



**HAL**  
open science

# Hamiltonian Monte Carlo based elastic full-waveform inversion of wide-angle seismic data

Nirmit Dhabaria, Satish C Singh

► **To cite this version:**

Nirmit Dhabaria, Satish C Singh. Hamiltonian Monte Carlo based elastic full-waveform inversion of wide-angle seismic data. *Geophysical Journal International*, 2024, 237 (3), pp.1384-1399. 10.1093/gji/ggae112 . hal-04581681

**HAL Id: hal-04581681**

**<https://u-paris.hal.science/hal-04581681>**

Submitted on 21 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# Hamiltonian Monte Carlo based elastic full-waveform inversion of wide-angle seismic data

Nirmit Dhabaria<sup>1</sup> and Satish C. Singh

*Institut de Physique du Globe de Paris, Université Paris Cité, 1 rue Jussieu, 75238 Paris 05, France. E-mail: [nirmit.dhabaria@gmail.com](mailto:nirmit.dhabaria@gmail.com)*

Accepted 2024 March 12. Received 2024 March 1; in original form 2022 September 19

## SUMMARY

Full-waveform inversion (FWI) of seismic data provides quantitative constraints on subsurface structures. Despite its widespread success, FWI of data around the critical angle is challenging because of the abrupt change in amplitude and phase at the critical angle and the complex waveforms, especially in the presence of a sharp velocity contrast, such as at the Moho transition zone (MTZ). Furthermore, the interference of refracted lower crustal (Pg) and upper mantle (Pn) arrivals with the critically reflected Moho (PmP) arrivals in crustal and mantle studies makes the application of conventional FWI based on linearized model updates difficult. To address such a complex relationship between the model and data, one should use an inversion method based on a Bayesian formulation. Here, we propose to use a Hamiltonian Monte Carlo (HMC) method for FWI of wide-angle seismic data. HMC is a non-linear inversion technique where model updates follow the Hamiltonian mechanics while using the gradient information present in the probability distribution, making it similar to iterative gradient techniques like FWI. It also involves procedures for generating distant models for sampling the posterior distribution, making it a Bayesian method. We test the performance and applicability of HMC based elastic FWI by inverting the non-linear part of the synthetic seismic data from a three-layer and a complex velocity model, followed by the inversion of wide-angle seismic data recorded by two ocean bottom seismometers over a 70 Ma old oceanic crustal segment in the equatorial Atlantic Ocean. The inversion results from both synthetic and real data suggest that HMC based FWI is an appropriate method for inverting the non-linear part of seismic data for crustal studies.

**Key words:** Inverse theory; Probability distributions; Body waves; Crustal imaging; Waveform inversion.

## 1 INTRODUCTION

Seismic tomography methods based on matching of traveltimes of different seismic arrivals are commonly used to determine large-scale velocity structures on both regional (Zelt & Smith 1992; Van der Voo *et al.* 1999; Korenaga *et al.* 2000; Hobro *et al.* 2005) and global (Dziewonski *et al.* 1977; Dziewonski 1984; Zhou 1996; Van der Hilst *et al.* 1997; Bijwaard *et al.* 1998) scales. The introduction of full-waveform inversion (FWI) in the 1980s (Bamberger *et al.* 1982; Lailly & Bednar 1983; Tarantola 1984a, b) made it theoretically possible to recover finer scale structures of the subsurface (Shaw & Orcutt 1985; Chapman & Orcutt 1985; Operto *et al.* 2004). Although this advancement came at the expense of increased computational requirements, the growth in computational capabilities in the last two decades has allowed rapid development in the applications of FWI (Ravaut *et al.* 2004; Capdeville *et al.* 2005; Operto *et al.* 2006; Brossier *et al.* 2009; Fichtner *et al.* 2009; Virieux &

Operto 2009; Morgan *et al.* 2013, 2016; Warner *et al.* 2013; French & Romanowicz 2014; Borisov & Singh 2015).

Full-waveform inversion (FWI) is a local optimization technique for estimating high-resolution physical parameters of the subsurface based on matching waveforms of the synthetic and observed data. The non-linear relationship of the misfit function with the model parameters makes FWI dependent on an appropriate initial model close enough to the final model to avoid cycle skipping of modelled waveforms (Bunks *et al.* 1995) and getting entrapped in local minima. The iterative inversion scheme of FWI implements gradient optimization methods to compute model updates, ideally defined as a product of the inverse Hessian (the second derivative of the misfit function with respect to the model parameters) and the gradient of the misfit function (Mora 1987; Pratt *et al.* 1998; Virieux & Operto 2009). In this context, Newton and Gauss–Newton methods (Pratt *et al.* 1998; Sheen *et al.* 2006; Askan *et al.* 2007) of model updates are often considered impractical for realistic

problems due to the high cost of computing the Hessian. On the other hand, inversion schemes based on either iteratively updating the Hessian using quasi-Newton methods such as the limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS) method (Nocedal 1980) or determining the gradient direction by replacing the Hessian with a scalar parameter called step-length (Pica *et al.* 1990; Virieux & Operto 2009), have shown to provide reliable and stable results while also being computationally less demanding.

Over the years, several approaches have been developed to mitigate the non-linearity of the misfit function. Several of these approaches included modification of the misfit functions (Fichtner *et al.* 2008; Brossier *et al.* 2010; Bozdağ *et al.* 2011; Choi & Alkhalifah 2012; Métivier *et al.* 2016) implemented in both the time and frequency domains, with an increasing time window and frequency bandwidth of the inverted seismic data (Ravaut *et al.* 2004; Sears *et al.* 2008; Brossier *et al.* 2009). These approaches also introduced the concept of hybrid cross-correlation between the synthetic and observed data (Luo & Schuster 1991; Gee & Jordan 1992), source-encoded inversions (Krebs *et al.* 2009; Ben-Hadj-Ali *et al.* 2011; Choi & Alkhalifah 2011) and hierarchical layer-stripping strategies (Shipp & Singh 2002; Sirgue & Pratt 2004) for convergence to an optimum model.

Another problem in FWI is associated with the non-linear behaviour of the amplitude and phase of seismic reflections around the critical and post-critical incident angles. This non-linear characteristic of the data arises due to the transformation of the reflection coefficients into complex numbers starting from the critical angle (Aki & Richards 2002; Stein & Wysession 2009) in the presence of a sharp high-velocity contrast in the model, such as at the Moho transition zone (MTZ). As a result, waveforms at the critical angle become the Hilbert transform of the incident wave (Choy & Richards 1975). The additional interference of refracted lower crustal (Pg) and upper mantle (Pn) arrivals with the Moho reflections (PmP), and their triplications results in very complex waveforms and could lead to an inaccurate estimation of elastic parameters while using FWI based on the linearized model updates. As the inverse problem of finding the model parameters is non-linear, its solution can be better represented with a probabilistic Bayesian formulation (Keilis-Borok & Yanovskaja 1967; Press 1968; Backus & Gilbert 1970; Tarantola & Valette 1982; Mosegaard & Tarantola 1995; Tarantola 2005).

Bayesian inversion based on Markov Chain Monte Carlo (MCMC; Metropolis *et al.* 1953) methods can estimate the model and its uncertainties by constructing the posterior probability distribution (PPD) of all the sampled models. The PPD associates uncertainties of all the sampled models in the framework of existing prior knowledge, allowing us to quantify and determine statistical properties associated with all the model parameters. Bayes' theorem provides a convenient basis for evaluating this PPD by incorporating the prior probability distribution with a likelihood function of the data. Over the years, several adaptations of MCMC methods have been used for uncertainty estimation of the model parameters (Cary & Chapman 1988; Sen & Stoffa 1991; Malinverno 2002; Sambridge *et al.* 2006; Hong & Sen 2009). The successful application of these methods, along with rapid growth in the computational facilities, has led to the development of advanced MCMC strategies, such as using temperature ladders, as in parallel tempering (Marinari & Parisi 1992; Sambridge 2014) and trans-dimensional methods (Green 1995; Malinverno 2002; Bodin & Sambridge 2009; Piana Agostinetti *et al.* 2015; Bottero *et al.* 2016; Guo *et al.* 2021). Nevertheless, the computational cost and necessity of sampling a large number of models for higher-dimensional problems remain critical factors in inverting large seismic data. This phenomenon is

known as the curse of dimensionality. It explains the diminishing performance of an MCMC method with increasing dimensions of the model parameters, making an unbiased Markov chain too long for a suitable convergence (Backus 1988; Scales & Tenorio 2001).

The above limitations of both inversion methods encourage us to invert seismic data using a more appropriate non-linear inversion technique. Based on sampling a particle in a multidimensional physical system using Hamiltonian mechanics, Hamiltonian Monte Carlo (HMC) is an MCMC method that uses the gradient information in the probability distributions to generate less correlated samples in the model space. As a result, it has a high sample acceptance rate in comparison to the random walk MCMC method. Introduced initially for its application in lattice quantum chromodynamics by Duane *et al.* (1987), HMC has found its use in many fields, including machine learning and neural networks (Yang *et al.* 2021), molecular simulations (Widom *et al.* 2014), quantum mechanics (Seah *et al.* 2015) and many more. It has also found its use in the geophysical fields, including inversion of electrical resistivity data (Maiti *et al.* 2011), amplitude versus angle analyses (Sen & Biswas 2017), seismic source inversions (Fichtner & Simutè 2018), seismic tomography and FWI (Fichtner *et al.* 2019; Gebraad *et al.* 2020; Fichtner *et al.* 2021). However, its application in solving non-linear problems associated with the complexities of inverting wide-angle seismic data remains untested.

Here, we present an application of HMC based elastic FWI of wide-angle seismic data. We test the algorithm by inverting seismic data from two groups of laterally homogeneous velocity profiles, followed by the inversion of OBS data from the Atlantic Ocean.

## 2 THEORY

### 2.1 Bayesian inference

Bayes' theorem relates prior information available on a model with data to construct the PPD and can be expressed as

$$P(\mathbf{m} | \mathbf{d}_{\text{obs}}) = \frac{P(\mathbf{m}) P(\mathbf{d}_{\text{obs}} | \mathbf{m})}{P(\mathbf{d}_{\text{obs}})}, \quad (1)$$

where  $\mathbf{m}$  and  $\mathbf{d}_{\text{obs}}$  represent the model and the observed data, with  $P(A)$  and  $P(A|B)$  representing the probability distribution and the conditional probability of A with respect to B, respectively.  $P(\mathbf{m} | \mathbf{d}_{\text{obs}})$  is the PPD,  $P(\mathbf{m})$  is the prior probability of all constraints on the model parameters prior to any calculation and  $P(\mathbf{d}_{\text{obs}} | \mathbf{m})$  is the likelihood function which represents the probability of the observed data  $\mathbf{d}_{\text{obs}}$  given the model  $\mathbf{m}$ . The denominator term  $P(\mathbf{d}_{\text{obs}})$  is known as the evidence or the marginal likelihood, representing an integral of prior with the likelihood function over an entire model space such that  $P(\mathbf{d}) = \int P(\mathbf{m}) P(\mathbf{d}_{\text{obs}} | \mathbf{m}) d\mathbf{m}$ . As the evidence  $P(\mathbf{d}_{\text{obs}})$  remains a constant normalization factor for the PPD, eq. (1) can be expressed as,

$$P(\mathbf{m} | \mathbf{d}_{\text{obs}}) \propto P(\mathbf{m}) P(\mathbf{d}_{\text{obs}} | \mathbf{m}). \quad (2)$$

### 2.2 Hamiltonian Monte Carlo

In the HMC method, the inverse problem of quantifying an  $n$ -dimensional model ( $\mathbf{m}$ ) is transformed into the sampling of particles in a multidimensional physical system governed by the Hamiltonian mechanics (Neal 1993, 1996) and defined in terms of the total energy referred to as the Hamiltonian  $H(\mathbf{m}, \mathbf{p})$ , consisting of two terms; potential energy  $U(\mathbf{m})$  and kinetic energy  $K(\mathbf{p})$ ,

$$H(\mathbf{m}, \mathbf{p}) = U(\mathbf{m}) + K(\mathbf{p}), \quad (3)$$

where the  $n$ -dimensional vector  $\mathbf{p}$  is the momentum of the particle. The  $n$ -dimensional vector  $\mathbf{m}$  is the model sampled from the PPD and represents the position vector of the particle. The potential energy  $U(\mathbf{m})$  defines the energy associated with the position of this particle and can be expressed as,

$$U(\mathbf{m}) = -\ln P(\mathbf{m} | \mathbf{d}_{\text{obs}}). \quad (4)$$

The kinetic energy  $K(\mathbf{p})$  of a particle due to its motion is expressed in terms of an auxiliary  $n$ -dimensional momentum variable  $\mathbf{p}$  such that

$$K(\mathbf{p}) = \frac{1}{2} \mathbf{p}^T \mathbf{M}^{-1} \mathbf{p}, \quad (5)$$

where  $\mathbf{M}$  is a symmetric and positive definite mass matrix of dimension  $n \times n$  and is analogous to a physical mass linking each variable pair of the position ( $\mathbf{m}$ ) and momentum ( $\mathbf{p}$ ) vector. Together, the position and momentum vectors ( $\mathbf{m}$ ,  $\mathbf{p}$ ) define the state of a particle inside the physical system of an  $n$ -dimensional model and momentum space, also called the phase space.

Hamilton's equations (eq. 6) govern the dynamics of the particle such that,

$$\frac{d\mathbf{m}}{d\tau} = \frac{\partial H}{\partial \mathbf{p}} = \mathbf{M}^{-1} \mathbf{p}, \quad \frac{d\mathbf{p}}{d\tau} = -\frac{\partial H}{\partial \mathbf{m}} = -\frac{\partial U}{\partial \mathbf{m}}, \quad (6)$$

where  $\tau$  is a time variable. Hamilton's equations are conservative, and thus moving a particle from an initial state ( $\mathbf{m}$ ,  $\mathbf{p}$ ) over time  $\tau$  results in the generation of a new state ( $\mathbf{m}_\tau$ ,  $\mathbf{p}_\tau$ ) corresponding to new values of potential and kinetic energies and an unchanged Hamiltonian. The initialization of HMC at every sample requires the generation of a random momentum vector from a multivariate Gaussian distribution with the covariance matrix  $\mathbf{M}$ , thus defining the initial state of the particle ( $\mathbf{m}$ ,  $\mathbf{p}$ ). HMC uses Hamiltonian mechanics by propagating the particle towards the new states and incorporates Bayesian inference by defining the potential and kinetic energies in terms of the probability distribution such that every state of the particle is sampled from a joint probability distribution  $P(\mathbf{m}, \mathbf{p} | \mathbf{d}_{\text{obs}})$  of position and momentum vector in the phase space. This joint probability distribution is also known as the canonical distribution and can be expressed in terms of Hamiltonian as  $P(\mathbf{m}, \mathbf{p} | \mathbf{d}_{\text{obs}}) = \exp(-H(\mathbf{m}, \mathbf{p}))$ . Following the construction of the joint probability distribution from the sampled states, the PPD of the sampled models can be constructed by marginalizing the momentum component of the joint probability distribution (Betancourt 2017).

Theoretically, the analytical solution of Hamilton's equations would have all the proposed states accepted in the joint probability distribution  $P(\mathbf{m}, \mathbf{p} | \mathbf{d}_{\text{obs}})$ . However, waveform inversion of seismic data using HMC requires discretization of Hamilton's equations using numerical integration schemes, which can preserve the properties of the Hamiltonian system, including the time-reversibility, preservation of volume and phase space partitioning (Neal *et al.* 2011). The integration schemes satisfying these properties are known as symplectic integrators (Simo *et al.* 1992). We have used a symplectic leapfrog integrator in this study for the generation of new particle states with a step size  $\epsilon$

(Betancourt 2017) as,

$$\begin{aligned} \mathbf{p}_i(\tau + \frac{\epsilon}{2}) &= \mathbf{p}_i(\tau) - \frac{\epsilon}{2} \frac{\partial U}{\partial \mathbf{m}_i} \Big|_{\tau}, \\ \mathbf{m}_i(\tau + \epsilon) &= \mathbf{m}_i(\tau) + \epsilon \frac{\partial K}{\partial \mathbf{p}_i} \Big|_{\tau+\epsilon/2}, \\ \mathbf{p}_i(\tau + \epsilon) &= \mathbf{p}_i(\tau + \frac{\epsilon}{2}) - \frac{\epsilon}{2} \frac{\partial U}{\partial \mathbf{m}_i} \Big|_{\tau+\epsilon}, \end{aligned} \quad (7)$$

where  $i$  is the vector index in the range  $(1, n)$ . A single sample workflow of HMC for an  $n$ -dimensional initial model  $\mathbf{m}$  can be described in a series of steps where first, a random  $n$ -dimensional momentum vector  $\mathbf{p}$  is drawn from a multivariate Gaussian distribution with zero mean and covariance  $\mathbf{M}$  defining the particle initial state ( $\mathbf{m}$ ,  $\mathbf{p}$ ) for the initial Hamiltonian  $H(\mathbf{m}, \mathbf{p})$ . The discretized Hamilton's equation integrates the initial state ( $\mathbf{m}$ ,  $\mathbf{p}$ ) over  $L$  leapfrog steps with step size  $\epsilon$  to generate a new state ( $\mathbf{m}_\tau$ ,  $\mathbf{p}_\tau$ ) and computes a new Hamiltonian  $H(\mathbf{m}_\tau, \mathbf{p}_\tau)$ . Though highly accurate in their properties, implementing discretized symplectic integration schemes introduces errors in the Hamiltonian while moving from one state to another. To correct errors introduced by the discretization scheme at every iteration, the new state is accepted with a probability,

$$P_{\text{accept}} = \min[1, \exp(H(\mathbf{m}, \mathbf{p}) - H(\mathbf{m}_\tau, \mathbf{p}_\tau))]. \quad (8)$$

Upon the acceptance of the new state, the new  $\mathbf{m}_\tau$  replaces the initial  $\mathbf{m}$  position vector, which otherwise remains the same and follows the same iterative workflow with a newly drawn random momentum vector.

### 2.3 HMC based FWI

In order to perform full-waveform inversion of seismic data using HMC, we associate the data misfit function with the likelihood function and in turn, the potential energy. We define the likelihood function as,

$$P(\mathbf{d}_{\text{obs}} | \mathbf{m}) = \frac{1}{\sqrt{(2\pi)^{N_r \times N_r} |C_d|}} \exp\left(-\frac{J_d(\mathbf{m})}{2}\right), \quad (9)$$

where  $N_r$  and  $N_t$  are the total number of receivers and trace length,  $C_d$  is the data covariance matrix and  $|\cdot|$  represents the determinant of the matrix.  $J_d(\mathbf{m}) = \sum_{i=1}^{N_r} \sum_{j=1}^{N_r} \Delta d_{i,j}(\mathbf{m})^T C_d^{-1} \Delta d_{i,j}(\mathbf{m})$  is the data misfit function (Mahalanobis 1936) with the data residual  $\Delta d(\mathbf{m})$  defined as a difference between synthetic data  $\mathbf{d}_{\text{syn}}$  and observed data  $\mathbf{d}_{\text{obs}}$ , such that  $\Delta d(\mathbf{m}) = \mathbf{d}_{\text{syn}} - \mathbf{d}_{\text{obs}}$ . With this definition, we can link the gradient of the potential energy computed in the HMC (eq. 7) to the gradient of the data misfit function computed in FWI. Similarly, we define the prior probability as,

$$P(\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^{N_z} |C_m|}} \exp\left(-\frac{J_m(\mathbf{m})}{2}\right), \quad (10)$$

where  $N_z$  is the number of model parameters and  $C_m$  is the model covariance matrix of the form  $C_m = \sigma_m^2 I$ , with  $\sigma_m$  representing the model standard deviation.  $J_m(\mathbf{m}) = (\mathbf{m} - \mathbf{m}_0)^T C_m^{-1} (\mathbf{m} - \mathbf{m}_0) / N_z$  is the model misfit function between the current  $\mathbf{m}$  and initial  $\mathbf{m}_0$  model.

### 2.4 Numerical implementation

The numerical strategy of HMC requires the computation of probability distributions and can be reduced to solving the forward problem for the computation of the synthetic data and gradients of the

misfit function, similar to the conventional FWI. The numerical modelling in this study uses a time-domain solution of the 2-D elastic wave equation discretized in a fourth-order adaptation of the staggered-grid finite difference approximation of the velocity and stress fields (Virieux 1986; Levander 1988) to perform laterally homogeneous (1-D) updates of the velocity model. We actively update the  $P$ -wave velocity ( $V_p$ ) and update the  $S$ -wave velocity ( $V_s$ ) and density using empirical relationships with  $V_p$  (Hamilton 1978; Castagna *et al.* 1985; Fliedner *et al.* 1998).

From eq. (4), we can compute the gradient of potential energy as half of the summation of gradients of the model and data misfit function. Computation of gradients of the data misfit function is performed using the adjoint-state method by combining the forward and adjoint wavefield using time-reversed signals of the adjoint source (Liu & Tromp 2006; Plessix 2006; Fichtner *et al.* 2006).

## 2.5 Tuning parameters and tempered HMC variant

The performance of HMC depends upon a careful selection of tuning parameters for convergence and quantification of the posterior distribution. These tuning parameters include leapfrog steps, step size and an appropriate estimate of the mass matrix. Numerical integration with a considerable step size impedes the conservative property of the Hamilton's equations, resulting in lower acceptance rates (Neal *et al.* 2011; Betancourt 2017). Therefore, a suitable value of step size  $\epsilon$  is empirically chosen to reduce numerical inaccuracies and drift arising due to the discretization of the Hamilton's equations.

The mass matrix is another tuning parameter which plays a crucial role in the convergence of the HMC algorithm and embodies propagation properties of the momentum vector in sampling new models from the model space. Hamilton's equations can be transformed into a second-order ordinary differential equation (ODE) of an undamped harmonic oscillation of the position vector under an external force. Solving this ODE for uniform oscillations of the position (model) vector components to efficiently sample the model space gives an estimate of the mass matrix as an inverse posterior covariance (Fichtner *et al.* 2019) such that,

$$\mathbf{M} = \left[ C_m^{-1} + \frac{\partial^2 J_d(\mathbf{m})}{\partial \mathbf{m}^2} \right], \quad (11)$$

where the term  $\partial^2 J_d(\mathbf{m})/\partial \mathbf{m}^2$  represents the Hessian computed at a given model  $\mathbf{m}$ . Due to the computational limitations in generating and storing Hessian at every specific intervals, a first-order estimation of Hessian is generated through the gradients of the misfit function for the computation of the mass matrix.

The tempering of a probability distribution is a process of rescaling an optimization function (Sambridge 2014). It can drastically increase the performance of an MCMC method by better sampling the flattened PPD. Gaussian-based distributions are very prone to tempering as their values are directly proportional to the size of variables. As such, this rescaling can become crucial, especially in real data applications where the original Gaussian-derived probability distributions can become highly variable, even for a minor change in the model parameters. A tempered probability distribution over an original posterior distribution can be written as

$$P_T(\mathbf{m} | \mathbf{d}_{\text{obs}}) = P(\mathbf{m} | \mathbf{d}_{\text{obs}})^{1/T} \propto P(\mathbf{m})^{1/T} P(\mathbf{d}_{\text{obs}} | \mathbf{m})^{1/T}, \quad (12)$$

where  $T$  is the temperature of the tempered distribution and can be applied separately to individual probability distributions (Sambridge 2014). The initialization of HMC in this study is based on a

methodology adapted from the works of Fichtner & Zunino (2019) and Fichtner *et al.* (2021) in sampling states of an effective nullspace around the converging solution of the tempered likelihood distribution by defining a certain tolerance bound within a range of the decreasing Hamiltonian.

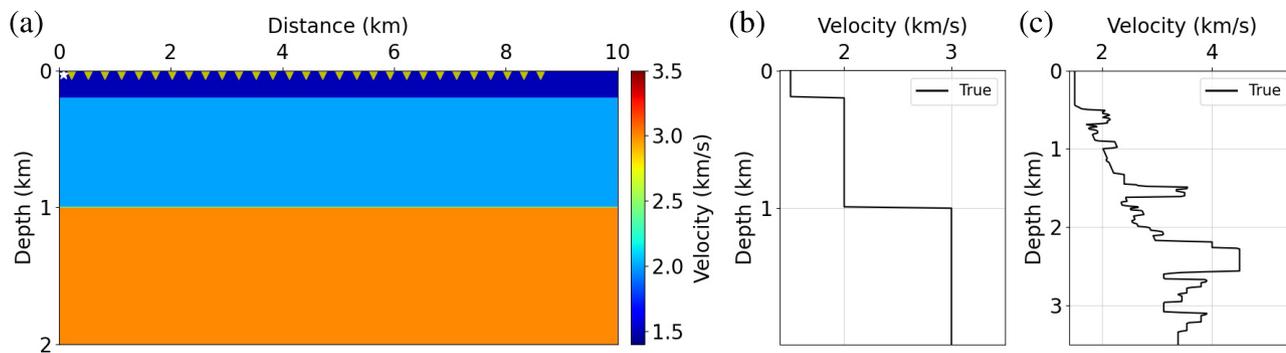
The expanded discretization of the Hamilton's equations can provide details of the sensitivity of a randomly drawn momentum vector and data misfit gradient on the model updates, where the sensitivity of the random momentum vector and gradient are proportional to  $\epsilon \mathbf{M}^{-1}$  and  $\epsilon^2 \mathbf{M}^{-1}/2$  (Fichtner & Zunino 2019). While the large number of leapfrog steps in HMC can be optimized to generate less correlated samples, the computational requirements limit its value in this study to only  $L = 1$  or  $2$ . Here, we have used  $L = 1$  on a test basis for both the synthetic and real data applications. Based on this value, the step size is empirically chosen in the range  $\epsilon \in (0.01, 0.1)$  with the maximum value in the mass matrix equal to 1. These parameters ensure an extensive sampling of the model space over the total number of samples varying between 3500 and 5000 based on the allocated cluster resources. Tuning of the step size is implemented through the variation of acceptance rate and is defined as a product of  $\alpha \epsilon_{\text{original}}$ , where  $\alpha$  is a ratio of acceptance rate averaged over the empirically chosen 25–50 samples with an optimal acceptance rate kept at 65 per cent.

## 3 SYNTHETIC TESTS

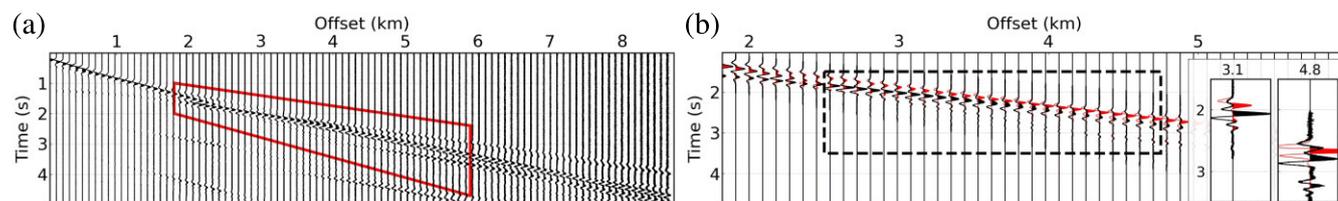
For the sake of simplicity, we will abbreviate HMC based elastic FWI with just HMC in the following sections. These synthetic studies aim to test the performance and applicability of the HMC against the conventional FWI in addressing non-linearities in wide-angle seismic data due to sharp velocity contrasts in the velocity models. For this objective, we conduct synthetic tests with two 1-D isotropic shallow velocity models representing acquisition geometries of small-scale wide-angle seismic data. We use a Ricker source wavelet with a dominant frequency of 6.5 Hz to compute the synthetic data. The marine acquisition geometry for both tests uses a single source at a depth of 20 m with a total of 290 receivers spaced equidistantly at an interval of 30 m at 30 m depth in an end-on spread with a minimum source–receiver distance of 50 m (Fig. 1). The real seismic data are emulated by adding a Gaussian random noise with standard deviation,  $\sigma_{\text{noise}}$  equal to 25 per cent of the mean amplitude. We have used the data covariance as an identity matrix of the form  $C_d = \sigma_d^2 I$ , with a constant noise variance  $\sigma_d^2 = 1$ . Given the cluster resources, the inversion is performed on  $\sim 20$  independent Markov chains for around 5000 samples with different values of mass matrix and step size to obtain the optimum model.

### 3.1 Test-1: 1-D three layered velocity model

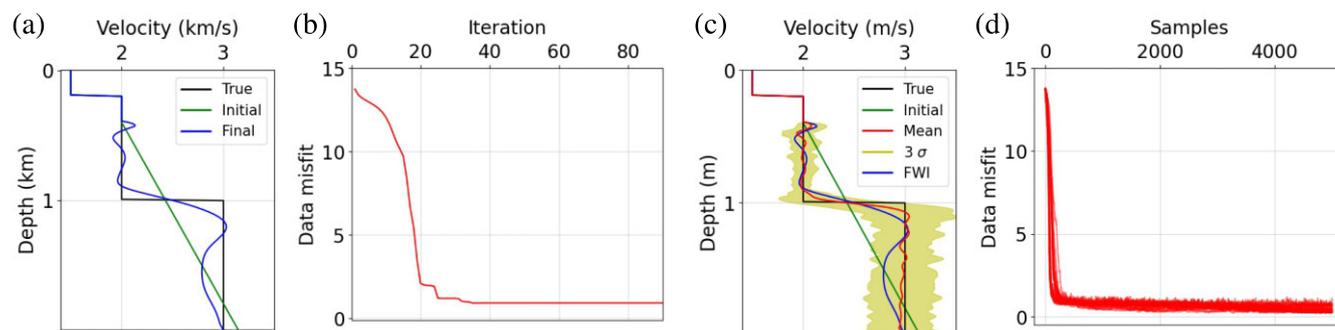
The first test is performed on a three-layered 1-D velocity model (Fig. 1b) to assess the applicability of HMC in inverting seismic data around the critical angle (Fig. 2a). The true model is 10.0 km long and 2.0 km deep, with the seabed at 0.2 km depth. It consists of two sharp interfaces at 0.2 km and 1.0 km depths with layer velocities of 2.0 km s<sup>-1</sup> and 3.0 km s<sup>-1</sup>. The initial model (Fig. 3a, green) follows the true model up to 0.4 km depth. Below this depth, it increases linearly from a velocity of 2.0 km s<sup>-1</sup> at 0.4 km depth to 3.2 km s<sup>-1</sup> at 2.0 km depth with a constant velocity gradient. The model is discretized at 0.01 km  $\times$  0.01 km grid interval leading to a total model dimension of 1000  $\times$  200 gridpoints. The offsets between 1.8 km and 6.0 km are used for the inversion, with a



**Figure 1.** (a) 2-D laterally homogeneous model (Test-1), and the 1-D representations of (b) Test-1 and (c) Test-2. The star (white) represents a single source at a depth of 20 m, and triangles (yellow) represent receivers placed at an equidistant interval of 30 m at 30 m depth.



**Figure 2.** (a) Resulting seismic data (normalized) from the 1-D velocity model (Fig. 1b) with the selected data enclosed within the red box. (b) Comparison of the observed (black) and initial (red) seismic data, with the black box enclosing the data with the major residual difference (observed - synthetic). The inset plot shows the blow-up of the observed (black) and initial (red) seismic data at 3.1 km and 4.8 km offsets.



**Figure 3.** Result of a three-layered velocity model (black) with an initial velocity model (green) for inverting seismic data with (a) FWI (blue) and (c) HMC mean (red) with the yellow fill representing an interval of three standard deviations. Panels (b) and (d) shows the respective variation of data misfit function of FWI and HMC chains with increasing iterations and samples.

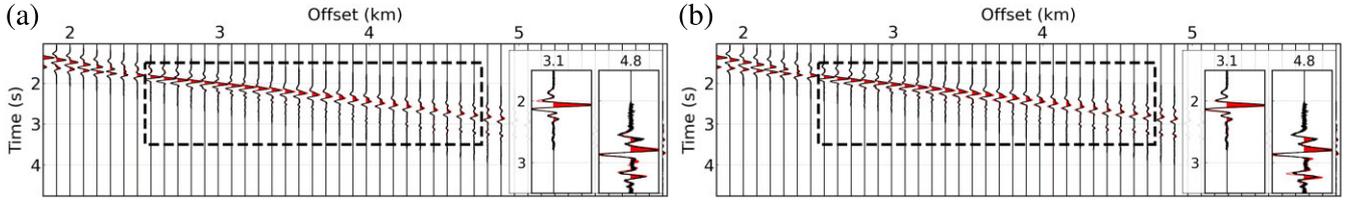
critical angle offset around 2.5 km. Fig. 2(b) shows the comparison of the observed (black) and initial synthetic (red) data from the true and initial velocity models. The initial synthetic data show an absence of refractions from the third layer due to a constant velocity gradient. Assuming the velocity model up to the 0.4 km depth is known, we perform 1-D model updates from 0.4 km to 2.0 km depth containing 160  $V_p$  model parameters for both FWI and HMC. The prior distribution is a Gaussian distribution (eq. 10) with a broad model standard deviation  $\sigma_m$  of  $1.0 \text{ km s}^{-1}$  for efficient sampling of the model space performed through multiple independent chains.

Fig. 3(a) shows the result of FWI (blue), and Fig. 3(c) shows the converged solution of HMC with the mean (red) and an interval of three standard deviations (yellow). The variation of data misfit as a function of iterations and samples for FWI and HMC are shown in Figs 3(b) and (d), respectively. From these figures, we observe that the final data misfits for both inversions are similar, but there is a difference in the inverted velocity models below 1.2 km depth, which indicates the presence of local minima. FWI converges to a local minimum, as it cannot completely recover the velocity in

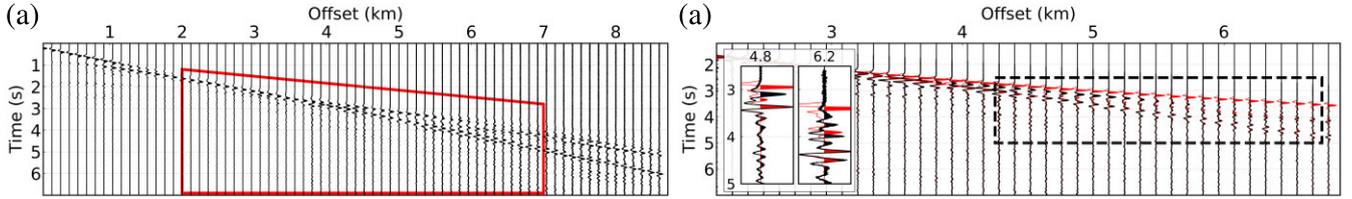
the third layer. The application of HMC provides better results than FWI by recovering both the long and short-wavelength structures of the velocity model, with partial recovery of the third layer due to minimal gradient information present in the observed seismic data. Comparison of the observed (black) and synthetic (red) data for both FWI and HMC are shown in Figs 4(a) and (b), with the enclosed black box showing the large data residuals for the result of FWI compared to HMC.

### 3.2 Test-2: 1-D complex velocity model

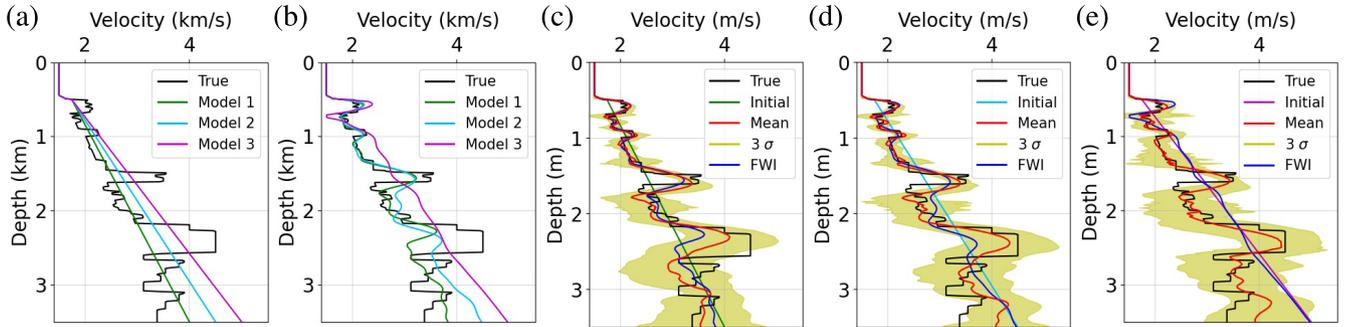
We performed the second test on a 1-D complex velocity model (Fig. 1c) to determine the applicability of HMC in inverting the seismic data (Fig. 5a) with three different initial velocity models. The true model is 10.0 km long and 3.5 km deep, with the seabed at 0.45 km depth, and contains three major sharp interfaces around 0.52 km, 1.5 km and 2.3 km depths. The initial models (green, blue, magenta; Fig. 6a) follow the true model up to the depth of 0.52 km. Below this depth, the velocities increase linearly to  $4.0 \text{ km s}^{-1}$



**Figure 4.** Comparison of the observed (black) and synthetic (red) seismic data from the inversion results (Fig. 3) of (a) FWI and (b) HMC. The enclosed black box shows the seismic arrivals from the third velocity layer with the major residual difference present in FWI compared to HMC. The inset plot in (a) and (b) shows the blow-up of the observed (black) and synthetic (red) seismic data from FWI and HMC at 3.1 km and 4.8 km offsets.



**Figure 5.** (a) Resulting seismic data (normalized) from the 1-D velocity model (Fig. 1c) with the selected data enclosed within the red box. (b) Comparison of the observed (black) and initial (red) seismic data from the true and initial velocity Model 3, with the black box enclosing the data with the major residual difference. The inset plot shows the blow-up of the observed (black) and initial (red) seismic data at 4.8 km and 6.2 km offsets.



**Figure 6.** (a) Initial velocity models used for the inversion of the seismic data from the complex layered velocity model (black). Different starting models are highlighted with different colours. (b) Results of FWI on the seismic data from the three initial velocity models shown in (a). Mean (red) of HMC after inverting the seismic data from the initial velocity Model 1 (c), Model 2 (d) and Model 3 (e) with the yellow fill representing intervals of three standard deviations. The results of FWI are shown in blue.

(Model 1),  $4.5 \text{ km s}^{-1}$  (Model 2) and  $5.0 \text{ km s}^{-1}$  (Model 3) at 3.5 km depth. The offsets used for the inversion of the seismic data are from 2.0 km to 7.0 km, with the critical angle offsets around 2.3 km and 3.7 km from the second and third layer high-velocity contrasts. Fig. 5(b) compares the observed (black) and initial synthetic (red) data from the true and initial velocity Model 3. We use the same grid interval as the previous test, leading to a total model dimension of  $1000 \times 351$  gridpoints. The 1-D model updates are performed from 0.45 km to 3.5 km depth containing 306  $V_p$  model parameters for both the FWI and HMC, with a model standard deviation of  $1.0 \text{ km s}^{-1}$  for sampling the model space through  $\sim 20$  independent chains.

Fig. 6(b) shows the result of FWI for Model 1 (green), Model 2 (blue) and Model 3 (magenta) and Fig. 7(a) shows the variation of their data misfit as a function of iterations. Figs 6(c), (d) and (e) show the solutions of HMC with a mean (red) and three standard deviations (yellow) for Model 1, 2 and 3, respectively. The variation of their data misfit as a function of increasing samples is shown in Figs 7(b), (c) and (d). The results of FWI of the seismic data from Model 1 and Model 2 show only a partial recovery of the true velocity model (Fig. 6b). In contrast, the FWI of the seismic data from Model 3 does not converge due to a significant difference between

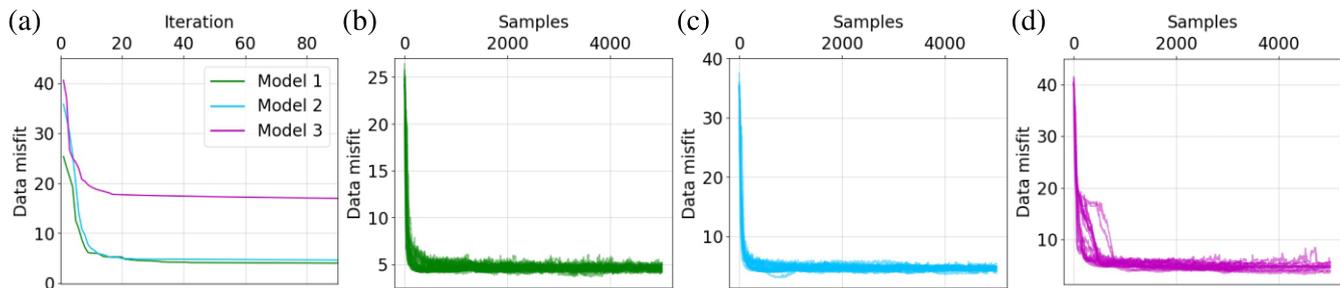
the initial and true velocity model. The HMC, on the other hand, provides a much better solution than the FWI for all three velocity models and recovers both short and long-wavelength structures of the true velocity model, especially in the case of Model 3, where the FWI gives a very poor result. The comparison of observed (black) and synthetic (red) data from the solution of the FWI (Fig. 8a) and HMC (Fig. 8b) of Model 3 confirms this observation. These results indicate a degree of independence of the HMC on the initial velocity model for the convergence to an optimum model.

#### 4 APPLICATION OF THE HMC TO WIDE-ANGLE OBS DATA

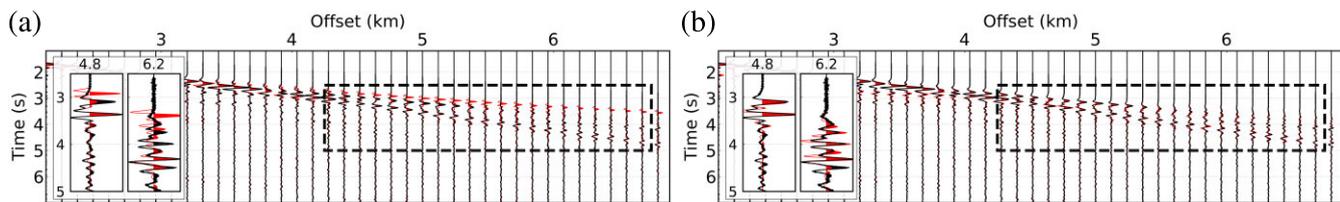
The results of synthetic tests show the potential of the HMC in inverting seismic data sets with critically reflected arrivals. Here we present the application of the HMC to ocean bottom seismometer (OBS) data from the Atlantic Ocean.

##### 4.1 Study area and OBS data set

We use the OBS data that was acquired during the 2018 ILAB-SPARC seismic survey (Marjanović *et al.* 2020; Grove *et al.* 2021) in the Atlantic Ocean over the crust formed at the



**Figure 7.** (a) Variation of the data misfit functions with increasing iterations of FWI on the seismic data from the initial velocity models (Fig. 6a). Variation of data misfit function with increasing samples of HMC (red) on the seismic data from the velocity (b) Model 1, (c) Model 2 and (d) Model 3.



**Figure 8.** Comparison of observed (black) and synthetic (red) seismic data from the inversion results of (a) FWI and (b) HMC on Model 3 shown in Fig. 6. The black box encloses the data with the major residual difference present in FWI compared to HMC. The inset plot in (a) and (b) shows the blow-up of the observed (black) and synthetic (red) seismic data from FWI and HMC at 4.8 km and 6.2 km offsets.

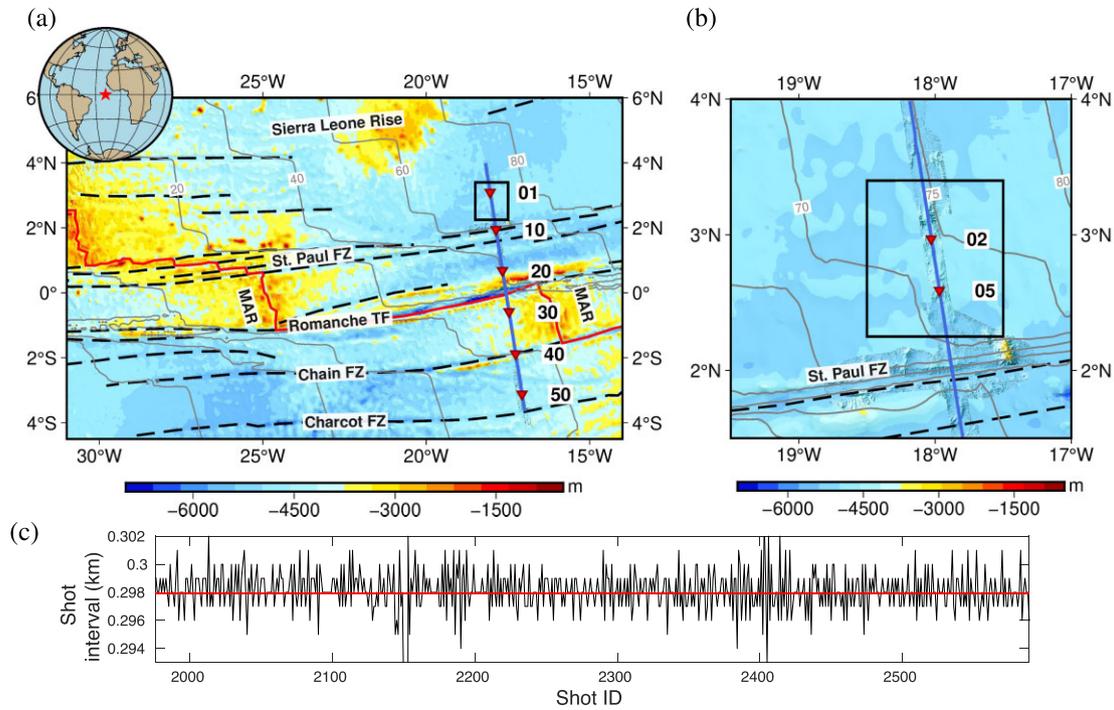
slow-spreading Mid Atlantic Ridge (MAR) with a full spreading rate of  $\sim 46 \text{ mm yr}^{-1}$  (Müller *et al.* 2008; Seton *et al.* 2020). An approximately 855 km long N–S seismic profile running parallel to the MAR at  $18^\circ \text{W}$  was shot with a total of 50 OBSs deployed at an average interval of 14.2 km to record wide-angle seismic data covering 8–70 Ma crust (Wang & Singh 2022; Fig. 9a). An air gun array consisting of 16 guns with a total volume of 4999 cubic inches was towed at 10 m depth below the sea level and shot at an approximate interval of 297.9 m (Fig. 9c). Our study area is located over a 70 Ma crustal segment at  $2^\circ \text{N}$ ,  $18^\circ \text{W}$  (Fig. 9b).

We perform first-arrival traveltome tomography using the inversion methods of Van Avendonk *et al.* (1998, 2004) to obtain a long-wavelength  $P$ -wave velocity structure of the subsurface that could serve as an initial velocity model for the waveform inversion. Before performing any operation, the OBS records were corrected for drift due to the oceanic currents and repositioned for the shots and OBSs locations (Marjanović *et al.* 2020; Wang & Singh 2022). The data were bandpass filtered from 3 Hz to 10 Hz to enhance the signal-to-noise ratio with a dominant frequency between 5 Hz and 7 Hz. We picked around 2500 first arrivals (Pg, Pn) from OBS 1 to OBS 11 and assigned an empirically picking uncertainty of 30 ms and 50 ms to the Pg and Pn arrivals, respectively, based on the quality of the data and visual inspection of errors present in OBSs and shots locations, bathymetry, ray path computations and the picking errors (Wang & Singh 2022). The traveltome is computed using the shortest ray path method (Moser 1991), where the model is discretized at a uniform horizontal and vertical grid spacing of 300 m and 60 m. The inversion is performed using an LSQR method (Paige & Saunders 1982) by iteratively minimizing the cost function defined by the traveltome residuals and picking uncertainties up to a pre-defined threshold value by updating the slowness model with the smoothing and damping constraints to avoid the generation of artefacts. The starting model is constructed by generating a smooth 1-D velocity profile of the slow-spreading MAR (White *et al.* 1992; Grevemeyer *et al.* 2018; Christeson *et al.* 2019) and hung below

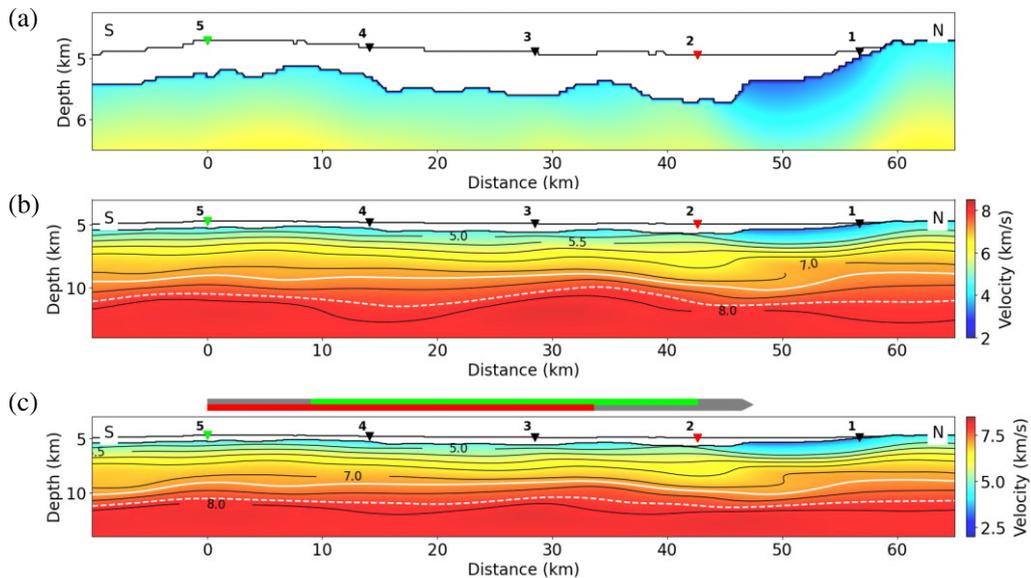
the basement to generate a smooth 2-D velocity model. We used a top-down inversion strategy to obtain a smooth 2-D velocity model from tomography by inverting the traveltome of Pg arrivals followed by joint inversion of the traveltomes of Pg and Pn arrivals (Fig. 10b). The initial 2-D model for FWI was obtained from the tomography model after covering the regions with no ray coverage (Fig. 10c).

The OBS data in this study is taken from a profile north of the St. Paul fracture zone between OBS 5 and OBS 2, marked with the black rectangle on the bathymetric map (Fig. 9). The average seabed depth in the study area is  $\sim 4.84$  km, and the sediment thickness slightly increases from  $\sim 595$  m at OBS 5 to  $\sim 775$  m at OBS 2, with an average thickness of  $\sim 582$  m (Fig. 10a). The OBS data is bandpass filtered from 3 Hz to 6 Hz for a numerically stable inversion on a uniform grid spacing of 49.65 m with a 3 ms time sampling rate. A 3-D to 2-D correction for geometrical spreading is applied to the processed data by multiplying the amplitudes of the observed data by the square root of the two-way traveltome and convolving with the inverse square root of the two-way traveltome to correct for deviations in the resulting phase and amplitude (Pica *et al.* 1990; Shipp & Singh 2002).

We specifically use the seismic data of OBS 2 and OBS 5 to test the application of the HMC in estimating the 1-D approximation of the velocity structure. The two OBSs are separated by 42.6 km. The Pg arrivals on the left-hand side of OBS 2 are observed from 7.65 km to 22.56 km, PmP reflections from 20.18 km to 27.34 km, and Pn arrivals from 25.85 km to 41.66 km (Fig. 11b). The Pg arrivals on the right side of OBS 5 are observed from 6.01 km to 24.21 km, PmP reflections from 22.42 km to 32.86 km and Pn arrivals from 24.51 km to 40.92 km (Fig. 11a). We observe a slight increase in the slope of the Pn arrivals as we move north to south from OBS 2 to OBS 5 and have chosen the data from 9.5 km to 42 km offset in a time window of 0.60 s from the arrival time picks. For the inversion of real seismic data, we require a balancing of amplitudes between the observed and synthetic data for a correct computation of data



**Figure 9.** (a) Bathymetric map of the ILAB-SPARC seismic survey over an 8–70 Ma crust in the equatorial Atlantic Ocean marked with a red star on a globe inset. The seismic profile is shown with a blue line around 18°W, with every tenth OBS marked with a red triangle. (b) The black rectangle shows the area used in this study for the application of HMC on OBS 2 and OBS 5, marked with red triangles. (c) The plot of the shot intervals (black) and their average (red) from shot point 1986 (35 km south of OBS 10) to 2589 (16.8 km north of OBS 1) of the ILAB-SPARC seismic survey with shot points 2343 and 2486 aligned with the position of OBS 5 and OBS 2 on the seabed.

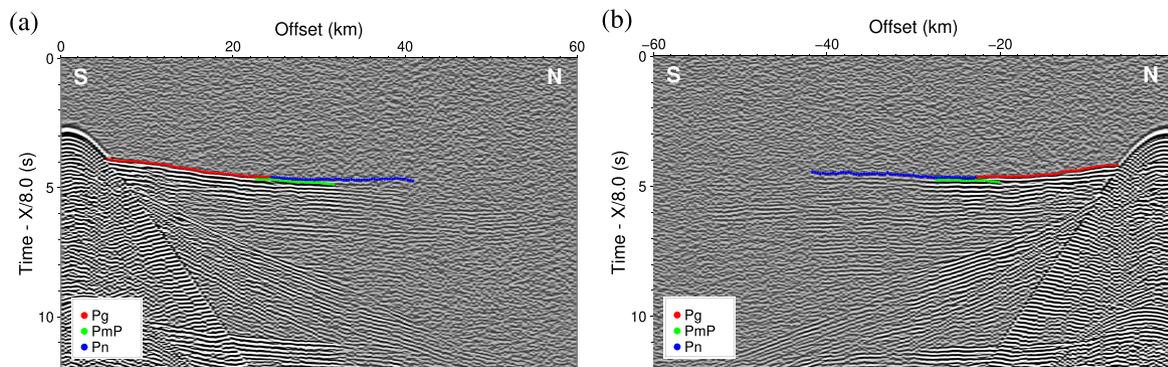


**Figure 10.** (a) Variation of the seabed and basement in the study area. (b) The velocity model from the first-arrival traveltimes tomography. The MTZ is defined between the solid and dashed white line representing the velocity contours of  $7.2 \text{ km s}^{-1}$  and  $7.8 \text{ km s}^{-1}$ . (c) The modified initial velocity model after covering the low-velocity anomalies below the MTZ with no ray coverage. The inverted triangles represent the location of OBS on the selected section of the ILAB-SPARC seismic survey. The grey arrow shows the S–N direction of increasing model distance with the origin at OBS 5. The green and red bands represent the regions used for inverting the seismic data of OBS 5 and OBS 2, marked with green and red inverted triangles, respectively.

misfit and its gradients. We perform this operation by normalizing the synthetic and observed data with their maximum amplitudes in the offset range from 9.5 km to 42 km.

We estimated the source wavelet by extracting the direct wave arrivals of OBS 1 to OBS 10 in the near offsets range from  $-0.6 \text{ km}$

to  $0.6 \text{ km}$ . Then, these arrivals are corrected for normal moveout using a constant water velocity of  $1.5 \text{ km s}^{-1}$  and aligned together to generate the stacked trace. The stacked trace is bandpass filtered from 3 Hz to 6 Hz, and a source wavelet of  $\sim 0.75 \text{ s}$  interval is extracted to avoid contamination from the sedimentary layers. This



**Figure 11.** Seismic data gather of (a) OBS 5 (right-hand side) and (b) OBS 2 (left-hand side) with marked Pg (red), PmP (green) and Pn (blue) seismic arrivals.

source wavelet is verified by comparing the synthetic and observed data at near offsets.

We performed seismic modelling on a 45.0 km long and 14.0 km deep velocity model discretized at a uniform grid spacing of 49.65 m, leading to a total model dimension of  $907 \times 282$  grid-points. We used a data covariance matrix based on an exponential correlation law (Bodin *et al.* 2012) with a constant noise variance  $\sigma_d^2 = 1$  and a correlation factor of 0.25 to account for sparse acquisition geometry with a receiver interval of 297.9 m. A tempered likelihood with temperatures  $T \in [20 - 100]$  is used to flatten the variable likelihood function to increase the efficiency rate of the HMC. The prior distribution is a Gaussian distribution (eq. 10) with a model standard deviation  $\sigma_m$  of  $\{750, 1000\}$  m s<sup>-1</sup> for sampling the model space through 20–30 independent chains.

## 4.2 Results

We use the 1-D initial velocity model (Fig. 13a, black) extracted by averaging the 2-D velocity model from 3.0 km to 10.0 km left of OBS 2. Figs 12(a) and (b) show the comparison of observed (black) and synthetic (red) seismic data from the initial velocity model and the data residuals. We performed the HMC to estimate the model parameters from  $\sim 5.5$  km to 14.0 km depth containing 171  $V_p$  model parameters, keeping the rest of the model above unchanged. Fig. 13(b) shows the results of the HMC mean (red), three standard deviations (yellow) and the maximum a posteriori (MAP) estimate (cyan) with the maximum probability. Fig. 13(c) and (d) show the variation of the posterior probability of each chain and its normalized data misfit function with increasing samples.

The result of HMC provides a good fit to the observed data, as can be seen from the comparison of the observed (black) and synthetic (red) seismic data (Fig. 14a) from the estimated mean and the data residuals (Fig. 14b). The mean velocity model shows a layered structure, marked by changes in velocity gradients. The thickness of these layers varies from  $\sim 500$  m to  $\sim 1000$  m. As the frequency bandwidth of the seismic data is from 3 Hz to 6 Hz, the thickness of these layers corresponds to half of the wavelength of these frequencies in the crust and upper mantle (Virieux & Operto 2009). An increasing gradient around 10 km depth till the maximum velocity of 7.8 km s<sup>-1</sup> indicates  $\sim 0.7$  km thick MTZ. We further discuss the details of this result in the next section.

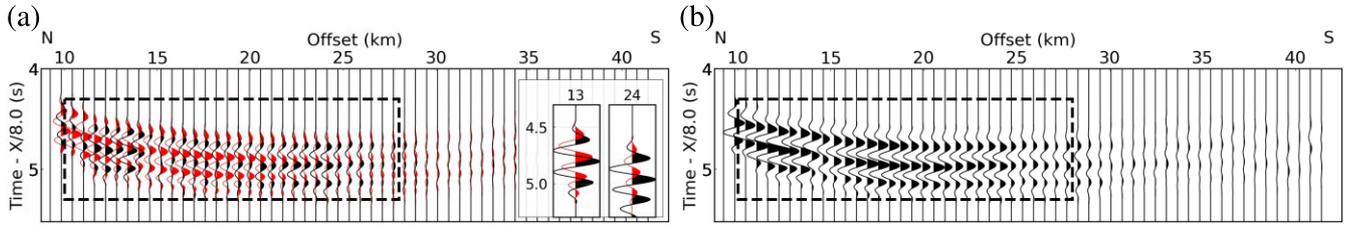
Fig. 16(a) shows the initial 1-D velocity model (black) for OBS 5 extracted by averaging the 2-D velocity model from 3 km to 10 km right of OBS 5. Figs 15(a) and (b) show the comparison of observed (black) and synthetic (red) data from the initial velocity model and the data residuals. Following the same strategy used in

inverting OBS 2, we performed the HMC to estimate the model parameters from  $\sim 5.5$  km to 14.0 km depth containing 173  $V_p$  model parameters, keeping the rest of the model unchanged. Fig. 16(b) shows the results of the HMC mean (red) with three standard deviations (yellow) and the MAP estimate (cyan). Figs 16(c) and (d) show the variation of the posterior probability of each chain and its normalized data misfit function with increasing samples.

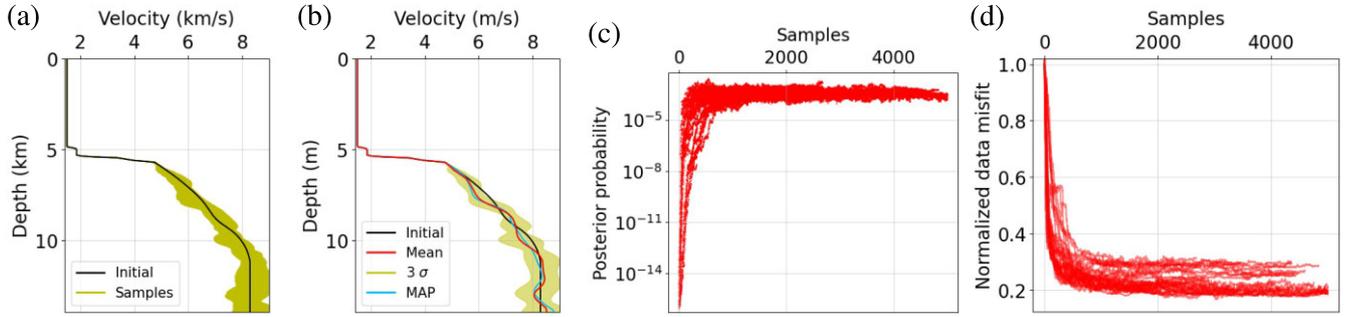
Figs 17(a) and (b) show the comparison of the observed (black) and synthetic (red) seismic data from the estimated mean and the data residuals. The resulting mean velocity model shows a slight decrease in the velocity in both the upper and lower crust from  $\sim 6.3$  km to  $\sim 9.8$  km depth compared to the initial velocity model and a smooth variation of velocity around the MTZ from 10.1 km to 11.2 km depth, indicating  $\sim 1.1$  km thick MTZ.

## 5 DISCUSSION

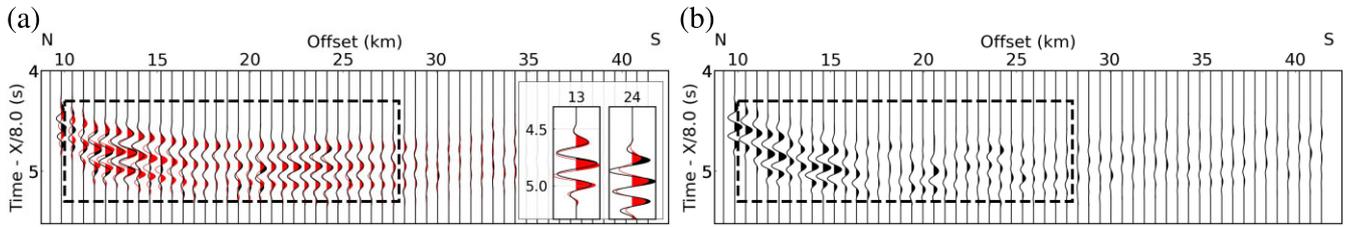
While the estimated mean and standard deviation are used to characterize the centre and spread of model parameters in the sampled posterior distribution, HMC can also provide additional information on the statistical distribution of the model parameters. This information can be obtained from a scatter plot matrix of model parameters, where the diagonal plots showing the 1-D marginal distribution of model parameters provide information on the multimodality of the model parameters, and the off-diagonal plots showing 2-D marginal distribution provide information on the distribution and correlation of two model parameters at different depths. The scatter plot matrix computed from the posterior distribution of OBS 5 (Fig. 19b) shows an approximated Gaussian distribution of the model parameters in the diagonal plots. Conversely, the presence of multiple peaks on the diagonal plots of the scatter plot matrix for OBS 2 (Fig. 19a) indicates deviation from the assumed Gaussian distribution of the PPD. The off-diagonal plots of both matrices show a complex distribution, indicating the non-linear relationship of model parameters with each other. This variation can also be observed from a joint distribution plot of velocities of OBS 5 (Fig. 20b) and OBS 2 (Fig. 20a) at depths of 8.0 km and 10.0 km, marking a shift from an approximated Gaussian to a non-Gaussian distribution of the posterior samples. The deviation from the Gaussian distribution of velocities observed in these plots could result from the computationally constrained sampling of the model space affecting the convergence of the Markov chains, as discussed in the later paragraph. Alternatively, it could also result from complex trade-offs of model parameters in the PPD, arising from the non-uniqueness of the inversion problem, the limited data window used for inversion, and the laterally homogeneous assumption of the 2-D velocity model used in this study.



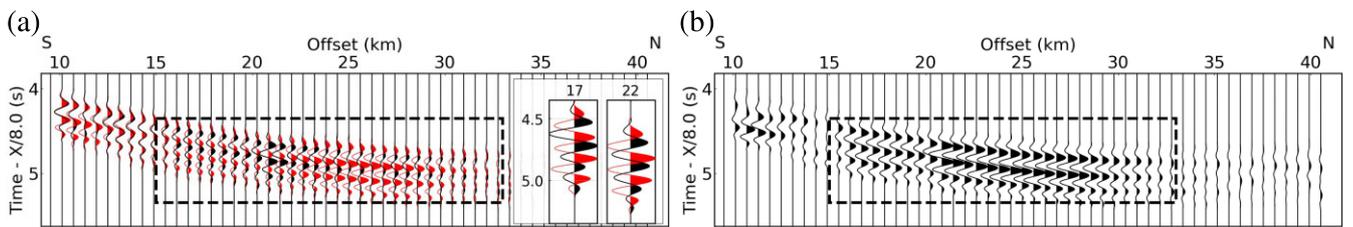
**Figure 12.** (a) Comparison of the observed (black) and the synthetic (red) seismic data from the initial velocity model of OBS 2 and (b) the data residuals plotted with an increasing data offset in the N–S direction. The black box marks the data with large data residuals. The inset plot in panel (a) shows the blow-up of the observed (black) and synthetic (red) seismic data at 13 km and 24 km offsets.



**Figure 13.** (a) Plot of random models sampled on every independent chain. (b) Mean velocity model (red) and three standard deviations (yellow) from the HMC of the OBS 2 data and the MAP estimate (cyan). (c) Variation of the posterior probability of each chain and (d) its normalized data misfit function with increasing samples.



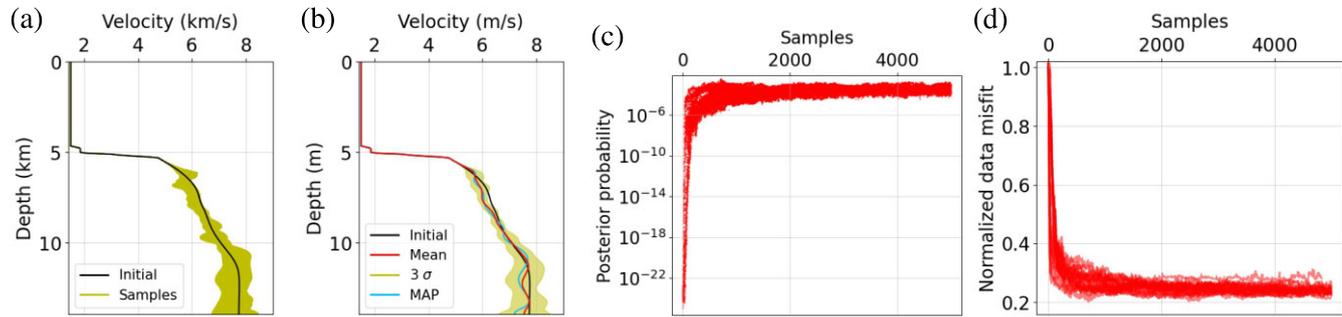
**Figure 14.** Comparison of the observed (black) and synthetic (red) seismic data from (a) the estimated mean of the HMC on OBS 2 data and (b) the data residuals plotted with an increasing data offset in the N–S direction. The black box marks the data with a significant decrease in the data residual. The inset plot in (a) shows the blow-up of the observed (black) and synthetic (red) seismic data at 13 km and 24 km offsets.



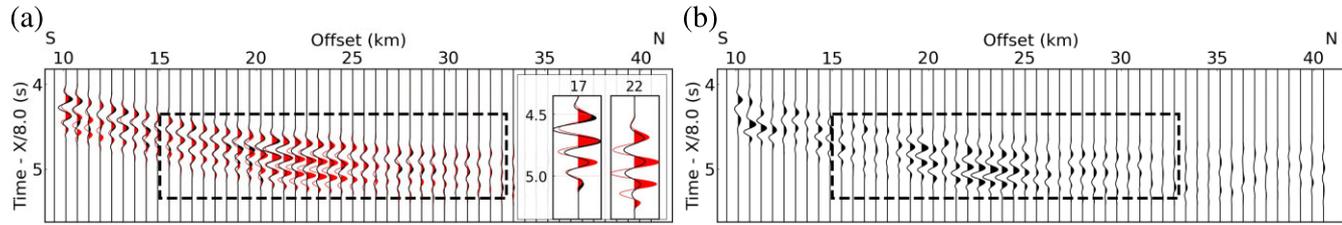
**Figure 15.** (a) Comparison of the observed (black) and synthetic (red) seismic data from the initial velocity model for OBS 5 and (b) the data residuals plotted with an increasing data offset in the S–N direction. The black box marks the data with large data residuals. The inset plot in (a) shows the blow-up of the observed (black) and synthetic (red) seismic data at 17 km and 22 km offsets.

Though the first arrivals of the final synthetics at both OBSs fit the observed data (Figs 14 and 17), the presence of comparatively large misfits of the later arrivals in the time window of 0.6 s suggests a degree of lateral heterogeneity present in the velocity model. Also, the structures in the MTZ and upper mantle for both OBS 2 and OBS 5 are constrained by the PmP and Pn seismic arrivals at offsets of >20 km and are limited in resolution with the frequency bandwidth of the data from 3 Hz to 6 Hz. Therefore, the 1-D approximations of a 2-D velocity model at both OBSs should be interpreted with caution. Fig. 18 shows the mean velocity models and three standard

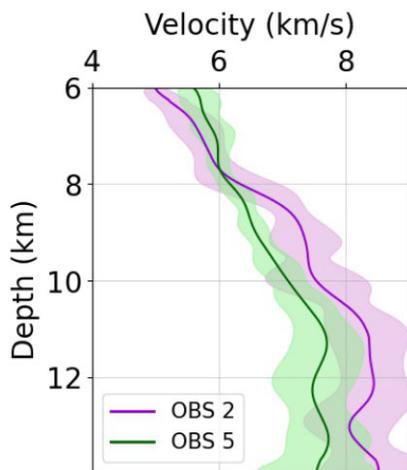
deviations from the result of the HMC for OBS 2 and OBS 5. The mean velocity model in the part of the upper crust up to ~7.7 km depth shows a slightly higher velocity from OBS 5 than the velocity from OBS 2. Below this depth, the velocity from OBS 5 is ~0.50 km s<sup>-1</sup> to 0.75 km s<sup>-1</sup> lower than that from OBS 2. Furthermore, the velocity depth profile shows more layered structures from OBS 2 than those observed from OBS 5. These differences could be due to the mapping of 2-D structures into two 1-D models that are 20–30 km apart. Assuming an increase in velocity gradient around the 10 km depth from 7.2 km s<sup>-1</sup> to a maximum velocity of 7.8 km s<sup>-1</sup>



**Figure 16.** (a) Plot of random models sampled on every independent chain. (b) Mean velocity model (red) and three standard deviations (yellow) from the HMC of the OBS 5 data and the MAP estimate (cyan). (c) Variation of the posterior probability of each chain and (d) its normalized data misfit function with increasing samples.



**Figure 17.** Comparison of observed (black) and synthetic (red) seismic data from (a) the estimated mean of the HMC on OBS 5, and (b) the data residuals plotted with an increasing data offset in the S–N direction. The black box marks the data with a significant decrease in the data residuals. The inset plot in (a) shows the blow-up of the observed (black) and synthetic (red) seismic data at 13 km and 24 km offsets.

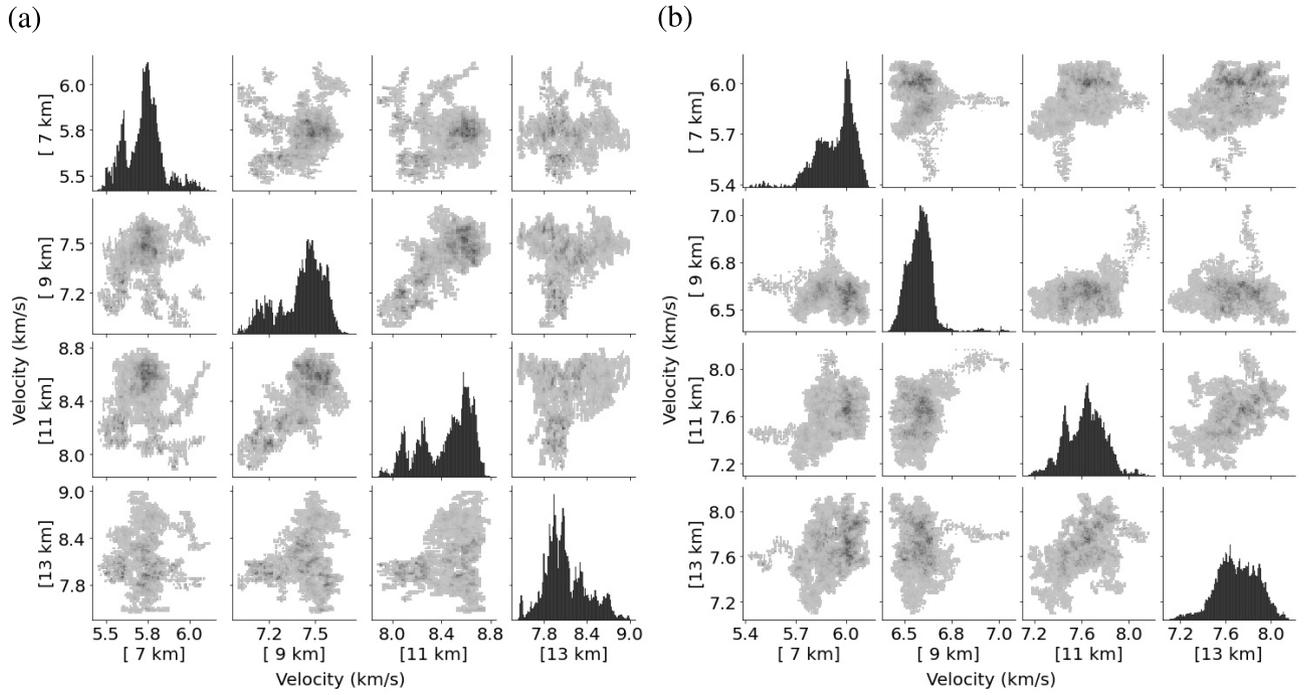


**Figure 18.** Blow-up of the result of the HMC from OBS 2 (purple) and OBS 5 (green) with a mean velocity model and three standard deviations.

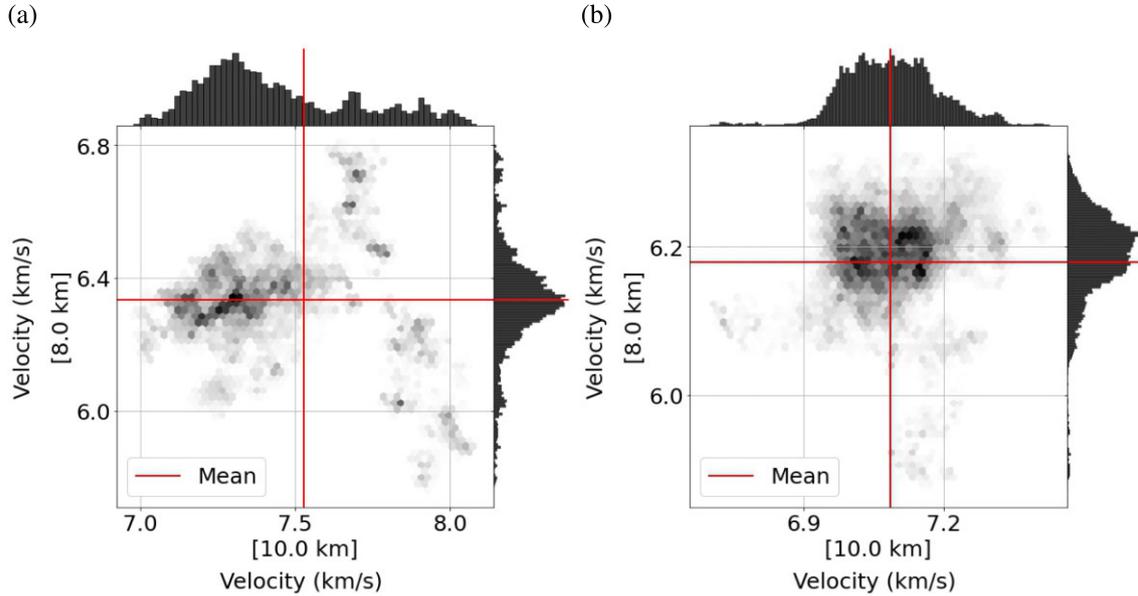
as the MTZ, the thickness of the MTZ would be  $\sim 1.1$  km for OBS 5 and  $\sim 0.7$  km for OBS 2, leading to a crustal thickness of  $\sim 5.1$  km for OBS 5 and  $\sim 4.3$  km for OBS 2. The higher velocity up to 7.7 km depth and the monotonous increase of velocity in the upper crust for OBS 5 might be related to ultramafic cumulates or an increase in serpentinization with depth. This interpretation would be consistent with a thick MTZ and a lower upper mantle velocity ( $7.8 \text{ km s}^{-1}$ ). For OBS 2, the uppermost high-velocity gradient layer down to  $\sim 6.2$  km depth might represent Layer 2A. The slightly low-velocity gradient layer down to  $\sim 7.7$  km depth might be due to the dike sequence and represent Layer 2B. The high-velocity gradient layer between  $\sim 7.7$  km and  $\sim 8.4$  km depth could be a transition from dike to gabbro. The nearly constant velocity below this depth down to the MTZ might be Layer 3 (gabbro).

Structures with alternate high and low-velocity layers have been observed by Guo *et al.* (2022) and interpreted as a result of *in situ* crystallization of the melt in the lower crust. However, we do not observe such structures, possibly due to the 1-D approximation used here. Below the MTZ, the velocities of  $\sim 8.3 \text{ km s}^{-1}$  are rather high for a normal mantle velocity. The high mantle velocities of  $\sim 8.3 \text{ km s}^{-1}$  were also observed by Vaddineni *et al.* (2021) and interpreted as a low degree of melting. In the upper mantle, both OBS results indicate the presence of low-velocity structures, which could be due to layering in the mantle or artefacts. However, the velocity structure for OBS 2 indicates that the crust is likely formed by a magmatic process (Wang & Singh 2022). These results can be further improved with a better probabilistic 2-D FWI of the OBSs while considering the lateral heterogeneity of the velocity model.

The numerical workflow of the HMC gives it an edge in generating less correlated samples with increasing dimensionality of the inverse problem. The computation time for generating a statistically independent sample in the HMC grows as a power of  $n^{5/4}$  compared to  $n^2$  in the standard Metropolis–Hastings algorithm (Neal *et al.* 2011) for  $n$ -dimensional parameters. To maintain this edge, tuning parameters play a crucial role in the efficiency and convergence of the HMC. The main computational cost of the HMC comes from the gradients computed  $L + 1$  times over a trajectory length of  $L\epsilon$  in every iterative workflow. This restriction severely limits the number of leapfrog steps for an efficient implementation of the algorithm. We performed numerical computations using the S-CAPAD/DANTE platform at IPGP with a single node consisting of 32–64 CPUs, each having 2/4 GB of memory and used  $L = 1$  to maintain the computational efficiency of the HMC. Thus, the step size and the mass matrix are the only tuning parameters used in this study. The computational time of the HMC with leapfrog steps  $L = 1$  for the synthetic test-1 and test-2 with model dimensions of  $1000 \times 200$  and  $1000 \times 351$  gridpoints, and  $160 V_p$  and  $306 V_p$  model



**Figure 19.** Scatter plot matrix computed from the posterior distribution of (a) OBS 2 and (b) OBS 5 at individual depths of 7.0 km, 9.0 km, 11.0 km and 13.0 km. The diagonal plots show the 1-D marginal distribution of model parameters, representing the multi-modality of the solution. On the other hand, off-diagonal plots represent the 2-D distribution of model parameters at different pairs of depths, with darker shades indicating increasing density of the sampled model parameters.

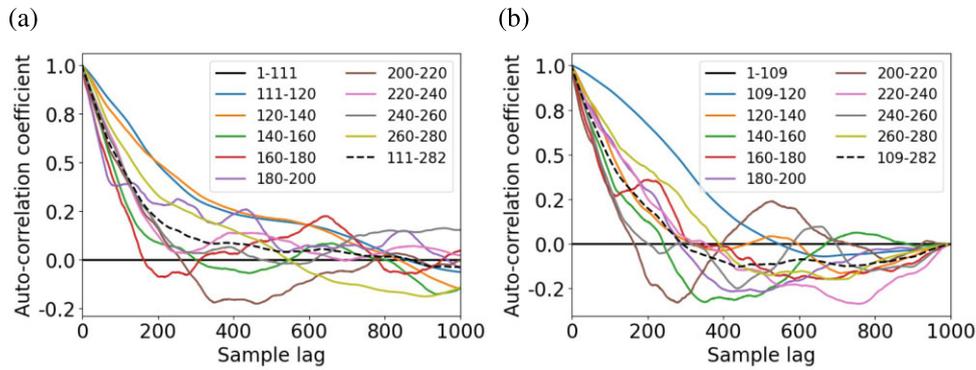


**Figure 20.** Joint distribution plots of (a) OBS 2 and (b) OBS 5 showing the deviation from the assumed Gaussian distribution of the PPD at depths of 8.0 km and 10.0 km. Red lines mark the mean velocities, while darker shades indicate increasing density of sampled velocities at the selected depths.

parameters takes  $\sim 56$  hr and  $\sim 98$  hr for sampling around 5000 models. In contrast, the computational time of FWI in test-1 and test-2 takes  $\sim 25$  min and  $\sim 45$  min for a total of 90 iterations. Similarly, the HMC of both OBS data sets with model dimensions of  $907 \times 282$  gridpoints and 171  $V_p$  and 173  $V_s$  model parameters takes 48 hr to sample around 3500–5000 models.

The value of step size  $\epsilon$  is empirically chosen based on factors of model updates associated with generating the random momentum variable and misfit gradient. We tune it at specific intervals

to maintain an appropriate acceptance rate. The remaining tuning parameter,  $\mathbf{M}$  (mass matrix), is inversely proportional to the sensitivity of the momentum vector in generating random velocity models by sampling new states along the particle trajectory. Therefore, effectively tuning the mass matrix can improve the convergence rate by increasing the range of model updates sampled in the HMC. Here, we have computed the diagonal mass matrix at regular intervals using the first-order estimation of the inverse posterior covariance and kept the maximum value of the mass



**Figure 21.** Sample autocorrelation plot of a specific individual chain of HMC for (a) OBS 2 and (b) OBS 5 averaged over all model parameters. The individual lines represent the average autocorrelation coefficients computed between the mentioned model parameters.

matrix to 1.0, corresponding to the minimum permissible model updates.

To assess the convergence of multiple chains, we computed the autocorrelation (Geyer 1992; Gelman *et al.* 2013) of the model parameters from the sampled PPD. In the context of MCMC, autocorrelation is a measure of independent samples and is indicative of the efficiency of the MCMC chain over an expansive model space. For this purpose, the effective sample size ( $N_{\text{eff}}$ ) reflects the number of statistically independent samples present in the sampled posterior distribution and is estimated as

$$N_{\text{eff}} = \frac{N}{1 + 2 \sum_{\gamma=1}^{\text{inf}} \rho(\gamma)}, \quad (13)$$

where  $N$  is the total number of samples and  $\rho(\gamma)$  is the normalized autocorrelation function, computed as  $c(\gamma)/c(0)$  for the autocorrelation  $c(\gamma)$  at a sample lag  $\gamma$ . Here, we have computed the autocorrelation of model parameters of an individual chain from the posterior samples of OBS 2 (Fig. 21a) and OBS 5 (Fig. 21b) starting from sample 500. We estimated the ratio of  $N_{\text{eff}}/N$  as a representative of uncorrelated samples present in the PPD by summing the normalized autocorrelation function  $\rho(\gamma)$  up to the lag where its first successive summed value becomes negative (Gelman *et al.* 2013; Fichtner *et al.* 2021). The average value of this ratio over the group of model parameters in our study varies from  $\sim 0.005$  to  $\sim 0.02$ , indicating the presence of statistically one independent sample for every  $\sim 50$  to  $\sim 200$  samples.

The results of the HMC on the synthetic and OBS data show its advantage in adequately sampling the posterior distribution. Compared to the FWI, the HMC shows considerable independence on the initial velocity model. We have, however, not explored the extent of this independence on the initial velocity models. For practical purposes, the construction of PPD in this study uses the Gaussian probability distributions to sample the model and momentum variables and to determine the multivariate Gaussian likelihood function. These distributions are prone to tempering and might create a bias in the resulting PPD. Using the probability distributions based on more robust norms of kinetic energy (Livingstone *et al.* 2019) and appropriate data misfit functions (Luo & Schuster 1991; Brossier *et al.* 2010; Bozdağ *et al.* 2011; Métivier *et al.* 2016) can increase the efficiency of the HMC. The constraints posed by the computational resources play a critical role in the extensive applicability of the algorithm and tuning strategies used in this study for the inversion of the observed seismic data using 1-D velocity models. However, the increased efficiency of the HMC with larger dimensional problems indicates its future potential in quantifying 2-D

and 3-D laterally heterogeneous structures with a better numerical framework and high-performance computational facilities.

## 6 CONCLUSION

FWI of the wide-angle seismic data in the presence of a high-velocity contrast (e.g. MTZ) is prone to converge to local minima due to the complex behaviour of reflection coefficients starting from the critical angle. The additional interference of different seismic arrivals and the inherent errors in the recorded seismic data increase the non-linearity of the inverse problem. In this study, we have presented the practical results of HMC based elastic FWI of wide-angle seismic data for understanding the nature of the lower crust and MTZ over a 70 Ma crustal segment in the equatorial Atlantic Ocean. The application was successfully tested against the limitations of FWI with the inversion of synthetic seismic data sets from two groups of 1-D velocity profiles using a time-domain finite difference solution of the 2-D elastic wave equation. The efficiency of the HMC in this study depends on an appropriate tuning strategy of the diagonal mass matrix and the empirically chosen step size limited by the computational constraints of the leapfrog steps  $L = 1$ .

The application of HMC to OBS data from the equatorial Atlantic Ocean indicates the presence of a  $\sim 1.1$  km thick MTZ for OBS 5 and  $\sim 0.7$  km thick MTZ for OBS 2. The presence of a high velocity in the upper crust, uniform and lower velocity in the crust and relatively low velocity in the mantle, combined with a thick MTZ from OBS 5, suggests a gabbroic or partially serpentinized crust with MTZ probably composed of ultramafic cumulates with a certain degree of serpentinization. On the other hand, the presence of layered structures in the crust, overall high velocities in the crust and upper mantle and a relatively thin MTZ from OBS 2 suggest that the crust is formed by a magmatic process.

## ACKNOWLEDGMENTS

This project was funded by the European Research Council Advanced Grant Agreement No. 339442 TransAtlantic ILAB. We thank the computational facilities provided by the S-CAPAD/DANTE platform at IPGP. Special thanks to Zhikai Wang for providing the OBS data and to Guo Peng from CSIRO for his advice during the initial stages of developing the HMC algorithm used in this study. We are grateful to Alexandrine Gesret for providing constructive reviews and to the two anonymous reviewers for

their valuable comments on improving the paper. This is an IPGP contribution number 4286.

## DATA AVAILABILITY

The OBS 2 and OBS 5 data used in this study are available online at <https://doi.pangaea.de/10.1594/PANGAEA.937195> (Grove *et al.* 2021).

## REFERENCES

- Aki, K. & Richards, P.G., 2002. *Quantitative Seismology*, 2nd edn, University Science Books.
- Askan, A., Akcelik, V., Bielak, J. & Ghattas, O., 2007. Full waveform inversion for seismic velocity and anelastic losses in heterogeneous structures, *Bull. seism. Soc. Am.*, **97**(6), 1990–2008.
- Backus, G. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross earth data, *Phil. Trans. R. Soc. Lond., A*, **266**(1173), 123–192.
- Backus, G.E., 1988. Comparing hard and soft prior bounds in geophysical inverse problems, *Geophys. J. Int.*, **94**(2), 249–261.
- Bamberger, A., Chavent, G., Hemon, C. & Lailly, P., 1982. Inversion of normal incidence seismograms, *Geophysics*, **47**(5), 757–770.
- Ben-Hadj-Ali, H., Operto, S. & Virieux, J., 2011. An efficient frequency-domain full waveform inversion method using simultaneous encoded sources, *Geophysics*, **76**(4), R109–R124.
- Betancourt, M., 2017. *A conceptual introduction to Hamiltonian Monte Carlo*, arXiv: 1701.02434.
- Bijwaard, H., Spakman, W. & Engdahl, E.R., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.*, **103**(B12), 30 055–30 078.
- Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, **178**(3), 1411–1436.
- Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.*, **117**(B2), doi:10.1029/2011JB008560.
- Borisov, D. & Singh, S.C., 2015. Three-dimensional elastic full waveform inversion in a marine environment using multicomponent ocean-bottom cables: a synthetic study, *Geophys. J. Int.*, **201**(3), 1215–1234.
- Bottero, A., Gesret, A., Romary, T., Noble, M. & Maisons, C., 2016. Stochastic seismic tomography by interacting Markov chains, *Geophys. J. Int.*, **207**(1), 374–392.
- Bozdağ, E., Trampert, J. & Tromp, J., 2011. Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements, *Geophys. J. Int.*, **185**(2), 845–870.
- Brossier, R., Operto, S. & Virieux, J., 2009. Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion, *Geophysics*, **74**(6), WCC105–WCC118.
- Brossier, R., Operto, S. & Virieux, J., 2010. Which data residual norm for robust elastic frequency-domain full waveform inversion?, *Geophysics*, **75**(3), R37–R46.
- Bunks, C., Saleck, F.M., Zaleski, S. & Chavent, G., 1995. Multiscale seismic waveform inversion, *Geophysics*, **60**(5), 1457–1473.
- Capdeville, Y., Gung, Y. & Romanowicz, B., 2005. Towards global earth tomography using the spectral element method: a technique based on source stacking, *Geophys. J. Int.*, **162**(2), 541–554.
- Cary, P. & Chapman, C., 1988. Automatic 1-D waveform inversion of marine seismic refraction data, *Geophys. J. Int.*, **93**(3), 527–546.
- Castagna, J.P., Batzle, M.L. & Eastwood, R.L., 1985. Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks, *Geophysics*, **50**(4), 571–581.
- Chapman, C. & Orcutt, J., 1985. Least-squares fitting of marine seismic refraction data, *Geophys. J. Int.*, **82**(3), 339–374.
- Choi, Y. & Alkhalifah, T., 2011. Source-independent time-domain waveform inversion using convolved wavefields: application to the encoded multisource waveform inversion, *Geophysics*, **76**(5), R125–R134.
- Choi, Y. & Alkhalifah, T., 2012. Application of multi-source waveform inversion to marine streamer data using the global correlation norm, *Geophys. Prospect.*, **60**(4), 748–758.
- Choy, G.L. & Richards, P.G., 1975. Pulse distortion and Hilbert transformation in multiply reflected and refracted body waves, *Bull. seism. Soc. Am.*, **65**(1), 55–70.
- Christeson, G., Goff, J. & Reece, R., 2019. Synthesis of oceanic crustal structure from two-dimensional seismic profiles, *Rev. Geophys.*, **57**(2), 504–529.
- Duane, S., Kennedy, A.D., Pendleton, B.J. & Roweth, D., 1987. Hybrid Monte Carlo, *Phys. Lett. B*, **195**(2), 216–222.
- Dziewonski, A.M., 1984. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.*, **89**(B7), 5929–5952.
- Dziewonski, A.M., Hager, B.H. & O’Connell, R.J., 1977. Large-scale heterogeneities in the lower mantle, *J. geophys. Res.*, **82**(2), 239–255.
- Fichtner, A. & Simutè, S., 2018. Hamiltonian Monte Carlo inversion of seismic sources in complex media, *J. geophys. Res.*, **123**(4), 2984–2999.
- Fichtner, A. & Zunino, A., 2019. Hamiltonian nullspace shuttles, *Geophys. Res. Lett.*, **46**(2), 644–651.
- Fichtner, A., Bunge, H.-P. & Igel, H., 2006. The adjoint method in seismology: I. Theory, *Phys. Earth planet. Inter.*, **157**(1–2), 86–104.
- Fichtner, A., Kennett, B.L., Igel, H. & Bunge, H.-P., 2008. Theoretical background for continental-and global-scale full-waveform inversion in the time–frequency domain, *Geophys. J. Int.*, **175**(2), 665–685.
- Fichtner, A., Kennett, B.L., Igel, H. & Bunge, H.-P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.*, **179**(3), 1703–1725.
- Fichtner, A., Zunino, A. & Gebräad, L., 2019. Hamiltonian Monte Carlo solution of tomographic inverse problems, *Geophys. J. Int.*, **216**(2), 1344–1363.
- Fichtner, A., Zunino, A., Gebräad, L. & Boehm, C., 2021. Autotuning Hamiltonian Monte Carlo for efficient generalized nullspace exploration, *Geophys. J. Int.*, **227**(2), 941–968.
- Fliedner, M., White, R. & Smallwood, J., 1998. Seismic velocity structure of basalt flows, in *SEG Technical Program Expanded Abstracts*, pp. 1178–1181, Society of Exploration Geophysicists.
- French, S. & Romanowicz, B.A., 2014. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography, *Geophys. J. Int.*, **199**(3), 1303–1327.
- Gebräad, L., Boehm, C. & Fichtner, A., 2020. Bayesian elastic full-waveform inversion using Hamiltonian Monte Carlo, *J. geophys. Res.*, **125**(3), e2019JB018428.
- Gee, L.S. & Jordan, T.H., 1992. Generalized seismological data functionals, *Geophys. J. Int.*, **111**(2), 363–390.
- Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A. & Rubin, D.B., 2013. *Bayesian Data Analysis*, 3rd edn, Chapman and Hall/CRC.
- Geyer, C.J., 1992. Practical Markov Chain Monte Carlo, *Stat. Sci.*, **7**(4), 473–483.
- Green, P.J., 1995. Reversible jump Markov Chain Monte Carlo computation and Bayesian model determination, *Biometrika*, **82**(4), 711–732.
- Grevemeyer, I., Ranero, C.R. & Ivandic, M., 2018. Structure of oceanic crust and serpentinization at subduction trenches, *Geosphere*, **14**(2), 395–418.
- Grove, K. *et al.*, 2021. Seismic structure of the St. Paul fracture zone and late Cretaceous to Mid Eocene oceanic crust in the equatorial Atlantic ocean near 18°W, *J. geophys. Res.*, **126**(11), e2021JB022456.
- Guo, P., Singh, S.C., Vaddineni, V.A., Visser, G., Grevemeyer, I. & Saygin, E., 2021. Nonlinear full waveform inversion of wide-aperture OBS data for Moho structure using a trans-dimensional bayesian method, *Geophys. J. Int.*, **224**(2), 1056–1078.
- Guo, P., Singh, S.C., Vaddineni, V.A., Grevemeyer, I. & Saygin, E., 2022. Lower oceanic crust formed by in situ melt crystallization revealed by seismic layering, *Nat. Geosci.*, **15**(7), 591–596.
- Hamilton, E.L., 1978. Sound velocity–density relations in sea-floor sediments and rocks, *J. acoust. Soc. Am.*, **63**(2), 366–377.
- Hobro, J., Minshull, T., Singh, S. & Chand, S., 2005. A three-dimensional seismic tomographic study of the gas hydrate stability zone, offshore Vancouver island, *J. geophys. Res.*, **110**(B9), doi:10.1029/2004JB003477.

- Hong, T. & Sen, M.K., 2009. A new MCMC algorithm for seismic waveform inversion and corresponding uncertainty analysis, *Geophys. J. Int.*, **177**(1), 14–32.
- Keilis-Borok, V. & Yanovskaja, T., 1967. Inverse problems of seismology (structural review), *Geophys. J. Int.*, **13**(1–3), 223–234.
- Korenaga, J., Holbrook, W., Kent, G., Kelemen, P., Detrick, R., Larsen, H.-C., Hopper, J. & Dahl-Jensen, T., 2000. Crustal structure of the southeast Greenland margin from joint refraction and reflection seismic tomography, *J. geophys. Res.*, **105**(B9), 21 591–21 614.
- Krebs, J.R., Anderson, J.E., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A. & Lacasse, M.-D., 2009. Fast full-wavefield seismic inversion using encoded sources, *Geophysics*, **74**(6), WCC177–WCC188.
- Lailly, P. & Bednar, J., 1983. The seismic inverse problem as a sequence of before stack migrations, in *Proceedings of the Conference on Inverse Scattering, Theory and Application*, Society for Industrial and Applied Mathematics, Expanded Abstracts, pp. 206–220.
- Levander, A.R., 1988. Fourth-order finite-difference P-SV seismograms, *Geophysics*, **53**(11), 1425–1436.
- Liu, Q. & Tromp, J., 2006. Finite-frequency kernels based on adjoint methods, *Bull. seism. Soc. Am.*, **96**(6), 2383–2397.
- Livingstone, S., Faulkner, M.F. & Roberts, G.O., 2019. Kinetic energy choice in Hamiltonian/Hybrid Monte Carlo, *Biometrika*, **106**(2), 303–319.
- Luo, Y. & Schuster, G.T., 1991. Wave-equation traveltime inversion, *Geophysics*, **56**(5), 645–653.
- Mahalanobis, P.C., 1936. On the generalized distance in statistics, *Proc. Natl. Inst. Sci. India*, **2**, 49–55.
- Maiti, S., Gupta, G., Erram, V.C. & Tiwari, R., 2011. Inversion of Schlumberger resistivity sounding data from the critically dynamic Koyna region using the hybrid Monte Carlo-based neural network approach, *Nonlin. Process. Geophys.*, **18**(2), 179–192.
- Malinverno, A., 2002. Parsimonious bayesian Markov Chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, **151**(3), 675–688.
- Marinari, E. & Parisi, G., 1992. Simulated tempering: a new Monte Carlo scheme, *EPL (Europhys. Lett.)*, **19**(6), doi:10.1209/0295-5075/19/6/002.
- Marjanović, M. et al., 2020. Seismic crustal structure and morphotectonic features associated with the chain fracture zone and their role in the evolution of the equatorial Atlantic region, *J. geophys. Res.*, **125**(10), e2020JB020275.
- Métivier, L., Brossier, R., Méritog, Q., Oudet, E. & Virieux, J., 2016. Measuring the misfit between seismograms using an optimal transport distance: application to full waveform inversion, *Geophys. Suppl. Mon. Not. R. astron. Soc.*, **205**(1), 345–377.
- Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.*, **21**(6), 1087–1092.
- Mora, P., 1987. Nonlinear two-dimensional elastic inversion of multioffset seismic data, *Geophysics*, **52**(9), 1211–1228.
- Morgan, J., Warner, M., Bell, R., Ashley, J., Barnes, D., Little, R., Roele, K. & Jones, C., 2013. Next-generation seismic experiments: wide-angle, multi-azimuth, three-dimensional, full-waveform inversion, *Geophys. J. Int.*, **195**(3), 1657–1678.
- Morgan, J., Warner, M., Arnoux, G., Hooft, E., Toomey, D., VanderBeek, B. & Wilcock, W., 2016. Next-generation seismic experiments—II: wide-angle, multi-azimuth, 3-D, full-waveform inversion of sparse field data, *Geophys. J. Int.*, **204**(2), 1342–1363.
- Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.*, **100**(B7), 12 431–12 447.
- Moser, T., 1991. Shortest path calculation of seismic rays, *Geophysics*, **56**(1), 59–67.
- Müller, R.D., Sdrolias, M., Gaina, C. & Roest, W.R., 2008. Age, spreading rates, and spreading asymmetry of the world's ocean crust, *Geochem., Geophys., Geosyst.*, **9**(4), doi:10.1029/2007GC001743.
- Neal, R.M., 1993. Probabilistic inference using Markov Chain Monte Carlo methods, Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto Toronto, ON, Canada.
- Neal, R.M., 1996. *Bayesian Learning for Neural Networks, Lecture Notes in Statistics*, Springer.
- Neal, R.M. et al., 2011. MCMC using Hamiltonian dynamics, in *Handbook of Markov Chain Monte Carlo*, pp. 113–162, eds Brooks, S., Gelman, A., Jones, G. & Meng, X.-L., Chapman & Hall/CRC Press.
- Nocedal, J., 1980. Updating quasi-Newton matrices with limited storage, *Math. Comput.*, **35**(151), 773–782.
- Operto, S., Ravaut, C., Imbrota, L., Virieux, J., Herrero, A. & Dell'Aversana, P., 2004. Quantitative imaging of complex structures from dense wide-aperture seismic data by multiscale traveltime and waveform inversions: a case study, *Geophys. Prospect.*, **52**(6), 625–651.
- Operto, S., Virieux, J., Dessa, J.-X. & Pascal, G., 2006. Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: application to the eastern Nankai Trough, *J. geophys. Res.*, **111**(B9), doi:10.1029/2005JB003835.
- Paige, C.C. & Saunders, M.A., 1982. LSQR: an algorithm for sparse linear equations and sparse least squares, *ACM Trans. Math. Softw. (TOMS)*, **8**(1), 43–71.
- Piana Agostinetti, N., Giacomuzzi, G. & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.*, **201**(3), 1598–1617.
- Pica, A., Diet, J. & Tarantola, A., 1990. Nonlinear inversion of seismic reflection data in a laterally invariant medium, *Geophysics*, **55**(3), 284–292.
- Plessix, R.-E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.*, **167**(2), 495–503.
- Pratt, R.G., Shin, C. & Hick, G., 1998. Gauss–Newton and full newton methods in frequency–space seismic waveform inversion, *Geophys. J. Int.*, **133**(2), 341–362.
- Press, F., 1968. Earth models obtained by Monte Carlo inversion, *J. geophys. Res.*, **73**(16), 5223–5234.
- Ravaut, C., Operto, S., Imbrota, L., Virieux, J., Herrero, A. & Dell'Aversana, P., 2004. Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt, *Geophys. J. Int.*, **159**(3), 1032–1056.
- Sambridge, M., 2014. A parallel tempering algorithm for probabilistic sampling and multimodal optimization, *Geophys. J. Int.*, **196**(1), 357–374.
- Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.*, **167**(2), 528–542.
- Scales, J.A. & Tenorio, L., 2001. Prior information and uncertainty in inverse problems, *Geophysics*, **66**(2), 389–397.
- Seah, Y.-L., Shang, J., Ng, H.K., Nott, D.J. & Englert, B.-G., 2015. Monte Carlo sampling from the quantum state space. II, *New J. Phys.*, **17**(4), doi:10.1088/1367-2630/17/4/043018.
- Sears, T.J., Singh, S. & Barton, P., 2008. Elastic full waveform inversion of multi-component OBC seismic data, *Geophys. Prospect.*, **56**(6), 843–862.
- Sen, M.K. & Biswas, R., 2017. Transdimensional seismic inversion using the reversible jump Hamiltonian Monte Carlo algorithm, *Geophysics*, **82**(3), R119–R134.
- Sen, M.K. & Stoffa, P.L., 1991. Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics*, **56**(10), 1624–1638.
- Seton, M. et al., 2020. A global data set of present-day oceanic crustal age and seafloor spreading parameters, *Geochem., Geophys., Geosyst.*, **21**(10), doi:10.1029/2020GC009214.
- Shaw, P.R. & Orcutt, J.A., 1985. Waveform inversion of seismic refraction data and applications to young pacific crust, *Geophys. J. Int.*, **82**(3), 375–414.
- Sheen, D.-H., Tuncay, K., Baag, C.-E. & Ortoleva, P.J., 2006. Time domain Gauss–Newton seismic waveform inversion in elastic media, *Geophys. J. Int.*, **167**(3), 1373–1384.
- Shipp, R.M. & Singh, S.C., 2002. Two-dimensional full wavefield inversion of wide-aperture marine seismic streamer data, *Geophys. J. Int.*, **151**(2), 325–344.
- Simo, J.C., Tarnow, N. & Wong, K., 1992. Exact energy-momentum conserving algorithms and symplectic schemes for nonlinear dynamics, *Comput. Methods Appl. Mech. Eng.*, **100**(1), 63–116.

- Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: a strategy for selecting temporal frequencies, *Geophysics*, **69**(1), 231–248.
- Stein, S. & Wysession, M., 2009. *An Introduction to Seismology, Earthquakes, and Earth Structure*, John Wiley & Sons.
- Tarantola, A., 1984a. Inversion of seismic data in acoustic approximation, *Geophysics*, **49**, 1259–1266.
- Tarantola, A., 1984b. Linearized inversion of seismic reflection data, *Geophys. Prospect.*, **32**(6), 998–1015.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.
- Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.*, **20**(2), 219–232.
- Vaddineni, V.A., Singh, S.C., Grevemeyer, I., Audhkhasi, P. & Papenberg, C., 2021. Evolution of the crustal and upper mantle seismic structure from 0–27 Ma in the equatorial Atlantic Ocean at 2 43° S, *J. geophys. Res.*, **126**(6), e2020JB021390.
- Van Avendonk, H.J., Harding, A.J., Orcutt, J.A. & McClain, J.S., 1998. A two-dimensional tomographic study of the Clipperton transform fault, *J. geophys. Res.*, **103**(B8), 17 885–17 899.
- Van Avendonk, H.J., Shillington, D.J., Holbrook, W.S. & Hornbach, M.J., 2004. Inferring crustal structure in the Aleutian Island Arc from a sparse wide-angle seismic data set, *Geochem., Geophys., Geosyst.*, **5**(8), doi:10.1029/2003GC000664.
- Van der Hilst, R.D., Widiyantoro, S. & Engdahl, E., 1997. Evidence for deep mantle circulation from global tomography, *Nature*, **386**(6625), 578–584.
- Van der Voo, R., Spakman, W. & Bijwaard, H., 1999. Tethyan subducted slabs under India, *Earth planet. Sci. Lett.*, **171**(1), 7–20.
- Virieux, J., 1986. P-SV wave propagation in heterogeneous media: velocity-stress finite-difference method, *Geophysics*, **51**(4), 889–901.
- Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics*, **74**(6), WCC1–WCC26.
- Wang, Z. & Singh, S.C., 2022. Seismic evidence for uniform crustal accretion along slow-spreading ridges in the equatorial Atlantic Ocean, *Nat. Commun.*, **13**(1), doi:10.1038/s41467-022-35459-z.
- Warner, M. *et al.*, 2013. Anisotropic 3D full-waveform inversion, *Geophysics*, **78**(2), R59–R80.
- White, R.S., McKenzie, D. & O’Nions, R.K., 1992. Oceanic crustal thickness from seismic measurements and rare earth element inversions, *J. geophys. Res.*, **97**(B13), 19 683–19 715.
- Widom, M., Huhn, W.P., Maiti, S. & Steurer, W., 2014. Hybrid Monte Carlo/molecular dynamics simulation of a refractory metal high entropy alloy, *Metall. Mater. Trans., A*, **45**(1), 196–200.
- Yang, L., Meng, X. & Karniadakis, G.E., 2021. B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data, *J. Comput. Phys.*, **425**, doi:10.1016/j.jcp.2020.109913.
- Zelt, C. & Smith, R., 1992. Seismic traveltimes inversion for 2-D crustal velocity structure, *Geophys. J. Int.*, **108**(1), 16–34.
- Zhou, H.-w., 1996. A high-resolution P wave model for the top 1200 km of the mantle, *J. geophys. Res.*, **101**(B12), 27 791–27 810.