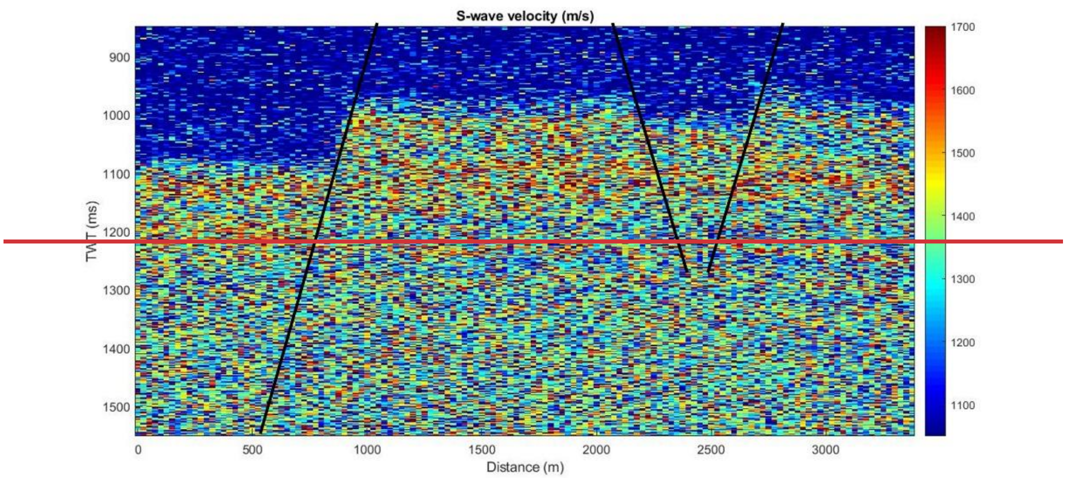
The two-dimensional inversion grid was designed with a vertical resolution of 351 samples, corresponding to the two-way time axis from 850 to 1550 ms, and a horizontal extent of 136 cells from 0 to 340 m. Three fundamental elastic parameters were considered: V_p, V_s and ρ. The prior model was built 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.

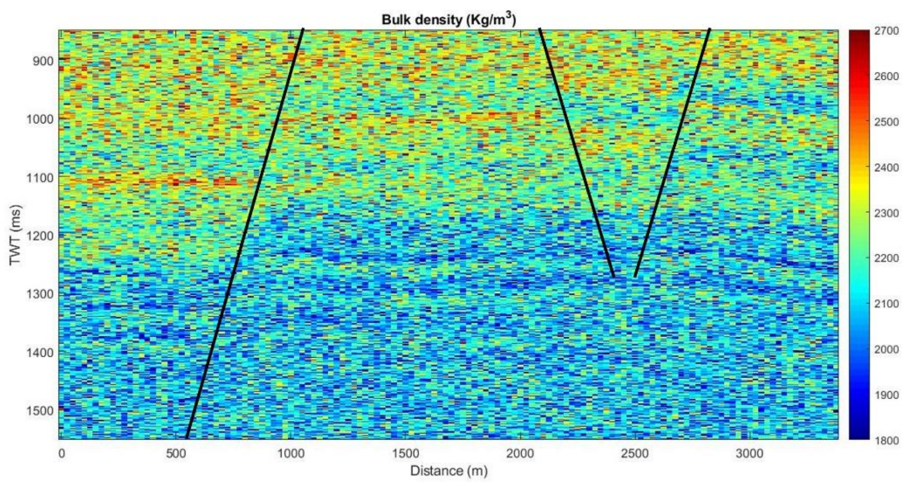
Table ~~9: Convergence test for~~10: Posterior standard deviations (2D real data~~.~~).

Method	**Sd V_p** ~~(m/s)~~ (m/s)	**Sd V_s** ~~(m/s)~~ (m/s)	**Sd ρ** ~~(Kg/m^3)~~ (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).


P-wave velocity (m/s)
TWT (ms)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
Distance (m)
3200
3100
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100


S-wave velocity (m/s)
TWT (ms)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
Distance (m)
1700
1600
1500
1400
1300
1200
1100


Bulk density (Kg/m3)
TWT (ms)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
Distance (m)
2700
2600
2500
2400
2300
2200
2100
2000
1900
1800

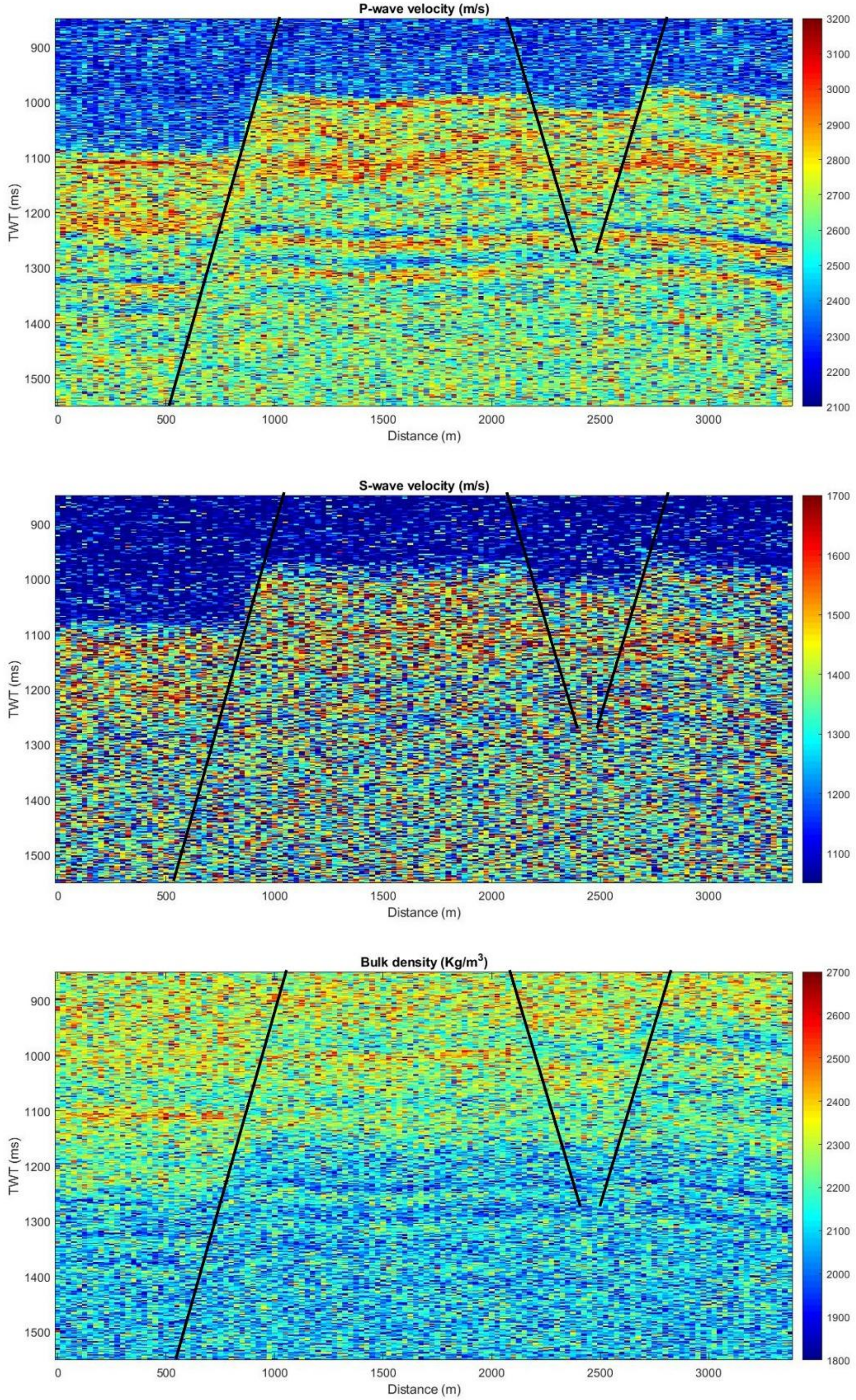
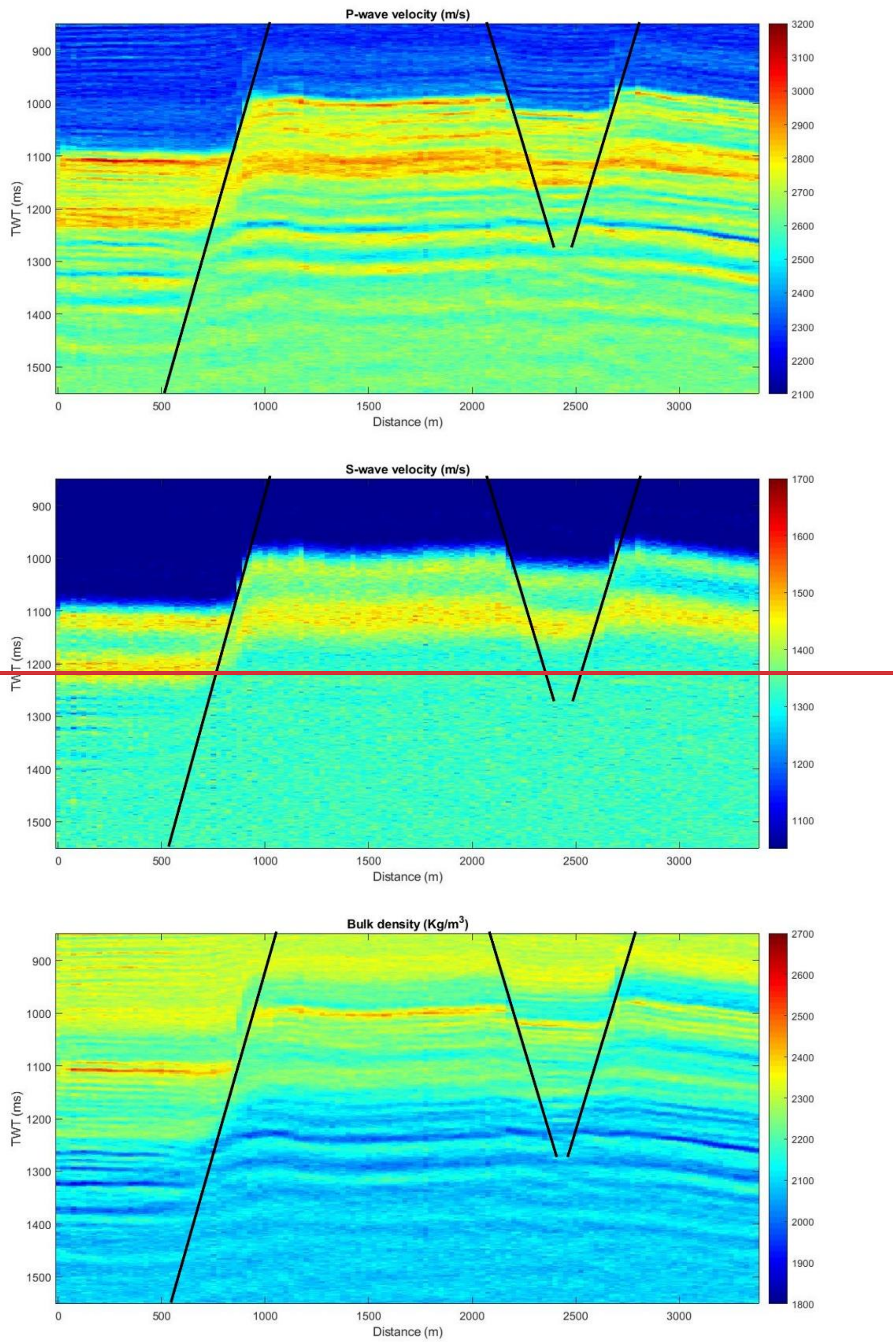

P-wave velocity (m/s)
TWT (ms)
Distance (m)
S-wave velocity (m/s)
TWT (ms)
Distance (m)
Bulk density (Kg/m3)
TWT (ms)
Distance (m)

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


P-wave velocity (m/s)
TWT (ms)
Distance (m)
S-wave velocity (m/s)
TWT (ms)
Distance (m)
Bulk density (Kg/m3)
TWT (ms)
Distance (m)

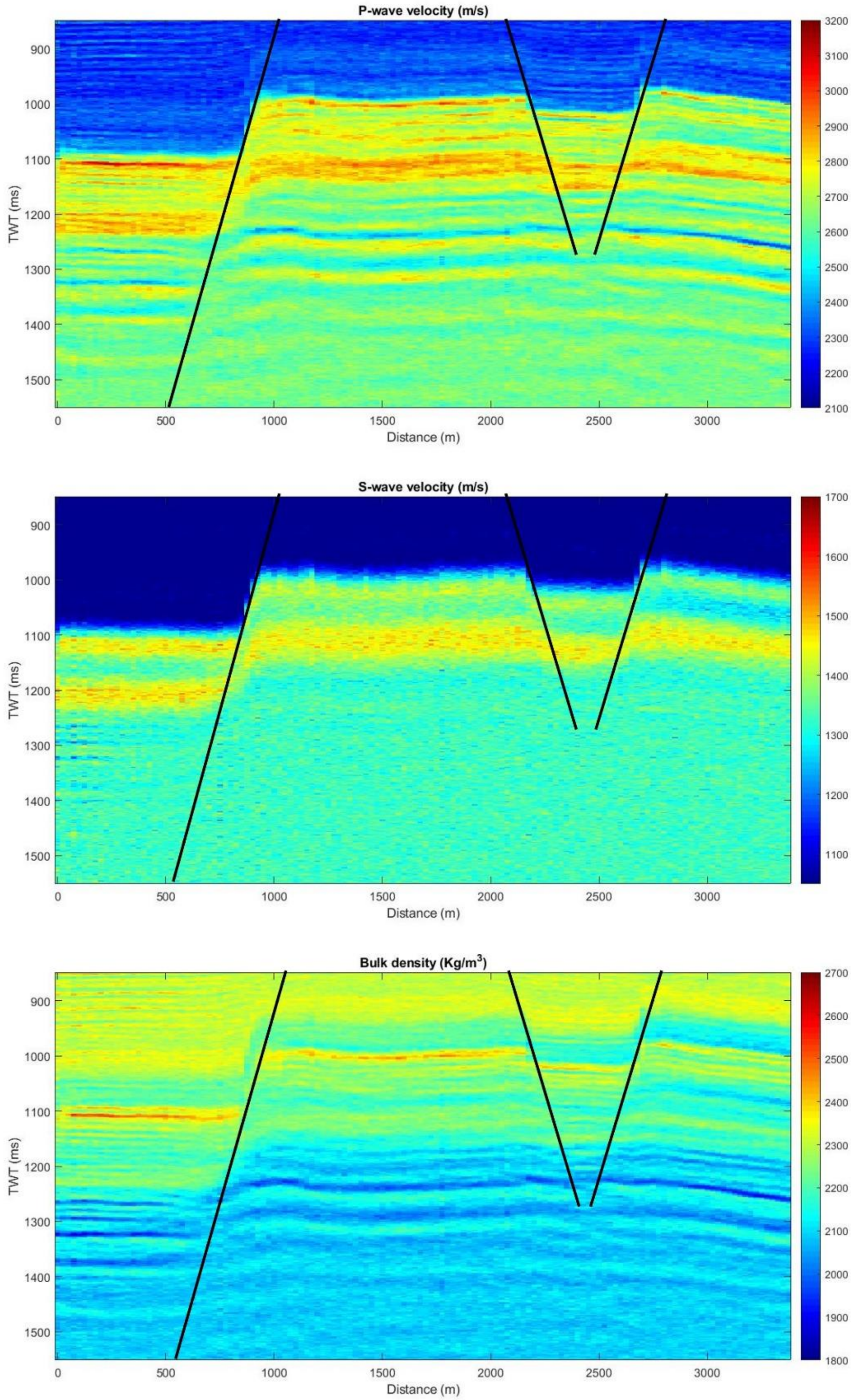
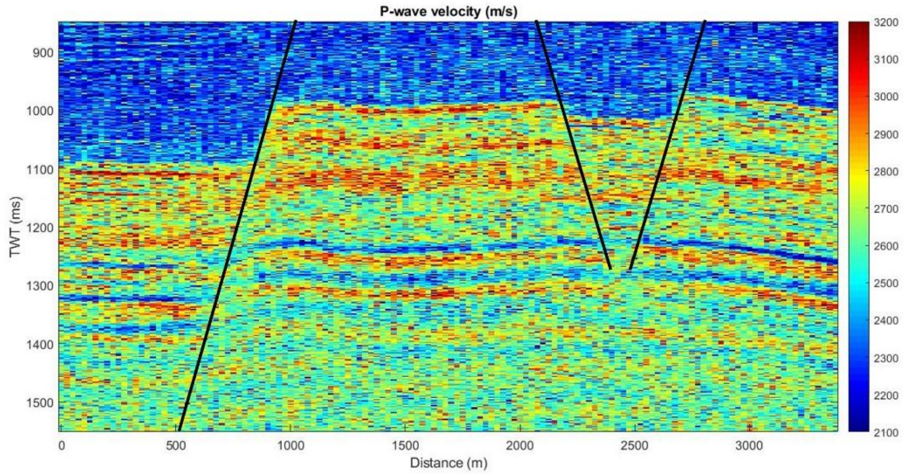
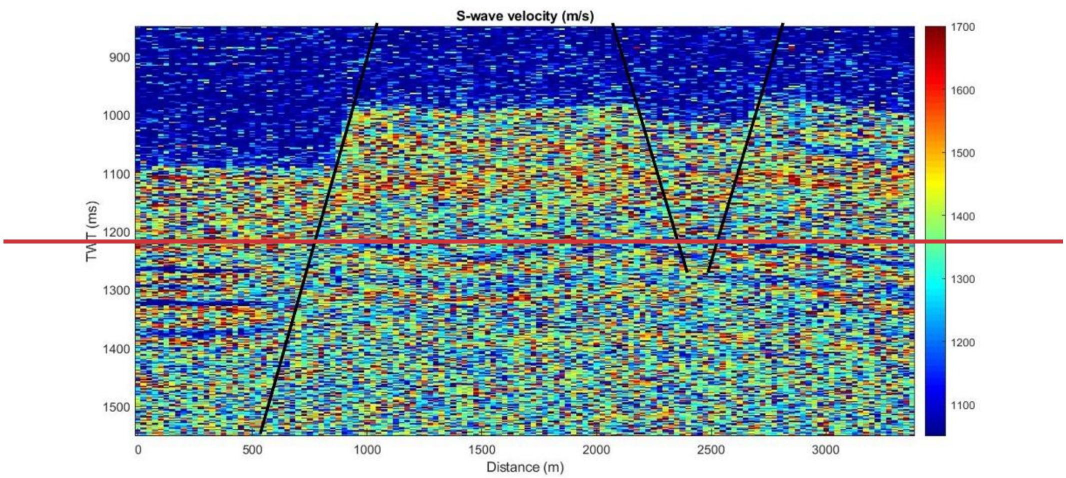
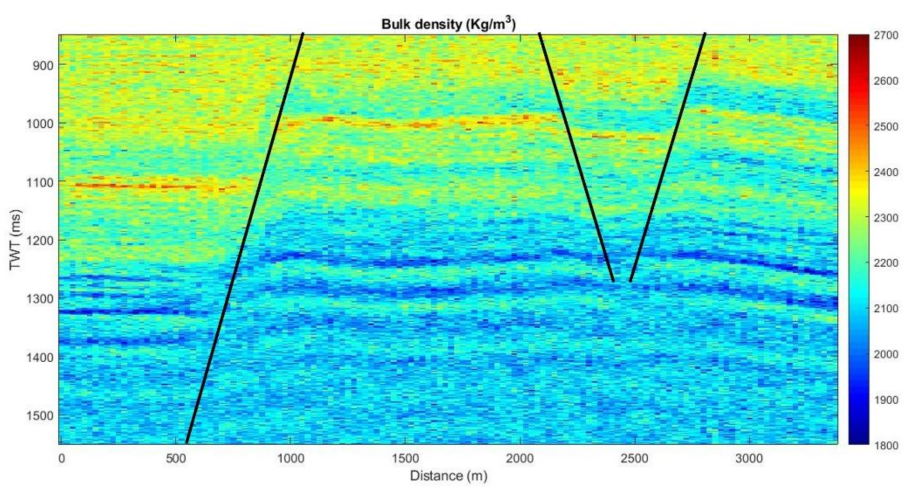

P-wave velocity (m/s)
TWT (ms)
Distance (m)
S-wave velocity (m/s)
TWT (ms)
Distance (m)
Bulk density (Kg/m^3)
TWT (ms)
Distance (m)

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


P-wave velocity (m/s)
TWT (ms)
Distance (m)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
3200
3100
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100


S-wave velocity (m/s)
TWT (ms)
Distance (m)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
1700
1600
1500
1400
1300
1200
1100


Bulk density (Kg/m3)
TWT (ms)
Distance (m)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
2700
2600
2500
2400
2300
2200
2100
2000
1900
1800

44

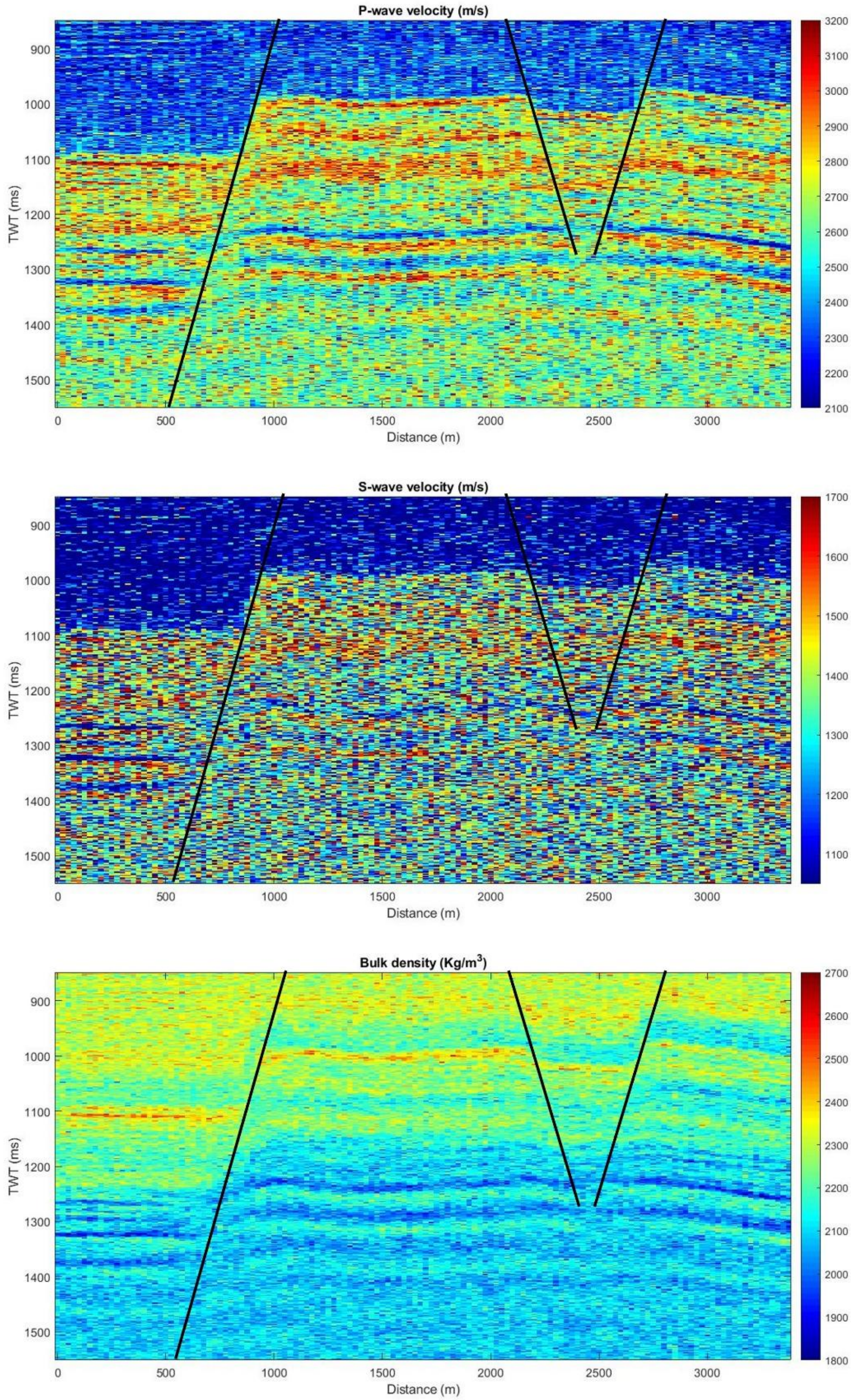
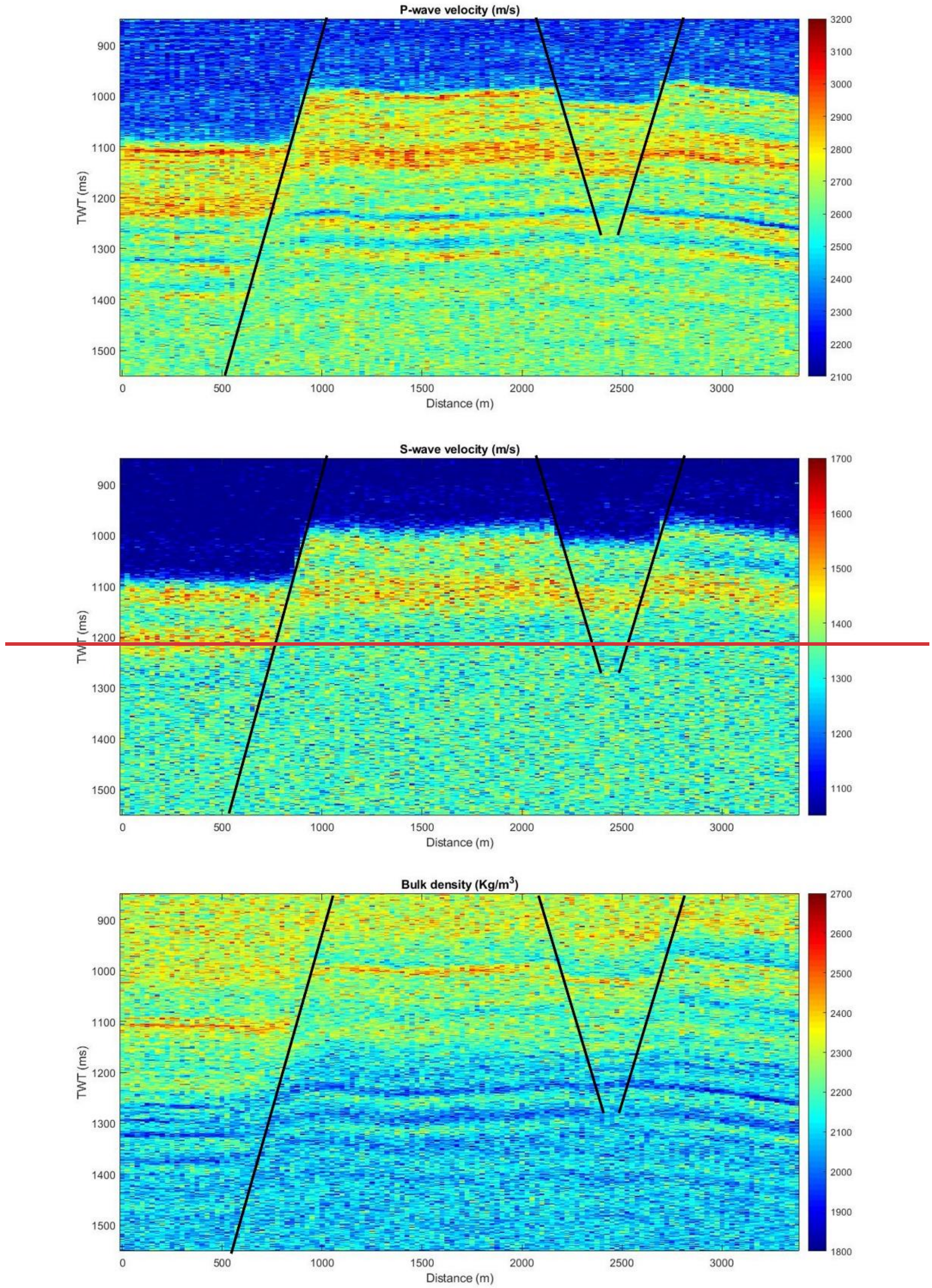

P-wave velocity (m/s)
TWT (ms)
Distance (m)
S-wave velocity (m/s)
TWT (ms)
Distance (m)
Bulk density (Kg/m³)
TWT (ms)
Distance (m)

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


P-wave velocity (m/s)
TWT (ms)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
Distance (m)
3200
3100
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
S-wave velocity (m/s)
TWT (ms)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
Distance (m)
1700
1600
1500
1400
1300
1200
1100
Bulk density (Kg/m3)
TWT (ms)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
Distance (m)
2700
2600
2500
2400
2300
2200
2100
2000
1900
1800

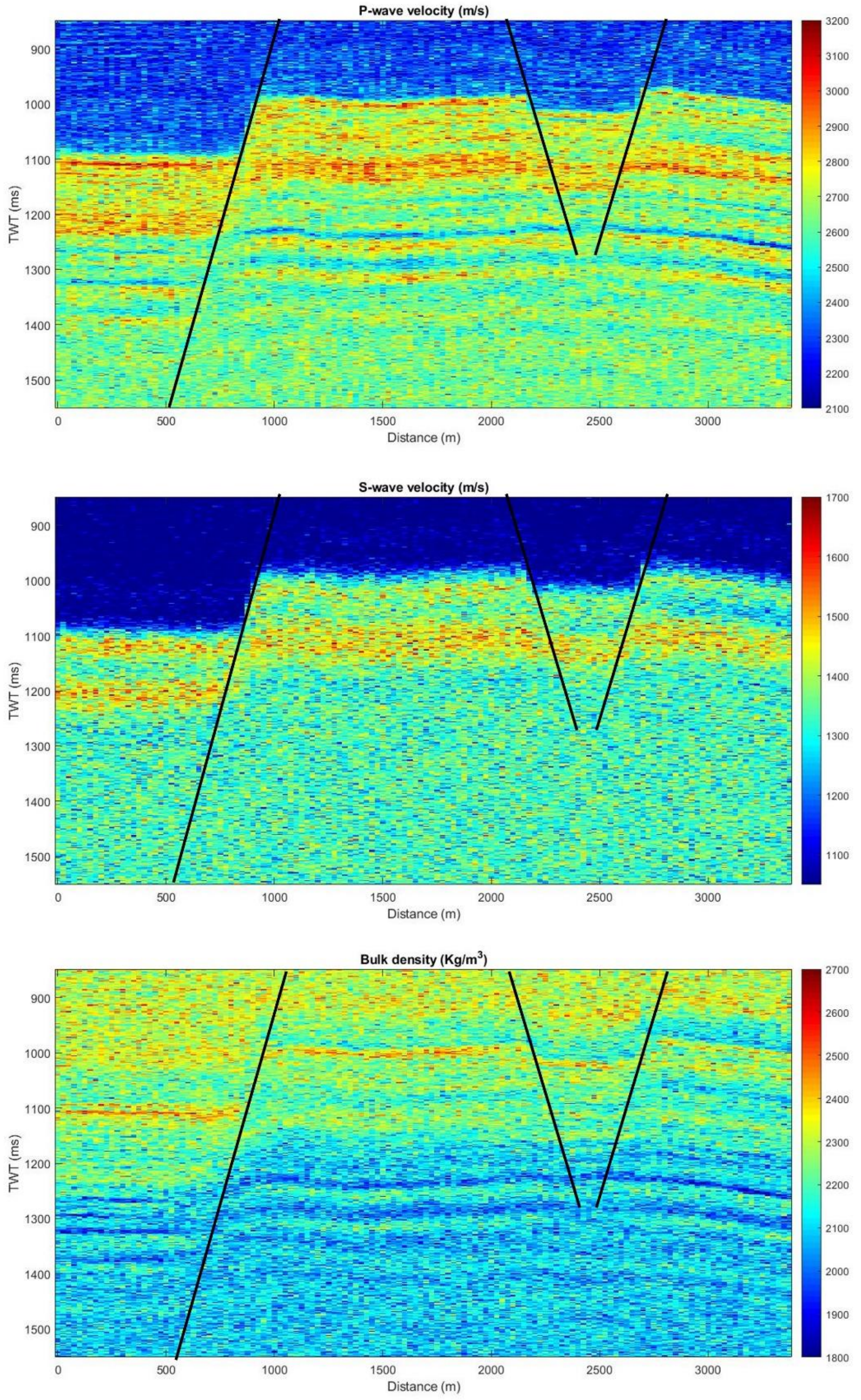

P-wave velocity (m/s)
TWT (ms)
Distance (m)
900
1000
1100
1200
1300
1400
1500
0
500
1000
1500
2000
2500
3000
3200
3100
3000
2900
2800
2700
2600
2500
2400
2300
2200
2100
S-wave velocity (m/s)
1700
1600
1500
1400
1300
1200
1100
Bulk density (Kg/m3)
2700
2600
2500
2400
2300
2200
2100
2000
1900
1800

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

Quantitative analysis of the figures 10 to 13 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 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) ~~highlight~~high-lights 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 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 () and density (), 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 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.

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.

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., 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. https://arxiv.org/abs/1504.01896v3

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., 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.

54