

## **Response to Anonymous Referee #1**

*This is mostly a survey paper on MCMC methods that does not contain much novelty,*

R: Thank you for your comment, but we do not agree with your opinion, because for us it is novelty because:

### 1. Comparative Analysis of Advanced MCMC Algorithms:

The study systematically evaluates and compares four Markov Chain Monte Carlo (MCMC) algorithms—Metropolis-Hastings (MH), Hamiltonian Monte Carlo (HMC), Metropolis-Adjusted Langevin Algorithm (MALA), and a locally Lipschitz adaptive variant (Lip-MALA)—for prestack seismic inversion. While these methods are individually well-documented, the comparative assessment within the specific context of prestack seismic data, using both synthetic and real-world datasets, provides a fresh perspective on their relative strengths and weaknesses.

### 2. Comprehensive Evaluation Metrics:

The study not only evaluates algorithmic performance in terms of accuracy (e.g., RMSE and correlations) but also considers computational efficiency (execution time and acceptance rates) and convergence behavior (mESS vs. minESS).

### 3. Application to Real-World Data:

The research validates the proposed methods using real seismic data from an oil field in eastern Venezuela, demonstrating practical applicability and robustness in resolving P-wave, S-wave velocities, and density.

### 4. Insightful Findings for Practical Applications:

The findings provide actionable insights: gradient-based methods (HMC, MALA, Lip-MALA) are shown to outperform MH in velocity estimation, while MALA and Lip-MALA exhibit superior performance for density estimation. These results offer practical guidance for selecting inversion methods based on specific geophysical needs.

### 5. Potential for Future Advancements:

The study identifies promising directions for future work, such as integrating well log conditioning for enhanced vertical resolution and extending the framework to incorporate more advanced seismic simulations or petrophysical models.

***especially when considering the target (4) is a generalised Normal model.***

R: A normal distribution was used because this distribution incorporates the uncertainties in seismic data and model parameters. In seismic inversion, it is usually assumed that the likelihood of the data, the prior distribution of the parameters, and the observational error obey a Gaussian distribution. The goal is to obtain a posterior Gaussian distribution with desirable mathematical properties (Bosch, 2004, 2007; Yang et al., 2022; Li & Ben-Zion, 2023; Sun et al., 2024; Izzatullah et al., 2021).

- Bosch, M. (2004). The optimization approach to lithological tomography: Combining seismic data and petrophysics for porosity prediction. *Geophysics*, 69(5), 1272–1282. <https://doi.org/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(5). <https://doi.org/10.1190/1.2783766>
- Yang, X., Mao, N., & Zhu, P. (2022). Bayesian linear seismic inversion integrating uncertainty of noise level estimation and wavelet extraction.

- Li, G., & Ben-Zion, Y. (2023). Daily and seasonal variations of shallow seismic velocities in Southern California from joint analysis of H/V ratios and autocorrelations of seismic waveforms.
- Sun, Q.-H., Zong, Z.-Y., & Li, X. (2024). Probabilistic seismic inversion based on physics-guided deep mixture density network.
  - 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(3), 1523–1553. <https://doi.org/10.1093/gji/ggab287>

***The style is poor at times (too many times to point all difficulties)***

R: Thank you for your comment, we are working on improving the style for the second version of the manuscript.

***The authors are mistakenly using the symmetric version of MH, which invalidates their comparisons.***

R: The symmetric version of the Metropolis-Hastings (MH) algorithm was used to simplify the computation of the acceptance ratio. The proposal distribution  $q(x'|x)$  satisfies the symmetry condition  $q(x'|x)=q(x|x')$ . A common choice for the proposal distribution is a Gaussian  $N(x,\sigma^2)$ , centered at the current point  $x$ , where  $\sigma$  controls the step size. Examples of this approach can be found in the works of Bosch (2007) and others.

***(p2) the HMC was not popularised by Betancourt (2018), there were already books on the topic by that year***

R: Your comment is absolutely correct, as the Hamiltonian Monte Carlo (HMC) method 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 correction has been made in the text.

- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2), 216–222. [https://doi.org/10.1016/0370-2693\(87\)91197-X](https://doi.org/10.1016/0370-2693(87)91197-X)
- MacKay, D. J. C. (2003). *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press.
- 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>

***(p4) one does not "estimate several samples in the parameter space"***

R: Thank you for your comment, the sentence was corrected in the manuscript.

***(p5) the first sentence of 3.1 misses a principal verb***

R: Thank you for your comment, it was reworded as follows: "The Metropolis algorithm is a Markov Chain Monte Carlo (MCMC) method used to generate samples from Markov chains. MH algorithm, introduced by Metropolis et al. (1953) and later generalized by Hastings (1970), defines a transition probability that ensures the Markov chain is ergodic, satisfies detailed balance, and exhibits reversibility. This algorithm generates a sequence of values  $m$  forming a Markov chain, which can be used to approximate a posterior density  $\sigma(m)$ ."

***(p5) one does not have to "assume that we have a chain that converges to the source distribution"***

R: Thank you for your comment, the Metropolis-Hastings (MH) algorithm defines a transition probability that ensures the Markov chain is ergodic, satisfies detailed balance, and exhibits reversibility (Chib & Greenberg, 1995; Robert, 2016).

- Chib, S., & Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4), 327–335. <https://doi.org/10.2307/2684568>
- Robert, C. P. (2016). The Metropolis–Hastings algorithm. arXiv Preprint. <https://arxiv.org/abs/1504.01896v3>

***(p5) the "transition rule of the convergent chain to the source probability density" is not defined and given (5) it should further be symmetric***

R: Thank you for your comment, we have already corrected what you suggested and it is reworded as follows: "Propose a new state: At iteration  $t$  generate a candidate values  $\tilde{m}$  from the proposal probability density  $\tilde{m} \sim q(\tilde{m}/m^{(t-1)})$ , and generate  $u \sim U(0,1)$ , where  $q(\tilde{m}/m^{(t-1)}) = q(m^{(t-1)}/\tilde{m})$ , is symmetric probability distribution."

***(p5) the wording "If [the new]  $\tilde{m}$  is better than the [old]  $m$ " is unclear and unnecessary***

R: Thank you for your comment, we have already corrected what you suggested and it is reworded as follows: ". If  $u \leq p_{\text{accept}}$ , accept  $\tilde{m}$  and set  $m^{(t)} = \tilde{m}$ . Otherwise reject  $\tilde{m}$  and set  $m^{(t)} = m^{(t-1)}$ ."

***(p7) it is not only "in this work, we use the leapfrog method for numerical integration" since this is the default scheme as e.g. in Stan. Furthermore, the leapfrog steps are not provided***

R: Thank you for your comment, it was corrected in the manuscript and we do not use STAN.

***(p7) as described, the HMC algorithm changes the momentum  $m$  at each iteration (in Step 1), which is not the case in general***

R: Thank you for your comment, we agree with your correction and the changes have already been made to the manuscript.

***(p8) the Langevin algorithm with the Metropolis correction is incorrect since the acceptance ratio does not involve  $t$***

R: The Unadjusted Langevin Algorithm (ULA) is simple to implement; however, it introduces bias. To address this, an acceptance-rejection step is introduced via the Metropolis-Hastings (MH) algorithm. By integrating the MH algorithm into ULA, the Metropolis-Adjusted Langevin Algorithm (MALA) is obtained (Izzatullah et al., 2020; Izzatullah et al., 2021).

The procedure involves constructing a Markov chain at each step  $t$ . Given  $m_{t-1}$ , a new observation  $m$  is generated from a proposal density  $q(m)$ . The candidate is accepted with a probability  $p_{\text{accept}}$  defined as:

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

- 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(3), 1523–1553. <https://doi.org/10.1093/gji/ggab287>

***he asymmetric proposals. Langevin diffusion is spelled Langevin diffusion***

R: Thank you for your comment, we agree with your correction and the changes have already been made to the manuscript.

*(p9) ULA was created to avoid rejection and has been deeply investigated in the past years, which makes one wonder at the appeal of MALA-MCMC (missing again the proposals in the acceptance ratio (20)*

R: ULA introduces a bias, need to introduce the acceptance-rejection step, see (Izzatullah et al., 2020, Izzatullah et al., 2021).

*(p11) the initial stage is not "called the burn stage" but the burn-in or warm-up stage*

R: Thank you for your comment, we agree with your correction and the changes have already been made to the manuscript.

*(p17) comparing raw acceptance rates in Table 3 is not appropriate since the different algorithms have different optimal acceptance rates*

R: We follow the methodology proposed by Muhammad Izzatullah, Tristan van Leeuwen, and Daniel Peter (2021), where Tables 2 and 3 in their study compare the acceptance rates of different methods.

- 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(3), 1523–1553. <https://doi.org/10.1093/gji/ggab287>.

## **Response to Anonymous Referee #2**

### ***Major comments:***

**1. Line 15: the full name of ‘MALA’ and ‘Lip-MALA’ should be provided as they firstly appear in the article.**

R: Thank you for your comment, it has already been corrected in the manuscript.

**2. Line 24: ‘...extract additional elastic parameters and reservoir properties...’ What additional elastic parameters and reservoir properties? Please clarify this.**

R: This article is part of a long-term research project and for a next phase we are working on the problem of joint inversion of petrophysical data (clay volume, porosity, water saturation) and seismic data ( $V_p$ ,  $V_s$ , density) these data are related by rock physics models.

**3. Line 25: ‘Integrating well log conditioning into the model could improve vertical resolution near wells and align the model with well data at drilling locations’ Integrating well log data, not only can improve the vertical resolution near wells, but also can improve the vertical resolution far away from wells. Cited here: e.g., Shi et al., 2024.**

R: Thank you for your comment and we believe that the article mentioned contributes to our work and will be incorporated into the references.

**4. Line 44: ‘...the parameters of the medium  $m$  with the observed data...’ Is parameters  $m$  or medium  $m$ ? And please clarify whether  $m$  is a scalar or a vector? There are many similar issues in other parts of the manuscript. Please carefully review the manuscript and unify the expression format.**

R:  $m$  is a vector and the notation in the manuscript was corrected.

**5. Line 54: ‘The approach of this work is based on computer statistics, which allows to include uncertainty in seismic data...’ what is its meaning?**

R: Thank you for your comment, it has already been corrected in the manuscript. The inclusion of randomness in the model implicitly includes uncertainty in the error, and many random samples can be generated that are used to estimate the parameters, from which we can make statistical summaries such as the mean, median and mode in the posterior, variance in the posterior, and construct confidence intervals.

**6. Line 66: ‘Bosch et al., (2007) solve an inverse problem...’ should be modified to ‘Bosch et al., (2007) solved an inverse problem...’ There are many similar issues in the article, please revise.**

R: Thank you for your comment, it has already been corrected in the manuscript.

**7. Line 83: Please provide additional literature related to MALA and Lip-MALA, and briefly describe them.**

R: Thank you for your comment, it has already been corrected in the manuscript. The following references have been included for MALA and Lip-MALA:

- Max Welling and Yee Whye Teh. 2011. Bayesian learning via stochastic gradient langevin dynamics. In Proceedings of the 28th International Conference on International Conference on Machine Learning (ICML'11). Omnipress, Madison, WI, USA, 681–688.
- 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>
- 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>
- Sarkka, Simo & Merktas, Christos & Karvonen, Toni. (2021). Gaussian Approximations of SDES in Metropolis-Adjusted Langevin Algorithms. 1-6. [10.1109/MLSP52302.2021.9596301](https://doi.org/10.1109/MLSP52302.2021.9596301).

**8. Line 104: Why do you emphasize the application in machine learning? Bayesian seismic inversion has always been a major topic.**

R: Thank you for your comment, we agree with what you say, which is why the paragraph was deleted from the manuscript.

**9. Please standardize the formulas in the article. E.g.,  $g:m \rightarrow$  does indicates a forward model?**

R: Thank you for your comment, the notation has been unified and corrected.

**10. Line 184: ‘particle’ is it a model parameter? And what does ‘artificial time variable’ mean?**

R: Thank you for your comment, the wording has been improved, and the analogy of the physical problem to the states of the parameters is made.

**11. Equation (14): Does ‘ $\epsilon$ ’ have the same meaning as ‘ $\epsilon$ ’ in Formula 1? And please explain  $n$ .**

R: No,  $\epsilon$  has different meanings, the notation has already been corrected, and  $n$  is the size of the vector of parameters of the medium ( $\mathbf{m}$ )

**12. Please confirm if formula 18 is written correctly.**

R: Thanks for the comment, the equation was incorrect and was corrected in the manuscript.

**13. Line 263: What do ‘ $\beta_0$ ’ and ‘ $L_c$ ’ represent?**

R:  $\beta_0$  is a ratio of the step length based on the Lipschitz condition,  $L_c$  is the Lipschitz constant

**14. In synthetic data test, the synthetic seismic trace is noise free or noisy?**

R: the synthetic seismic trace is noise free

**15. Please explain any figure to make it easier for readers to understand, e.g., what do Figures 1a-d represent respectively? And add legends.**

R: Thank you for your comment. The figures and their description have been improved in accordance with your suggestion.

**16. The table name should be above the table.**

R: Thanks for the comment, the title of the table was placed in the manuscript where you suggest.

**17. For Figure 1, how does authors distinguish between the sampling phase and the burn-in phase? During the sampling phase, the decrease of the objective function does not tend to a steady state. Will this sampling influence the accuracy of the inversion results? Please explain.**

R: Thanks for the comment, it is standard in Bayesian estimation to use a sample burn-in phase and keep the rest to make the inference and achieve parameter convergence more quickly, using the effective sample size criterion in our case we used multivariate effective sample size (mESS) .

**18. Is Figure 2 the initial samples obtained from a prior distribution? Are the average values of these samples the same as the initial model?**

R: Thanks for the comment, the samples used to make Figure 2 are those accepted according to the criterion of multivariate effective sample size (mESS) . The initial model is a smoothing model based on real data.

**19. In Figure 3, does the gray line represent the synthetic seismic records corresponding to the initial samples (Figure 2)?**

R: Thank you for your comment, the grey lines represent the synthetic seismic traces that were generated using the accepted realizations.

**20. In Figure 4, does the red line represent a random inversion result or the mean of multiple inversion results?**

R: Thank you for your comment, in Figure 4 the red line represents the average of the accepted realizations in the sampling phase.

**21. What does 'corr' mean in Table 2? Does it mean correlation coefficient? If it is the correlation coefficient, it can be seen that the correlation coefficients of the density inversion results obtained by the four algorithms are very low, especially based on the MH algorithm. I think the correlation coefficient based on the MH algorithm is relatively very low, this may unreasonable. Moreover, based on the MH algorithm, the correlation coefficient of the obtained three parameter inversion results may not be so low.**

R: Thanks for your comment, Corr represents the correlation and in our opinion, it is acceptable that the correlation for density is low due to weak sensitivity of seismic reflection amplitude to density as has been demonstrated by many authors.

**22. Please show the seismic gathers before and after well seismic tie?**

R: Thank you for your comment, this image will be added in the second version of the manuscript, but I can tell you that the correlation between the real data and synthetic data is 0.55, which is an acceptable value for seismic well tie.

**23. Line 382: How are 'uncertainty bounds' determined?**

R: Thanks for your comment, 'uncertainty bounds' are determined from the covariance of the data.

**24. *Line 382: The realizations of velocities and density better illustrates their characteristics and variability.***

R: Thank you for your comment, with this comment we wanted to say that the investment results maintain the variability that is observed in the real data.

**25. *Can authors provide a two-dimensional data test?***

R: Yes, it will be added in the second version of the manuscript.

***Some minor corrections:***

R: Thank you very much for the minor corrections, they were corrected in the manuscript and you will be able to see them in the second version of it.

### **Response to Anonymous Referee #3**

***(1) I recommend that the authors briefly explain the reason for selecting the four inversion methods before outlining the research content. This additional context will help readers better understand the study's scope and significance.***

R: Thank you very much for your comment. The selection of these four methods aims to provide an estimation methodology based on Bayesian techniques at different levels of sophistication and efficiency, incorporating well-established theoretical properties and practical considerations to address the seismic data inversion problem.

The Metropolis-Hastings (MH) algorithm serves as a reference method for comparing other approaches. This MCMC method is widely applicable across various scientific fields; however, its main drawback is its slow convergence in high-dimensional or ill-conditioned problems.

The Hamiltonian Monte Carlo (HMC) algorithm enhances MH by leveraging gradient information, enabling efficient exploration of high-dimensional parameter spaces and achieving better convergence rates for complex models.

The Metropolis-Adjusted Langevin Algorithm (MALA) represents an intermediate approach, combining gradient information with stochastic diffusion processes. This balance of computational cost and convergence speed makes MALA well-suited for estimating parameters in moderately complex posterior distributions.

Lip-MALA extends the MALA algorithm by enforcing continuity through the Lipschitz condition on the gradient. This improves the theoretical robustness and performance, particularly in scenarios where the posterior distribution is multimodal.

These four proposed methods enable a comprehensive analysis that combines simplicity, efficiency, and robustness in parameter estimation. These considerations will be included in the second version of the manuscript.

***(2) I suggest the authors add a summary table or figure comparing the performance (inversion accuracy, computational cost, etc.) of the four algorithms across all tested parameters to provide readers with a concise reference.***

R: Thank you very much for your suggestion, we agree and the table will be added in the second version of the manuscript.

*(3) The variable  $\tau$  is used to represent time in subsection 3.2 and step size in subsection 3.3. To avoid confusion, I suggest renaming one of these variables for greater clarity and ease of understanding.*

R: Thank you very much for your correction, it was corrected in the manuscript and you will be able to see it in version 2 of the manuscript.

*(4) The discussion could be strengthened by emphasizing the broader impact of the study. For example, discuss how these methods could be adapted for other types of geophysical data. Besides, add comments on the challenges observed with poor density estimation and propose potential strategies for improvement.*

R: Thank you very much for your comment, it will improve the discussion regarding what you mention, for example, to improve the estimation of density some authors often use relationships such as the Willie equation or the Gardner equation.

*(5) Some specific edits:*

R: Thank you for your comment, the corrections you indicate will be made.